On Logics and Homomorphism Closure

Manuel Bodirsky*, Thomas Feller†, Simon Knäuer*, and Sebastian Rudolph†

* Institut für Algebra, TU Dresden

† Computational Logic Group, TU Dresden

Abstract—Predicate logic is the premier choice for specifying classes of relational structures. Homomorphisms are key to describing correspondences between relational structures. Questions concerning the interdependencies between these two means of characterizing (classes of) structures are of fundamental interest and can be highly non-trivial to answer. We investigate several problems regarding the homomorphism closure (homclosure) of the class of all (finite or arbitrary) models of logical sentences: membership of structures in a sentence's homclosure; sentence homclosedness; homclosure characterizability in a logic; normal forms for homclosed sentences in certain logics. For a wide variety of fragments of first- and second-order predicate logic, we clarify these problems' computational properties.

I. Introduction

The *homomorphism preservation theorem* states that, for any first-order sentence Φ, the property of being a model of Φ is preserved under homomorphisms[1] if and only if Φ is equivalent to an existential positive sentence. For arbitrary models, this correspondence goes back to Łoś, Tarski, and Lyndon in the 1950s, whereas Rossman showed only rather recently that it also holds when considering *finite* models only [22].

Classes of homclosed first- or second-order sentences are ubiquitous in computer science. In databases, they often serve as intuitive *query languages* and check for the presence of a certain "pattern" in a given database [24]. In particular, existential positive sentences, which can be expressed as so-called *unions of conjunctive queries*, are encountered very frequently in practice. Homclosed classes of finite structures also appear naturally in other areas of computer science; for example, the complement of any constraint satisfaction problem, viewed as a class of finite structures, is homclosed.

This paper poses and comprehensively answers the following four fundamental questions concerning homclosures:

1) **Homclosure membership.** Given a sentence Φ from some logic and a finite structure $\mathfrak{A}$, we want to decide whether $\mathfrak{A}$ belongs to the homclosure of Φ. How difficult is this problem computationally? We are interested both in the *combined complexity* (where both $\mathfrak{A}$ and Φ are given) and in the *data complexity* (where $\mathfrak{A}$ is given and Φ is fixed).
2) **Homclosedness.** Given a sentence Φ from some logic, we want to decide whether Φ is homclosed. How difficult is this problem computationally?

3) **Homclosure characterizability.** Given a sentence Φ from some logic, does there exist a sentence Ψ (from the same or a different logic) such that the class of Ψ's models is precisely the homclosure of the class of Φ's models?
4) **Homclosed normal forms.** For which logics exists a *'syntactic normal form'*, i.e., a subset of easily recognizable sentences such that an arbitrary sentence from that logic is equivalent to a sentence in normal form if and only if the former is homclosed?

This paper provides answers to these questions for (fragments of) first- and second-order logic of varying expressivity, with a particular emphasis on very expressive but decidable formalisms (including the two-variable fragment, the guarded (negation) fragment, and the triguarded fragment), popular formalisms in database theory (such as existential positive sentences and tuple-generating dependencies), and diverse prefix classes. We next summarize our findings and methods. Our results hold both for the finite- and the arbitrary-model case, so we will not make this distinction here, even if distinct proofs are sometimes required.

As for Question 1), Section IV presents results regarding decidability and combined complexity. We show that the homclosure membership problem is generically interreducible with the considered logic's satisfiability problem in a wide variety of cases, thus tightly tying decidability and complexity of homclosure membership to satisfiability; the reduction from the former to the latter is shown by logically encoding model *colorings*, inspired by similar ideas from constraint satisfaction. As notable exception, we find that for tuple-generating dependencies, the problem is undecidable despite satisfiability being trivial. Regarding the data complexity of homclosure membership, we find that question intimately related to Question 3), since in the presence of a logical characterization of a sentence's homclosure, checking homclosure membership boils down to model checking against the characterizing sentence. Along those lines, Section VI provides a variety of results, the most striking of which is probably Theorem 47, asserting polynomial time data complexity of checking membership in the homclosure of sentences in the guarded (negation) fragment of first-order logic.

Regarding Question 2) addressed in Section V, we find that under fairly weak assumptions, the satisfiability problem for the logic under consideration can be reduced to its homclosedness problem. Yet, a similar reduction can be found from satisfiability in the "dual" logic, identifying several prefix classes where homclosedness is undecidable, despite satisfiability be-

---

[1]Henceforth, we will simply call such a Φ (and its model class) *homclosed*.

ing decidable. For the remaining cases, we propose a generic reduction from homclosedness to unsatisfiability, based on the notion of *spoiler* – a homomorphism from a model to a non-model, i.e., a witness for non-homclosedness – and a way of creating spoiler-detecting sentences. Again, tuple-generating dependencies need to be coped with differently; we find that for the plain version, homclosedness is NP-complete, while adding disjunction turns the problem undecidable.

For Question 3), which we focus on in Section VI, it is important to note that already for very basic first-order sentences (including tuple-generating dependencies), their homclosure is not expressible in first-order logic, and even the question whether a given first-order sentence is expressible in any logic with a decidable model checking problem is undecidable. On the other hand, we show that the homclosure of any sentence from the Bernays-Schönfinkel class can even be characterized in existential positive first-order logic. Using elaborate model-theoretic characterizations based on *types*, we are able to show that the two-variable fragment, the triguarded fragment and the Gödel class allow for homclosure-characterization in second-order logic (in fact, even in decidable fragments thereof), while sentences in the guarded and the guarded negation fragment can be homclosure-characterized even in first-order logic with least fixed points. Employing arguments from descriptive complexity we show that these characterizability results are optimal in a certain sense.

Question 4) is dealt with in Section VII. For fragments of first-order logics, we exploit the homomorphism-preservation theorem to settle the case for all fragments fully encompassing existential positive first-order logic. We can also give a syntactic normal form for Bernays-Schönfinkel subclasses with a bounded number of quantifiers. For fragments with a decidable homclosedness problem, we can resort to this (usually costly) check to identify normal form sentences. A few cases, however, remain open. Finally, turning to full second-order logic, we are indeed able to establish a normal form for sentences whose models are closed under homomorphisms, too. As opposed to first-order logic, however, this normal form can be even computed from the given second-order sentence in polynomial time.

Detailed proofs can be found in the appendix.

II. PRELIMINARIES

We assume the reader to be familiar with standard notions regarding model theory, first- and second-order logic, decidability, and complexity theory.

Structures and Homomorphisms. We let capital Fraktur letters denote structures and the corresponding capital Roman letters their domains; i.e., A denotes the domain of $\mathfrak{A}$. We assume that structures have non-empty domains. A structure is called finite if its domain is finite. We work with signatures τ that consist of finitely many constant and predicate symbols (written in typewriter font, often simply called *constants* and *predicates* for brevity), and let $\mathrm{ar}(\mathtt{P})$ denote the arity of a

predicate $\mathtt{P} \in \tau$. For convenience, we assume that our signatures contain at least one constant symbol; this is not a severe restriction since we may always add a new constant symbol to the signature without affecting our results. The *size* $|\mathfrak{A}|$ of a finite τ-structure $\mathfrak{A}$ is the number of bits needed to represent $\mathfrak{A}$ up to isomorphism, and bounded by $(|\tau|+1)(|A|+1)^{m+1}$ where m is the maximal arity of the predicates in τ.

A *homomorphism* $h\colon \mathfrak{A} \to \mathfrak{B}$ from a τ-structure $\mathfrak{A}$ to a τ-structure $\mathfrak{B}$ is a function $h\colon A \to B$ where
- for every predicate symbol $\mathtt{P} \in \tau$, $(a_1,\ldots,a_{\mathrm{ar}(\mathtt{P})}) \in \mathtt{P}^{\mathfrak{A}}$ implies $(h(a_1),\ldots,h(a_{\mathrm{ar}(\mathtt{P})})) \in \mathtt{P}^{\mathfrak{B}}$, and
- for every constant symbol $\mathtt{c} \in \tau$, we have $h(\mathtt{c}^{\mathfrak{A}}) = \mathtt{c}^{\mathfrak{B}}$.

A homomorphism $h\colon \mathfrak{A} \to \mathfrak{B}$ is called *strong* if for every predicate symbol $\mathtt{P} \in \tau$ we have $(a_1,\ldots,a_{\mathrm{ar}(\mathtt{P})}) \in \mathtt{P}^{\mathfrak{A}}$ *exactly if* $(h(a_1),\ldots,h(a_{\mathrm{ar}(\mathtt{P})})) \in \mathtt{P}^{\mathfrak{B}}$. We use the common symbols $\hookrightarrow$ and $\twoheadrightarrow$ for injective and surjective homomorphisms, respectively, and add a horizontal stroke ($\Rightarrow$, $\hookrightarrow$, $\twoheadrightarrow$) when they are strong. Injective strong homomorphisms are also called *embeddings*.

Given a class $\mathcal{C}$ of τ-structures, let $\mathcal{C}^{\to}$ denote the *homomorphism closure of $\mathcal{C}$*, i.e., the class of all τ-structures into which some structure $\mathfrak{A} \in \mathcal{C}$ has a homomorphism. We use $\mathcal{C}^{\hookrightarrow}$ and $\mathcal{C}^{\twoheadrightarrow}$ analogously. We write $\mathcal{C}^{\to<\omega}$ for the class obtained by removing all infinite structures from $\mathcal{C}^{\to}$.

Given a τ-structure $\mathfrak{A}$ and some $\sigma \subseteq \tau$, let $\mathfrak{A}|_{\sigma}$ denote the σ-reduct of $\mathfrak{A}$. Moreover, for some constant-free signature ρ disjoint from τ, we let $\mathfrak{A} \cdot \widehat{\rho}$ denote the $\tau \uplus \rho$-structure expanding $\mathfrak{A}$ in which every element $\mathtt{P}$ of ρ is mapped to the full relation $A^{\mathrm{ar}(\mathtt{P})}$. Likewise, we let $\mathfrak{A} \cdot \widecheck{\rho}$ denote the $\tau \uplus \rho$-structure expanding $\mathfrak{A}$ in which every element of ρ is mapped to the empty relation.

For a signature ρ of only constants, let $\hat{\mathfrak{D}}_{\rho}$ denote the one-element ρ-structure (i.e., the up-to-isomorphism unique structure with all constants mapped to the only element), while $\check{\mathfrak{D}}_{\rho}$ denotes the $|\rho|$-element ρ-structure where each constant is interpreted separately. For a signature $\tau = \rho \uplus \sigma$ of constants ρ and predicates σ, we let $\mathfrak{I}_{\tau} = \check{\mathfrak{D}}_{\rho} \cdot \widecheck{\sigma}$ and $\mathfrak{F}_{\tau} = \hat{\mathfrak{D}}_{\rho} \cdot \widehat{\sigma}$.[2] We omit signature subscripts if they are clear from the context.

Logics. We give a brief description of the logics we study in this paper. By default, equality is not allowed in a logic $\mathbb{L}$; the variant of $\mathbb{L}$ where equality $=$ is allowed is denoted by $\mathbb{L}_{=}$. *First-order logic* ($\mathbb{FO}_{=}$) formulae are built from atomic formulae using existential quantification $\exists$, universal quantification $\forall$, negation $\neg$, conjunction $\wedge$, and disjunction $\vee$ in the usual way. We use the symbol $\Rightarrow$ for implication and $\Leftrightarrow$ for equivalence as the usual shortcuts for better readability. All the considered logics are fragments of *second-order logic* ($\mathbb{SO}$),[3] which extends first-order logic by *second-order quantifiers*, of the form $\exists \mathtt{P}$ or $\forall \mathtt{P}$ for a predicate symbol $\mathtt{P}$. A formula's *quantifier rank* is the maximal nesting depth of its quantifiers.

---

[2] Note that $\mathfrak{I}_{\tau}$ has a homomorphism into every τ-structure, and every τ-structure has a homomorphism into $\mathfrak{F}_{\tau}$, i.e., $\{\mathfrak{I}_{\tau}\}^{\to}$ is the class of all τ-structures, whereas $\mathfrak{F}_{\tau} \in \mathcal{C}^{\to}$ for every non-empty class $\mathcal{C}$ of τ-structures.

[3] As $=$ can be axiomatized in $\mathbb{SO}$, we do not distinguish $\mathbb{SO}$ from $\mathbb{SO}_{=}$.

- A *term* is either a variable or a constant symbol (as per our assumption that the signature may contain constant symbols but no function symbols of arities greater than zero).
- We use bold variable/term/element symbols (e.g., $\boldsymbol{x}$, $\boldsymbol{y}$, or $\boldsymbol{t}$, or $\boldsymbol{a}$) to denote finite sequences of such entities (e.g., $x_1, \ldots, x_n$) of appropriate length. Whenever order and multiplicity are irrelevant, we take the liberty to consider such sequences as sets, justifying expressions like $\boldsymbol{x} \cup \boldsymbol{y}$.
- For a (sub)formula φ, we write $\varphi[\boldsymbol{x}]$ to indicate that all free first-order variables in φ are from $\boldsymbol{x}$. We write $\varphi(\boldsymbol{x})$ to indicate a formula with free occurrences of $\boldsymbol{x}$ and use $\varphi(\boldsymbol{t})$ to denote φ with all occurrences of free variables from $\boldsymbol{x}$ replaced by the corresponding element from $\boldsymbol{t}$.

We study the following fragments of first-order logic:

- *Prefix classes* of $\mathbb{FO}$ / $\mathbb{FO}_=$. For these logics, we assume that the formulae are in prenex normal form and that the quantifier prefix is restricted by a regular expression, following standard nomenclature as used by Börger, Grädel and Gurevich [7].
- *Existential positive* $\mathbb{FO}$ ($\exists^* \mathbb{FO}_=^+$). In this fragment, $\forall$ and $\neg$ are disallowed; however, the special sentence $\bot$ is allowed in order to express falsity (it could be read as empty disjunction). We obtain (Boolean) conjunctive queries ($\mathbb{CQ}$) from $\exists^* \mathbb{FO}_=^+$ by disallowing $\vee$. The *canonical query* for a finite structure $\mathfrak{A}$, denoted by $\mathrm{cq}(\mathfrak{A})$, is the $\mathbb{CQ}$ sentence obtained by existentially quantifying over each variable x_a, for all $a \in A$, in

$$\bigwedge_{\mathsf{c} \in \tau} \mathsf{c} = x_{\mathsf{c}^{\mathfrak{A}}} \wedge \bigwedge_{\substack{\mathsf{P} \in \tau,\, k = \mathrm{ar}(\mathsf{P}) \\ (a_1, \ldots, a_k) \in \mathsf{P}^{\mathfrak{A}}}} \mathsf{P}(x_{a_1}, \ldots, x_{a_k}). \tag{1}$$

- *The guarded fragment of* $\mathbb{FO}_=$ ($\mathbb{GFO}_=$). All occurrences of $\forall$ have the form $\forall \boldsymbol{x}.(\mathsf{P}(\boldsymbol{y}) \Rightarrow \varphi[\boldsymbol{y}])$, and all of $\exists$ the form $\exists \boldsymbol{x}.(\mathsf{P}(\boldsymbol{y}) \wedge \varphi[\boldsymbol{y}])$ for $\boldsymbol{x} \subseteq \boldsymbol{y}$ and predicates P. This shape is called *guarded quantification* and $\mathsf{P}(\boldsymbol{y})$ the respective *guard*.[4]
- *The triguarded fragment of* $\mathbb{FO}$ ($\mathbb{TGF}$). Every quantification over formulae with at least two free variables must be guarded.
- *The guarded negation fragment of* $\mathbb{FO}_=$ ($\mathbb{GNFO}_=$). The symbol $\forall$ is disallowed and every occurrence of $\neg$ must be in the form of $\mathsf{P}(\boldsymbol{x}) \wedge \neg\varphi[\boldsymbol{x}]$.
- *The n-variable fragment of* $\mathbb{FO}_=$ / $\mathbb{FO}$. All variables must be from $\{x_1, \ldots, x_n\}$. As special cases we obtain the 2-variable fragment with equality ($\mathbb{FO}_=^2$) and without it ($\mathbb{FO}^2$); here we normally use the variable names x, y instead of x_1, x_2.
- *(Disjunctive) tuple-generating dependencies* (TGDs). A *disjunctive TGD* is a $\mathbb{FO}$ sentence of the form

$$\forall \boldsymbol{x}.\varphi(\boldsymbol{x}) \Rightarrow \exists \boldsymbol{y}.\psi(\boldsymbol{x}, \boldsymbol{y}), \tag{2}$$

where φ is a (possibly empty) conjunction over atoms while ψ is a non-empty disjunction over conjunctions of atoms.[5] Note that for finite structures $\mathfrak{A}, \mathfrak{B}$, the formula $\mathrm{cq}(\mathfrak{A}) \Rightarrow \mathrm{cq}(\mathfrak{B})$

---

[4]It is innocuous to conjunctively add further atoms to the guard, as long as they use only variables from $\boldsymbol{x}$. We will consider such variants part of the language; they can be LOGSPACE-transformed into the plain form. The same holds for unguarded quantification over formulae with just one free variable.

[5]Non-emptiness of disjunction is common in database theory and a deliberate choice: it ensures desirable properties like guaranteed satisfiability. Empty disjunction would make disjunctive TGDs as expressive as $\mathbb{FO}$.

---

can be easily transformed into an equivalent TGD (by pulling out the existential quantifiers of $\mathrm{cq}(\mathfrak{A})$ thus turning them into universal ones, using that the variable sets of $\mathrm{cq}(\mathfrak{A})$ and $\mathrm{cq}(\mathfrak{B})$ are disjoint), so we will consider it as such right away.

We obtain $\mathbb{DTGD}$ (sentences) as finite conjunctions over disjunctive TGDs. Likewise, we obtain $\mathbb{TGD}$ (sentences) as finite conjunctions over (non-disjunctive) TGDs, wherein the ψ are plain conjunctions over atoms. As a middle ground between $\mathbb{TGD}$ and $\mathbb{DTGD}$, we introduce *mildly disjunctive TGD sentences* ($\mathbb{MDTGD}$) as sentences of the form $\Phi \vee \Psi$ where $\Phi \in \mathbb{TGD}$ and $\Psi \in \mathbb{CQ}$. Note that any $\mathbb{MDTGD}$ sentence can be equivalently rewritten into $\mathbb{DTGD}$ in polynomial time.

Existential $\mathbb{SO}$ ($\exists \mathbb{SO}$) is the fragment of $\mathbb{SO}$ where all second-order quantifiers are existential and prenex.

We apply the standard model-theoretic semantics. For a sentence Φ, we let $\llbracket \Phi \rrbracket$ denote the class of all models of Φ and $\llbracket \Phi \rrbracket_{<\omega}$ the class of all finite models of Φ. A sentence in some logic is (finitely) satisfiable if it has a (finite) model. A logic has the *finite model property (FMP)* if satisfiability and finite satisfiability coincide. Table I gives results on the FMP for logics considered by us. A sentence Φ *(finitely) characterizes* a class $\mathcal{C}$ of structures, if $\llbracket \Phi \rrbracket_{(<\omega)} = \mathcal{C}$. For $\tau' \subseteq \tau$, we say that a τ-sentence Φ *(finitely) projectively characterizes* a class $\mathcal{C}$ of τ'-structures if $\llbracket \Phi \rrbracket_{(<\omega)}|_{\tau'} = \mathcal{C}$. A logic is *non-degenerate* if there exists an unsatisfiable sentence $\Phi_{\mathbf{false}}$.

Tilings. We introduce a variant of the $\mathbb{N} \times \mathbb{N}$ tiling problem, which serves as a versatile and useful vehicle for various subsequent undecidability arguments.

Definition 1 (margin-constrained tiling problem). *A (margin-constrained) domino system $\mathcal{D} = (D, B, L, H, V)$ consists of a finite set D of* dominoes*, some of which are* bottom-margin *dominoes $B \subseteq D$ and some* left-margin *ones $L \subseteq D$. We refer to $L \cap B$ as seed dominoes. The two binary relations $H, V \subseteq D \times D$ specify* horizontal *and* vertical *compatibility.*[6]

We call $\mathcal{D} = (D, B, L, H, V)$ deterministic if

1) *there is at most one seed domino (i.e., $|B \cap L| \leq 1$),*
2) *for every $d \in B$ there is at most one e such that $(d, e) \in H$, and in this case $e \in B$,*
3) *for every $d \in L$ there is at most one e such that $(d, e) \in V$, and in this case $e \in L$, and*
4) *for any $d_1, d_2 \in D$ there is at most one e with $(d_1, e) \in H$ and $(d_2, e) \in V$.*

A $\mathcal{D}$-tiling (of $\mathbb{N} \times \mathbb{N}$) is a function $t : \mathbb{N} \times \mathbb{N} \to D$ such that for any $k, \ell \in \mathbb{N}$ hold $t(0, \ell) \in L$ and $t(k, 0) \in B$ as well as $\big(t(k, \ell), t(k+1, \ell)\big) \in H$ and $\big(t(k, \ell), t(k, \ell+1)\big) \in V$.

A tiling t is ultimately periodic*, if there are natural numbers $k_{\mathrm{init}}, k_{\mathrm{period}}, \ell_{\mathrm{init}}, \ell_{\mathrm{period}}$ such that, for every $k, \ell \in \mathbb{N}$,*

- $t(k, \ell) = t(k + k_{\mathrm{period}}, \ell)$ *whenever $k \geq k_{\mathrm{init}}$ and*
- $t(k, \ell) = t(k, \ell + \ell_{\mathrm{period}})$ *whenever $\ell \geq \ell_{\mathrm{init}}$.*

---

[6]In the context of domino systems, it will be convenient to let $R \in \{H, V\}$ also denote the map $R : D \to 2^D$ defined by $R(x) := \{d \in D \mid (x, d) \in R\}$. The intended meaning will always be clear from the context.

It is straightforward to see that for deterministic domino systems, tilings are unique if they exist. Also note that the "classical" $\mathbb{N} \times \mathbb{N}$ tiling problem without margin constraints is a special case of this variant (choose $D = L = B$).

Lemma 2. *The following problems are both undecidable even for deterministic margin-constrained domino systems $\mathcal{D}$:*

1) *Given $\mathcal{D}$, does there exist a $\mathcal{D}$-tiling?*
2) *Given $\mathcal{D}$, does there exist an ultimately periodic $\mathcal{D}$-tiling?*

III. SOME TOOLS

In this section, we present basic tools that will be used in the subsequent sections; we start with the folklore technique of *relativization* to make statements about a model's induced substructures. In the following, let $\mathtt{U}$ denote a unary predicate. Given a $\tau \uplus \{\mathtt{U}\}$-structure $\mathfrak{A}$ we write $[\mathfrak{A}]_{\mathtt{U}}$ for the τ-reduct of the substructure of $\mathfrak{A}$ induced by $\mathtt{U}^{\mathfrak{A}}$.

Given an $\mathbb{SO}$ $\rho \uplus \tau$-sentence Φ with constants ρ and predicates $\tau \not\ni \mathtt{U}$, let $\Phi^{\mathrm{rel}(\mathtt{U})}$ denote the $\mathtt{U}$-*relativization* of Φ, defined as

$$\Phi^{\mathrm{guard}(\mathtt{U})} \wedge \bigwedge_{c \in \rho} \mathtt{U}(c), \qquad (3)$$

where $\Phi^{\mathrm{guard}(\mathtt{U})}$ is obtained from Φ by (inside-out) replacing all subformulae $\forall x. \varphi$ by $\forall x. (\mathtt{U}(x) \Rightarrow \varphi)$ and all subformulae $\exists x. \varphi$ by $\exists x. (\mathtt{U}(x) \wedge \varphi)$. Relativization preserves many aspects of the sentence, such as the number of variables, the quantifier rank, and membership in the fragments $\mathbb{GFO}_=$, $\mathbb{GNFO}_=$, $\mathbb{TGF}$, $\mathbb{FO}_=^2$, $\mathbb{TGD}$, $\mathbb{DTGD}$, $\mathbb{MDTGD}$, $\mathbb{FO}_=$ (modulo some trivial syntactic equivalence-preserving transformations) as well as all prefix classes. Obviously, $\Phi^{\mathrm{rel}(\mathtt{U})}$ is LOGSPACE-computable from Φ. Moreover, the models of $\Phi^{\mathrm{rel}(\mathtt{U})}$ are precisely those $\rho \uplus \tau \uplus \{\mathtt{U}\}$-structures $\mathfrak{A}$ for which $[\mathfrak{A}]_{\mathtt{U}} \in [\![\Phi]\!]$.

Next, we introduce some machinery for labeling structures with numbers and using such endowed structures as compact representations of other structures.

Given a τ-structure $\mathfrak{A}$ and a number $n \in \mathbb{N}$, an n-*labeling of $\mathfrak{A}$* is a function $\lambda \colon A \to \{1, \ldots, n\}$. We will then call $(\mathfrak{A}, \lambda)$ an n-*labeled structure*. If $(\mathfrak{A}, \lambda)$ satisfies $\lambda(c^{\mathfrak{A}}) = 1$ for every constant $c \in \tau$, we call it *constant-sole*. Given an n-labeled τ-structure $(\mathfrak{A}, \lambda)$, its *implicit representation* $\mathfrak{A}_\lambda$ is the $\tau \uplus \{\mathtt{Bit}_0^\lambda, \ldots, \mathtt{Bit}_\ell^\lambda\}$-structure with $\ell = \lceil \log_2 n \rceil$ expanding $\mathfrak{A}$ by $(\mathtt{Bit}_i^\lambda)^{\mathfrak{A}_\lambda} = \{a \mid \lfloor \lambda(a)/2^i \rfloor \text{ is odd}\}$.

We next define $\mathbb{FO}_=$ formulae allowing to compare some natural number against the label of a domain element associated with some term t in $\mathfrak{A}_\lambda$. Given some $m \in \{1, \ldots, n\}$, assume $b_1 \cdots b_\ell \in \{0, 1\}^\ell$ is the fixed-length binary representation of $m-1$ (i.e., $m-1 = \sum_{i=1}^\ell b_i \cdot 2^{\ell-i}$). Then we use the following abbreviations for $\mathbb{FO}_=$ sentences:

$$[\lambda(t) = m] := \left(\bigwedge_{\substack{1 \leq j \leq \ell \\ b_j = 0}} \neg \mathtt{Bit}_j^\lambda(t) \right) \wedge \left(\bigwedge_{\substack{1 \leq j \leq \ell \\ b_j = 1}} \mathtt{Bit}_j^\lambda(t) \right), \quad (4)$$

$$[\lambda(t) \geq m] := \bigwedge_{\substack{1 \leq j \leq \ell \\ b_j = 1}} \left(\mathtt{Bit}_j^\lambda(t) \vee \bigvee_{\substack{1 \leq i \leq j-1 \\ b_i = 0}} \mathtt{Bit}_i^\lambda(t) \right). \quad (5)$$

These formulae's size is polynomial in ℓ, and, for some $\mathfrak{A}$ and a variable assignment $v \colon V \to A$, we get $\mathfrak{A}_\lambda, v \models [\lambda(t) = m]$ whenever $\lambda(t^{\mathfrak{A}, v}) = m$; analogously for $[\lambda(t) \geq m]$.

For a constant-sole $(\mathfrak{A}, \lambda)$, let its *unfolding* $\mathfrak{A}^\lambda$ be the τ-structure with $A^\lambda = \{(a, i) \mid a \in A,\ 1 \leq i \leq \lambda(a)\}$ where, for any $\mathtt{P} \in \tau$ of arity k, holds $((a_1, m_1), \ldots, (a_k, m_k)) \in \mathtt{P}^{\mathfrak{A}^\lambda}$ iff $(a_1, \ldots, a_k) \in \mathtt{P}^{\mathfrak{A}^\lambda}$. We define $\pi \colon A^\lambda \to A$ via $(a, m) \mapsto a$. Obviously, π is a strong surjective homomorphism from $\mathfrak{A}^\lambda$ to $\mathfrak{A}$. Conversely, for any strong surjective homomorphism $h \colon \mathfrak{B} \to \mathfrak{A}$ where every $a \in A$ satisfies $|h^{-1}(a)| \leq n$ (i.e., the size of each element's pre-image is bounded by n) there exists some $\lambda \colon A \to \{1, \ldots, n\}$ with $\mathfrak{B} \cong \mathfrak{A}^\lambda$ (witnessed by an isomorphism $g \colon \mathfrak{B} \to \mathfrak{A}^\lambda$ such that $h = \pi \circ g$).

For variable labelings $as \colon V \to \{1, \ldots, n\}$, let $as^+(t)$ denote $as(t)$ if t is a variable and 1 if t is a constant. We recursively define the functions tr_{as}^n mapping $\mathbb{FO}_=$ τ-formulae in negation normal form to $\mathbb{FO}_=$ $\tau \uplus \{\mathtt{Bit}_0^\lambda, \ldots, \mathtt{Bit}_{\lceil \log_2 n \rceil}^\lambda\}$-formulae as follows:

$$\mathrm{tr}_{as}^n(\mathtt{P}(\boldsymbol{t})) = \mathtt{P}(\boldsymbol{t}) \qquad (6)$$

$$\mathrm{tr}_{as}^n(\neg \mathtt{P}(\boldsymbol{t})) = \neg \mathtt{P}(\boldsymbol{t}) \qquad (7)$$

$$\mathrm{tr}_{as}^n(t_1 = t_2) = \begin{cases} t_1 = t_2 & \text{if } as^+(t_1) = as^+(t_2) \\ \bot & \text{otherwise} \end{cases} \qquad (8)$$

$$\mathrm{tr}_{as}^n(t_1 \neq t_2) = \begin{cases} t_1 \neq t_2 & \text{if } as^+(t_1) = as^+(t_2) \\ \top & \text{otherwise} \end{cases} \qquad (9)$$

$$\mathrm{tr}_{as}^n(\varphi \wedge \psi) = \mathrm{tr}_{as}^n(\varphi) \wedge \mathrm{tr}_{as}^n(\psi) \qquad (10)$$

$$\mathrm{tr}_{as}^n(\varphi \vee \psi) = \mathrm{tr}_{as}^n(\varphi) \vee \mathrm{tr}_{as}^n(\psi) \qquad (11)$$

$$\mathrm{tr}_{as}^n(\exists x. \varphi) = \bigvee_{1 \leq i \leq n} \exists x. [\lambda(x) \geq i] \wedge \mathrm{tr}_{as \cup \{x \mapsto i\}}^n(\varphi) \qquad (12)$$

$$\mathrm{tr}_{as}^n(\forall x. \varphi) = \bigwedge_{1 \leq i \leq n} \forall x. [\lambda(x) \geq i] \Rightarrow \mathrm{tr}_{as \cup \{x \mapsto i\}}^n(\varphi) \qquad (13)$$

Defining tr^n by letting $\mathrm{tr}^n(\Phi) = \mathrm{tr}_\emptyset^n(\mathrm{NNF}(\Phi))$, we obtain a lemma stating that tr^n truthfully rewrites sentences for their evaluation over implicit representations of unfoldings.

Lemma 3. *Let Φ be any $\mathbb{FO}_=$ sentence. Then $\mathfrak{A}^\lambda \models \Phi$ exactly if $\mathfrak{A}_\lambda \models \mathrm{tr}^n(\Phi)$.*

Note that the size of $\mathrm{tr}^n(\Phi)$ may be exponential in the size of Φ. On the positive side, tr^n preserves (negation-)guardedness, number of variables, quantifier prefixes, and quantifier rank.

IV. CHECKING HOMCLOSURE MEMBERSHIP

The first problem we address is determining whether a given finite structure is contained in the homclosure of the set of models (or finite models) of a given sentence Φ.

Problem: $\mathrm{InHomCl}_{(\mathrm{fin})}$
Input: τ, τ-sentence Φ, finite τ-structure $\mathfrak{A}$.
Output: YES, if $\mathfrak{A} \in [\![\Phi]\!]_{(<\omega)}^{\to}$, NO otherwise.

The fact that Φ is (finitely) satisfiable iff $\mathfrak{F}_\tau \in [\![\Phi]\!]_{(<\omega)}^{\to}$, already yields some lower bounds for this problem's difficulty.

Proposition 4. *(Finite) satisfiability can be* LOGSPACE-*reduced to* InHomCl$_{(\text{fin})}$.

This shows that InHomCl$_{(\text{fin})}$ is undecidable for any logic with an undecidable (finite) satisfiability problem, such as $\mathbb{FO}$ and every logic containing it, but also all $\mathbb{FO}$ reduction classes.

Toward some decidable cases and corresponding optimal upper bounds, we employ and formalize the straightforward idea of "coloring" a model with elements of the target structure.

Definition 5 (extrinsic/intrinsic $\mathfrak{A}$-coloring). *Let $\mathfrak{A}$ be a finite τ-structure over the domain $A = \{1, \ldots, n\}$. As before, we use a signature $\rho = \{\text{Bit}_1^\lambda, \ldots, \text{Bit}_\ell^\lambda\}$ of unary predicates with $\ell = \lceil \log_2 n \rceil$ to represent n-labeled structures.*

Given a predicate $\text{P} \in \tau$ of arity k and a k-tuple of terms $\boldsymbol{t} = (t_1, \ldots, t_k)$, we let $\chi_{\mathfrak{A}}^{\text{P}}(\boldsymbol{t})$ denote the formula

$$\text{P}(\boldsymbol{t}) \wedge \bigvee_{(m_1,\ldots,m_k) \in \text{P}^{\mathfrak{A}}} \bigwedge_{i=1}^{k} [\lambda(t_i) = m_i]. \tag{14}$$

We denote by $\Omega_{\mathfrak{A}}$ the following $\tau \uplus \rho$-sentence:

$$\bigwedge_{\substack{\text{P} \in \tau \\ k = ar(\text{P})}} \forall x_1 \ldots x_k . \text{P}(x_1, \ldots, x_k) \Rightarrow \chi_{\mathfrak{A}}^{\text{P}}(x_1, \ldots, x_k). \tag{15}$$

Given a τ-sentence Φ and a finite τ-structure $\mathfrak{A}$, we define the extrinsic $\mathfrak{A}$-coloring *of Φ by $\Phi_{\mathfrak{A}}^{\text{ext}} = \Phi \wedge \Omega_{\mathfrak{A}}$. The* intrinsic $\mathfrak{A}$-coloring *of Φ, denoted $\Phi_{\mathfrak{A}}^{\text{int}}$ is obtained from Φ by replacing, for all predicates $\text{P} \in \tau$, any occurrence of $\text{P}(\boldsymbol{t})$ by $\chi_{\mathfrak{A}}^{\text{P}}(\boldsymbol{t})$.*

Note that both $\Phi_{\mathfrak{A}}^{\text{ext}}$ and $\Phi_{\mathfrak{A}}^{\text{int}}$ are of polynomial size wrt. $|\mathfrak{A}|$ and the size of Φ. The sentence $\Omega_{\mathfrak{A}}$ projectively characterizes the class of τ-structures having a homomorphism into $\mathfrak{A}$, thus $\Phi_{\mathfrak{A}}^{\text{ext}}$ projectively characterizes the class of (finite) models of Φ having a homomorphism into $\mathfrak{A}$. Also, every model of Φ having a homomorphism into $\mathfrak{A}$ can be extended to a model of $\Phi_{\mathfrak{A}}^{\text{int}}$, and conversely, for every model $\mathfrak{B}$ of $\Phi_{\mathfrak{A}}^{\text{int}}$, the structure $\mathfrak{C}$ obtained by setting $C := B$, $\text{c}^{\mathfrak{C}} := \text{c}^{\mathfrak{B}}$ for all constants, and $\text{P}^{\mathfrak{C}} := (\chi_{\mathfrak{A}}^{\text{P}})^{\mathfrak{B}}$ is a model of Φ with a homomorphism into $\mathfrak{A}$. Hence we obtain the following lemma.

Lemma 6. *Let Φ be an arbitrary τ-sentence and let $\mathfrak{A}$ be a finite τ-structure. The following are equivalent:*

1) Φ has a (finite) model admitting a homomorphism into $\mathfrak{A}$.
2) $\Phi_{\mathfrak{A}}^{\text{int}}$ is (finitely) satisfiable.
3) $\Phi_{\mathfrak{A}}^{\text{ext}}$ is (finitely) satisfiable.

As a direct consequence, arbitrary and finite homclosure membership coincide for a wide range of fragments.

Theorem 7. *Let $\mathbb{L} \subseteq \mathbb{SO}$ and $\mathbb{L}^{\text{ext}} := \{\Phi_{\mathfrak{A}}^{\text{ext}} \mid \Phi \in \mathbb{L}, \mathfrak{A} \text{ finite}\}$ as well as $\mathbb{L}^{\text{int}} := \{\Phi_{\mathfrak{A}}^{\text{int}} \mid \Phi \in \mathbb{L}, \mathfrak{A} \text{ finite}\}$. If $\mathbb{L}^{\text{ext}}$ or $\mathbb{L}^{\text{int}}$ has the FMP, then $[\![\Psi]\!]^{\rightarrow <\omega} = [\![\Psi]\!]_{<\omega}^{\rightarrow <\omega}$ for every $\Psi \in \mathbb{L}$.*

This theorem applies to various fragments, including any (sublogic) of $\mathbb{DTGD}$, $\mathbb{GNFO}_=$, $\mathbb{TGF}$, $\mathbb{FO}_=^2$, $\exists^*\forall\forall\exists^*\mathbb{FO}$, or $\exists^*\forall^*\mathbb{FO}_=$. We obtain the following complexity results.

Theorem 8. InHomCl$_{(\text{fin})}$ *is complete for*

1) 2EXPTIME *for* $\mathbb{GNFO}_=$,
2) N2EXPTIME *for* $\mathbb{TGF}$,
3) NEXPTIME *for* $\mathbb{FO}_=^2$, $\exists^*\forall\forall\exists^*\mathbb{FO}$, $\exists^*\forall^*\mathbb{FO}_=$, $\forall^*\mathbb{FO}_=$,
4) NP *for* $\exists^*\mathbb{FO}_=$ *and* $\exists^*\mathbb{FO}_=^+$.

Next to the combined complexity addressed in the previous theorem, one might also ask for the complexity if the sentence is fixed and only the structure of interest varies. We will come back to this question in the course of Section VI.

We conclude with a case where the difficulty of InHomCl$_{(\text{fin})}$ does not coincide with that of satisfiability (in the most egregious way). We recap that $\mathbb{TGD}$ (just like $\mathbb{MDTGD}$ and $\mathbb{DTGD}$) has a trivial satisfiability problem (in that every $\mathbb{TGD}$ sentence is satisfiable and finitely satisfiable, with $\mathfrak{F}$ uniformly serving as a model), yet we present a fixed $\mathbb{TGD}$ sentence Φ_{grid} for which InHomCl$_{(\text{fin})}$ is undecidable.

Definition 9. *We let $\mathfrak{A}_{\mathbb{N} \times \mathbb{N}}$ denote the $\mathbb{N} \times \mathbb{N}$ grid, i.e., the $\{\text{H}, \text{V}\}$-structure with domain $A_{\mathbb{N} \times \mathbb{N}} = \mathbb{N} \times \mathbb{N}$ that satisfies $\text{H}^{\mathfrak{A}_{\mathbb{N} \times \mathbb{N}}} = \{((k,\ell),(k+1,\ell)) \mid k, \ell \in \mathbb{N}\}$ as well as $\text{V}^{\mathfrak{A}_{\mathbb{N} \times \mathbb{N}}} = \{((k,\ell),(k,\ell+1)) \mid k, \ell \in \mathbb{N}\}$.*

Let Φ_{grid} denote the $\mathbb{TGD}$ sentence

$$\begin{pmatrix} \forall x \exists y . \text{H}(x,y) \end{pmatrix} \wedge \begin{pmatrix} \forall x \exists y . \text{V}(x,y) \end{pmatrix} \wedge \\ \begin{pmatrix} \forall xyzv . \text{H}(x,y) \wedge \text{V}(x,z) \wedge \text{V}(y,v) \Rightarrow \text{H}(z,v) \end{pmatrix}. \tag{16}$$

For a given domino system $\mathcal{D} = (D, D, D, H, V)$, we define the finite $\{\text{H}, \text{V}\}$-structure $\mathfrak{A}_{\mathcal{D}}$ as follows: $A_{\mathcal{D}} = D$ as well as $\text{H}^{\mathfrak{A}_{\mathcal{D}}} = H$ and $\text{V}^{\mathfrak{A}_{\mathcal{D}}} = V$.

Noting that every model $\mathfrak{B}$ of Φ_{grid} allows for a homomorphism $h \colon \mathfrak{A}_{\mathbb{N} \times \mathbb{N}} \to \mathfrak{B}$, it follows that every homomorphism from some (finite) model of Φ_{grid} into $\mathfrak{A}_{\mathcal{D}}$ gives rise to a(n ultimately periodic) $\mathcal{D}$-tiling. Thus, by Lemma 2, both membership in $[\![\Phi_{\text{grid}}]\!]^{\rightarrow}$ and $[\![\Phi_{\text{grid}}]\!]_{<\omega}^{\rightarrow}$ are undecidable problems.

Theorem 10. InHomCl$_{(\text{fin})}$ *is undecidable for* $\mathbb{TGD}$ *(and hence for* $\mathbb{MDTGD}$ *and* $\mathbb{DTGD}$*).*

V. CHECKING HOMCLOSEDNESS

The next problem we investigate is the problem of determining whether a given sentence is homclosed (that is, the class of its models is closed under homomorphisms).

Problem: HomClosed
Input: τ, τ-sentence Φ.
Output: YES, if $[\![\Phi]\!] = [\![\Phi]\!]^{\rightarrow}$, NO otherwise.

For the finite-model version of this problem, note that any structure has a homomorphism into some infinite structure. Thus, any non-empty class $\mathcal{C}$ of finite structures cannot be homomorphism-closed (whence the question if $[\![\Phi]\!]_{<\omega} = [\![\Phi]\!]_{<\omega}^{\rightarrow}$ trivializes to the question if $[\![\Phi]\!]_{<\omega} = \emptyset$). This motivates the subsequent definition, focusing on finite target structures.

Problem: HomClosed$_{\text{fin}}$
Input: τ, τ-sentence Φ.
Output: YES, if $[\![\Phi]\!]_{<\omega} = [\![\Phi]\!]_{<\omega}^{\rightarrow <\omega}$, NO otherwise.

To begin with, obviously, both problems are trivial (i.e., the answer is always YES) for fragments where every sentence's model class is homomorphism-closed, such as $\exists^*\mathbb{FO}_=^\pm$.

Next, we will see that under very mild assumptions, (finite) satisfiability within the considered logic can be reduced to HomClosed$_{(\text{fin})}$, yielding lower bounds for the latter in many cases. To this end, we state the following technical result.

Lemma 11. *Let $\mathbb{L}, \mathbb{L}' \subseteq \mathbb{SO}$ and $f: \mathbb{L}' \to \mathbb{L}$ such that*
- *(finite) unsatisfiability in $\mathbb{L}'$ is C-hard and*
- *f is C'-computable and maps τ-sentences to $\tau \cup \{U\} \cup \tau'$-sentences with $[\![f(\Phi)]\!]_{(<\omega)}|_{\tau \cup \{U\}} = \{(\mathfrak{A}, \emptyset) \mid \mathfrak{A} \in [\![\Phi]\!]_{(<\omega)}\}$.*

Then HomClosed$_{(\text{fin})}$ *for $\mathbb{L}$ is C-hard under C'-reductions.*

In particular, Lemma 11 shows that for any logic allowing for extending any sentence Φ into $\Phi \wedge \forall x.\neg U(x)$, (finite) homclosedness must be as hard as (finite) unsatisfiability. As a direct consequence of this lemma, we also obtain undecidability for all logics containing $\mathbb{FO}$, but also for all undecidable $\mathbb{FO}_=$ prefix classes featuring a universal quantifier (such as $\forall\forall\forall\exists\mathbb{FO}$ or $\forall\forall\exists\mathbb{FO}_=$), even if the signature does not contain constants. As in the previous section, this justifies focusing our attention to logics with a decidable satisfiability problem. However, it turns out that even for many of those, HomClosed$_{(\text{fin})}$ is undecidable, as we will show now. To this end, we establish a companion result to Lemma 11.

Lemma 12. *Let $\mathbb{L}, \mathbb{L}' \subseteq \mathbb{SO}$ and $f: \mathbb{L}' \to \mathbb{L}$ such that*
- *(finite) unsatisfiability in $\mathbb{L}'$ is C-hard and*
- *f is C'-computable and maps τ-sentences to $\tau \cup \{U\}$-sentences with $[\![f(\Phi)]\!]_{(<\omega)} = \{(\mathfrak{A}, \tilde{A}) \mid \mathfrak{A} \in [\![\neg\Phi]\!]_{(<\omega)} \text{ or } \tilde{A} \neq A\}$.*

Then HomClosed$_{(\text{fin})}$ *for $\mathbb{L}$ is C-hard under C'-reductions.*

In particular, this result shows that for any logics $\mathbb{L}$ and $\mathbb{L}'$ where one can easily find for any $\mathbb{L}'$ sentence Φ an $\mathbb{L}$ sentence equivalent to $\neg\Phi \vee \exists x.\neg U(x)$, HomClosed$_{(\text{fin})}$ for $\mathbb{L}$ is at least as hard as unsatisfiability for $\mathbb{L}'$. Picking $\forall\forall\forall\exists\mathbb{FO}$ and $\forall\forall\exists\mathbb{FO}_=$ for $\mathbb{L}'$, this immediately entails undecidability of HomClosed$_{(\text{fin})}$ for the prefix classes $\exists\exists\exists\forall\mathbb{FO}$ and $\exists\exists\forall\mathbb{FO}_=$, despite decidability of satisfiability. Again, this even holds when constants are absent from the signature. Of course, this also entails undecidability of HomClosed$_{(\text{fin})}$ for the decidable prefix classes $\exists^*\forall\forall\exists^*\mathbb{FO}$ (the Gödel class) and $\exists^*\forall^*\mathbb{FO}_=$ (the Bernays-Schönfinkel class).

After these negative results, we will now introduce a generic method for establishing decidability and even tight complexity bounds of HomClosed$_{(\text{fin})}$ for a wide variety of logics.

Definition 13 (spoiler, FSP). *Let Φ be a τ-sentence. A (homclosedness) spoiler for Φ is a triple $(\mathfrak{A}, \mathfrak{B}, s)$, written as $\mathfrak{A} \xrightarrow{s} \mathfrak{B}$, where $\mathfrak{A} \in [\![\Phi]\!]$, $\mathfrak{B} \notin [\![\Phi]\!]$, and $s: \mathfrak{A} \to \mathfrak{B}$. We call a spoiler* injective, *surjective,* strong, *etc. if s is (and use*

the appropriate arrow symbol). We call it finite *if both $\mathfrak{A}$ and $\mathfrak{B}$ are. A logic $\mathbb{L}$ has the* finite spoiler property *(FSP) if each $\mathbb{L}$ sentence that has a spoiler also has a finite one.*

Obviously, $[\![\Phi]\!]^{\rightarrow} \neq [\![\Phi]\!]$ iff Φ has a spoiler, and $[\![\Phi]\!]_{<\omega}^{\rightarrow <\omega} \neq [\![\Phi]\!]_{<\omega}$ iff Φ has a finite spoiler. Hence, homclosedness is verified or refuted by the absence or presence, respectively, of such a spoiler, the search for which can be confined using a straightforward decomposition result.

Proposition 14. *For any two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$, any homomorphism $h: \mathfrak{A} \to \mathfrak{B}$ can be represented as the composition $h_2 \circ h_1$ of two homomorphisms where*
- *$h_1: \mathfrak{A} \hookrightarrow \mathfrak{C}$ is injective and*
- *$h_2: \mathfrak{C} \twoheadrightarrow \mathfrak{B}$ is surjective and strong.*

The same holds when consideration is restricted exclusively to (homomorphisms between) finite structures.

Corollary 15. *Let Φ be an arbitrary τ-sentence. Then $[\![\Phi]\!]_{(<\omega)} \neq [\![\Phi]\!]_{(<\omega)}^{\rightarrow (<\omega)}$ exactly if Φ has a (finite) injective spoiler or a (finite) strong surjective spoiler.*

We will first tend to the case of injective spoilers.

Definition 16. *Let Φ be a τ-sentence. Let τ' consist of copies P' for all predicates $P \in \tau$. Let Φ^- be the $\tau \cup \tau'$-sentence obtained from Φ by replacing every occurrence of some atom $P(t)$ by $P(t) \wedge P'(t)$. Then, let $\Phi^\natural_{\hookrightarrow}$ denote the $\tau \cup \tau' \cup \{U\}$-sentence $\neg\Phi \wedge (\Phi^-)^{\text{rel}(U)}$.*
A witness of an injective spoiler $\mathfrak{A} \xrightarrow{s} \mathfrak{B}$ is a $\tau \uplus \tau' \uplus \{U\}$-structure $\mathfrak{D}$ satisfying $\mathfrak{D}|_\tau = \mathfrak{B}$, $U^{\mathfrak{D}} = s(A)$, and $P^{\mathfrak{D}} \cap P'^{\mathfrak{D}} = \{(s(a_1), \ldots, s(a_k)) \mid (a_1, \ldots, a_k) \in P^{\mathfrak{A}}\}$ for every predicate P from τ.

Note that any (finite) injective spoiler of some Φ has a (finite) witness but also any such witness is a (finite) model of $\Phi^\natural_{\hookrightarrow}$. Conversely, any (finite) model of $\Phi^\natural_{\hookrightarrow}$ is clearly a witness of some (finite) injective spoiler of Φ. So we obtain the following.

Lemma 17. *$\Phi^\natural_{\hookrightarrow}$ is (finitely) satisfiable exactly if Φ has a (finite) injective spoiler.*

We next tend to the case of strong surjective spoilers, focusing on the case where Φ is from $\mathbb{FO}_=$. Before, note that for any equality-free logic (in particular any fragment of $\mathbb{FO}$), sentences are naturally preserved under strong surjective homomorphisms. Thus, for any such logic, $[\![\Phi]\!]_{(<\omega)}^{\rightarrow (<\omega)} = [\![\Phi]\!]_{(<\omega)}$ holds precisely if $\Phi^\natural_{\hookrightarrow}$ is (finitely) unsatisfiable.

Lemma 18. *Let $\mathfrak{B}$ and $\mathfrak{A}$ be τ-structures, let $h: \mathfrak{B} \twoheadrightarrow \mathfrak{A}$ be a strong surjective homomorphism. Let n be a natural number. Then there exists a function $\lambda: A \to \{1, \ldots, n\}$ such that*
- *there is a strong surjective homomorphism $h': \mathfrak{B} \twoheadrightarrow \mathfrak{A}^\lambda$ satisfying $h = \pi \circ h'$, and*
- *$\mathfrak{B}$ and $\mathfrak{A}^\lambda$ satisfy the same $\mathbb{FO}_=$ sentences of quantifier rank n.*

Corollary 19. *If an $\mathbb{FO}_=$ sentence Φ of quantifier rank n has a strong surjective spoiler, then it has a spoiler of the form $\mathfrak{A}^\lambda \xrightarrow{\pi} \mathfrak{A}$ for some $\lambda: A \to \{1, \ldots, n\}$ (which is automatically strong and surjective).*

Thus, in our quest to detect the existence of strong surjective spoilers, we can focus on spoilers of this specific type.

Definition 20. *For a $\mathbb{FO}_=$ sentence Φ, let $\Phi^{\maltese}_{\Rrightarrow}$ denote $\bot$ if Φ is equality-free, otherwise it denotes the sentence $\neg\Phi \wedge \mathrm{tr}^n_\emptyset(\Phi)$ where n is the quantifier rank of Φ.*
Given a strong surjective spoiler of Φ of the form $\mathfrak{A}^\lambda \xrightarrow{\pi}_{\Rrightarrow} \mathfrak{A}$, its witness *is the $\tau \uplus \{\mathtt{Bit}^\lambda_0, \ldots, \mathtt{Bit}^\lambda_{\lceil \log_2 n \rceil}\}$-structure $\mathfrak{A}_\lambda$.*

We note that $\Phi^{\maltese}_{\Rrightarrow}$ may be of exponential size in Φ. With help of Lemma 3 we obtain that, on the one hand, the witness of a (finite) Φ-spoiler of the form $\mathfrak{A}^\lambda \xrightarrow{\pi}_{\Rrightarrow} \mathfrak{A}$ is a (finite) model of $\Phi^{\maltese}_{\Rrightarrow}$, and, on the other hand, that any (finite) model of $\Phi^{\maltese}_{\Rrightarrow}$ is isomorphic to a witness of some Φ-spoiler of the form $\mathfrak{A}^\lambda \xrightarrow{\pi}_{\Rrightarrow} \mathfrak{A}$. This gives rise to the following characterization.

Lemma 21. $\Phi^{\maltese}_{\Rrightarrow}$ *is (finitely) satisfiable exactly if Φ has a (finite) strong surjective spoiler.*

As a next step, we combine the obtained characterizations.

Definition 22. *Given a $\mathbb{FO}_=$ sentence Φ, let $\Phi^{\maltese} = \Phi^{\maltese}_{\hookrightarrow} \vee \Phi^{\maltese}_{\Rrightarrow}$.*

We find that $\Phi^{\maltese}$ is in $\mathbb{GNFO}_=$, $\mathbb{TGF}$, and $\mathbb{FO}^2_=$ whenever Φ is; it can be easily rewritten into $\exists^*\forall\forall\exists^*\mathbb{FO}$ whenever Φ is in $\exists\exists\forall\forall\mathbb{FO}$ or $\forall\forall\exists\exists\mathbb{FO}$; and into $\exists^*\forall^*\mathbb{FO}_=$ whenever Φ is in $\exists^*\mathbb{FO}_=$ or $\forall^*\mathbb{FO}_=$. Thanks to Lemma 17, Lemma 21, and Proposition 14, $\Phi^{\maltese}$ is (finitely) satisfiable whenever a (finite) spoiler of Φ exists. So we obtain the following.

Theorem 23. Φ *is (finitely) homclosed if and only if $\Phi^{\maltese}$ is (finitely) unsatisfiable.*

Unfortunately, the size of $\Phi^{\maltese}$ is exponential in Φ and hence too big to directly yield tight complexity results. However, whenever we know that $\Phi^{\maltese}$ belongs to a fragment with the finite model property, Theorem 23 allows us to establish that the existence of arbitrary spoilers and finite spoilers coincide.

Theorem 24. $\mathbb{GNFO}_=$, $\mathbb{TGF}$, $\mathbb{FO}^2_=$, $\forall^*\mathbb{FO}_=$, $\exists^*\mathbb{FO}_=$, $\exists\exists\forall\forall\mathbb{FO}$, *and* $\forall\forall\exists\exists\mathbb{FO}$ *have the finite spoiler property. Thus,* HomClosed *and* HomClosed$_{\mathrm{fin}}$ *coincide for these fragments.*

This justifies, for all the mentioned fragments, to focus on the case of finite spoilers. To this end, we introduce a very specific kind of strong surjective homomorphism, which only merges two "indistinguishable" domain elements.

Definition 25 (monomerge). *An* monomerge *is a surjective strong homomorphism $h \colon \mathfrak{A} \twoheadrightarrow \mathfrak{B}$ satisfying $B = A \setminus \{a\}$ for some $a \in A$ and $h(a') = a'$ for every $a' \neq a$. We call a the* merge source *and $h(a)$ the* merge target *of h. We will use $\rightarrowtail$ to specifically refer to homomorphisms that are monomerges.*

It is not hard to see that in the finite case, every strong surjective homomorphism can be realized, up to isomorphism, by iteratively merging pairs of domain elements in this way.

Lemma 26. *Let $\mathfrak{A}_0$, $\mathfrak{B}$ be finite τ-structures and $h \colon \mathfrak{A}_0 \twoheadrightarrow \mathfrak{B}$ a strong surjective homomorphism. Then there is a finite sequence $h_1 \colon \mathfrak{A}_0 \rightarrowtail \mathfrak{A}_1, \ldots, h_m \colon \mathfrak{A}_{m-1} \rightarrowtail \mathfrak{A}_m$ of monomerges and an isomorphism $g \colon \mathfrak{A}_m \to \mathfrak{B}$ such that $h = g \circ h_m \circ \ldots \circ h_1$.*

Thus, in the finite, our search for strong surjective spoilers can be confined to the search for monomerge spoilers.

Corollary 27. *If a sentence Φ has a finite strong surjective spoiler, then it has a finite monomerge spoiler. Consequently, whenever Φ has an arbitrary finite spoiler, it must have an injective spoiler or a monomerge spoiler.*

Definition 28. *For a τ-sentence Φ, we let $\Phi^{\maltese}_{\rightarrowtail}$ denote $\bot$ if Φ is equality-free, otherwise it denotes the $\tau \uplus \{\mathsf{c}_s, \mathsf{c}_t, \dot{\mathsf{U}}\}$-sentence*

$$\Phi \wedge (\neg\Phi)^{\mathrm{rel}(\dot{\mathsf{U}})} \wedge \Omega_{\mathsf{c}_s \simeq \mathsf{c}_t} \wedge \forall x.\big(x = \mathsf{c}_s \Leftrightarrow \neg\dot{\mathsf{U}}(x)\big), \quad (17)$$

where $\Omega_{\mathsf{c}_s \simeq \mathsf{c}_t}$ stands for

$$\bigwedge_{\substack{\mathsf{P} \in \tau \\ k = ar(\mathsf{P})}} \bigwedge_{\substack{1 \le i \le k \\ \boldsymbol{x} = (x_1, \ldots, x_{i-1}) \\ \boldsymbol{y} = (x_{i+1}, \ldots, x_k)}} \forall \boldsymbol{x}\boldsymbol{y}.\mathsf{P}(\boldsymbol{x}, \mathsf{c}_s, \boldsymbol{y}) \Leftrightarrow \mathsf{P}(\boldsymbol{x}, \mathsf{c}_t, \boldsymbol{y}). \quad (18)$$

Let $\mathfrak{A} \xrightarrow{h}_{\rightarrowtail} \mathfrak{B}$ be a monomerge spoiler of Φ with merge source a_s and merge target a_t. Then, we call the $\tau \uplus \{\mathsf{c}_s, \mathsf{c}_t, \dot{\mathsf{U}}\}$-structure $\big(\mathfrak{A}, a_s, a_t, B\big)$ the witness *of $\mathfrak{A} \xrightarrow{h}_{\rightarrowtail} \mathfrak{B}$.*

We note that the witness of a finite monomerge spoiler of Φ is a finite model of $\Phi^{\maltese}_{\rightarrowtail}$. Conversely, any finite model of $\Phi^{\maltese}_{\rightarrowtail}$ is isomorphic to the witness of a finite monomerge spoiler of Φ. Thus we obtain the following characterization.

Lemma 29. $\Phi^{\maltese}_{\rightarrowtail}$ *is finitely satisfiable exactly if Φ has a finite monomerge spoiler.*

Definition 30. *Given a sentence Φ we let $\Phi^{\maltese}_{<\omega} = \Phi^{\maltese}_{\hookrightarrow} \vee \Phi^{\maltese}_{\rightarrowtail}$.*

We note a few properties of $\Phi^{\maltese}_{<\omega}$: It is polytime computable and of polynomial size with respect to Φ. It is in $\mathbb{GNFO}_=$, $\mathbb{TGF}$, and $\mathbb{FO}^2_=$ whenever Φ is. It is in $\exists^*\forall\forall\exists^*\mathbb{FO}$ whenever Φ is in $\exists\exists\forall\forall\mathbb{FO}$ or $\forall\forall\exists\exists\mathbb{FO}$ and it is in $\exists^*\forall^*\mathbb{FO}_=$ whenever Φ is in $\exists^*\mathbb{FO}_=$ or $\forall^*\mathbb{FO}_=$. By Corollary 27 as well as Lemma 17 and Lemma 29, we also obtain that $\Phi^{\maltese}_{<\omega}$ is finitely satisfiable whenever a finite spoiler of Φ exists, which gives rise to the following characterization.

Theorem 31. Φ *is finitely homclosed if and only if $\Phi^{\maltese}_{<\omega}$ is finitely unsatisfiable.*

Now we have all the bits and pieces together to establish a bunch of tight complexity results – both for the finite and the arbitrary case.

Theorem 32. HomClosed$_{(\mathrm{fin})}$ *for*

1) $\mathbb{GNFO}_=$ *is* 2EXPTIME-*complete,*
2) $\mathbb{TGF}$ *is* CON2EXPTIME-*complete,*
3) *any of $\mathbb{FO}^2_=$, $\forall^*\mathbb{FO}_=$, $\exists^*\mathbb{FO}_=$, $\forall\forall\exists\exists\mathbb{FO}$, and $\exists\exists\forall\forall\mathbb{FO}$ is* CONEXPTIME-*complete.*

Finally, we turn to $\mathbb{TGD}$, for which the above machinery fails to work (despite the trivial (un)satisfiability problem) as $\mathbb{TGD}$ is not closed under negation. We start with a positive result.

Theorem 33. HomClosed$_{(\mathrm{fin})}$ *for $\mathbb{TGD}$ is* NP-*complete.*

Intuitively, this result is facilitated by the fact that homclosed $\mathbb{TGD}$ sentences must have very small universal models, against which homclosedness violations can be checked on a per-rule basis. However, all is lost as soon as just a tad of disjunction is introduced.

Definition 34. *Given a deterministic domino system* $\mathcal{D} = (D, B, L, H, V)$, *we define the* $\mathbb{TGD}$ *sentence*

$$\Phi_{\mathcal{D}\text{-tiling}} = \Phi_{\text{grid}} \wedge \Phi_{\mathcal{D}}, \tag{19}$$

where $\Phi_{\mathcal{D}}$ *is the conjunction over the following rules (with unary predicates* $\mathtt{T}_{\emptyset}$ *and* $\mathtt{T}_{\{d\}}$ *for all the dominoes* $d \in D_{\mathcal{D}}$):

$$\exists v. \mathtt{T}_{B \cap L}(v) \tag{20}$$

$$\forall xy. \big(\mathtt{T}_{\{d\}}(x) \wedge \mathtt{V}(x,y) \Rightarrow \mathtt{T}_{V(d)}(y)\big) \qquad d \in L \tag{21}$$

$$\forall xy. \big(\mathtt{T}_{\{d\}}(x) \wedge \mathtt{H}(x,y) \Rightarrow \mathtt{T}_{H(d)}(y)\big) \qquad d \in B \tag{22}$$

$$\forall xyz. \big(\mathtt{T}_{\{d\}}(x) \wedge \mathtt{H}(x,z) \wedge \mathtt{T}_{\{d'\}}(y) \wedge \mathtt{V}(y,z)$$
$$\Rightarrow \mathtt{T}_{H(d) \cap V(d')}(z)\big) \quad d,d' \in D \tag{23}$$

$$\forall x. \big(\mathtt{T}_{\{d\}}(x) \wedge \mathtt{T}_{\{d'\}}(x) \Rightarrow \mathtt{T}_{\emptyset}(x)\big) \qquad d,d' \in D, d \neq d' \tag{24}$$

Note that a(n ultimately periodic) $\mathcal{D}$-tiling exists exactly if $\Phi_{\mathcal{D}\text{-tiling}}$ has a (finite) model $\mathfrak{A}$ with $\mathtt{T}_{\emptyset}^{\mathfrak{A}} = \emptyset$. This is the case exactly if $[\![\exists x. \mathtt{T}_{\emptyset}(x)]\!]_{(<\omega)} \neq [\![\Phi_{\mathcal{D}\text{-tiling}} \vee \exists x. \mathtt{T}_{\emptyset}(x)]\!]_{(<\omega)}$. Yet, note that any – finite or infinite – model $\mathfrak{A}$ of $\Phi_{\mathcal{D}\text{-tiling}}$ with $\mathtt{T}_{\emptyset}^{\mathfrak{A}} = \emptyset$ admits a homomorphism into $\mathfrak{A} \uplus \mathfrak{I}$, for which $\mathtt{T}_{\emptyset}^{\mathfrak{A} \uplus \mathfrak{I}} = \emptyset$ and it is not a model of $\Phi_{\mathcal{D}\text{-tiling}}$, thus $\mathfrak{A} \uplus \mathfrak{I} \notin [\![\Phi_{\mathcal{D}\text{-tiling}} \vee \exists x. \mathtt{T}_{\emptyset}(x)]\!]_{(<\omega)}$. Therefore, the class of (finite) models of the $\mathbb{MDTGD}$ sentence $\Psi = \Phi_{\mathcal{D}\text{-tiling}} \vee \exists x. \mathtt{T}_{\emptyset}(x)$ is closed under (finite-target) homomorphisms exactly if there exists no (ultimately periodic) $\mathcal{D}$-tiling. We get the following.

Theorem 35. $\text{HomClosed}_{(\text{fin})}$ *is undecidable for* $\mathbb{MDTGD}$ *(and hence for* $\mathbb{DTGD}$). *This even holds for* $\mathbb{DTGD}$ *sentences with* ≤ 4 *universally and* ≤ 1 *existentially quantified variables and predicates of arity* ≤ 2.

VI. CHARACTERIZABILITY OF HOMCLOSURES

In case a sentence Φ under consideration is not homclosed, one might still be interested in the (finite) homclosure $[\![\Phi]\!]_{(<\omega)}^{\rightarrow(<\omega)}$ of its (finite) models. In particular one might want to (finitely) characterize it logically, that is, find some sentence Ψ whose set of (finite) models is precisely Φ's (finite) homclosure.

We start with the observation that even for very basic $\mathbb{FO}_=$ sentences contained in very restricted fragments, it might be impossible to characterize their homclosure within $\mathbb{FO}_=$.

Definition 36. *Let* Φ_{∞} *denote the* $\{\mathtt{P}\}$-sentence $\forall x \exists y. \mathtt{P}(x,y)$.

We note that Φ_{∞} is simultaneously contained in $\mathbb{GFO}_=$, $\mathbb{FO}^2$, $\forall \exists \mathbb{FO}$ and $\mathbb{TGD}$. $[\![\Phi_{\infty}]\!]^{\rightarrow}$ consists of all structures containing an infinite P-path. In the finite, this condition is equivalent to containing a directed P-cycle. Identifying such structures is NLOGSPACE-complete, hence not expressible in $\mathbb{FO}_=$ [10].

Proposition 37. $[\![\Phi_{\infty}]\!]^{\rightarrow}$ *cannot be characterized within* $\mathbb{FO}_=$.

Given that homclosures are not guaranteed to be characterizable in $\mathbb{FO}_=$ even for very inexpressive $\mathbb{FO}_=$ fragments, we might ask whether one can at least tell sentences that have a $\mathbb{FO}_=$-characterizable homclosure from those that do not.

> **Problem:** $\text{ExCharHomCl}_{(<\omega)}^{\mathbb{L}}$
> **Input:** τ, τ-sentence Φ.
> **Output:** YES, if $[\![\Psi]\!]_{(<\omega)} = [\![\Phi]\!]_{(<\omega)}^{\rightarrow(<\omega)}$ for some $\mathbb{L}$-sentence Ψ, NO otherwise.

It turns out the answer is no. Also, there is no stronger logic capable of characterizing the homclosures of all $\mathbb{FO}_=$ sentences, nor one that would facilitate $\text{ExCharHomCl}_{(<\omega)}$.

Theorem 38. $\text{ExCharHomCl}_{(<\omega)}^{\mathbb{L}}$ *is undecidable for* $\mathbb{FO}_=$ *sentences whenever* $\mathbb{L}$ *is a non-degenerate logic with a decidable model checking problem.*

Theorem 39. $\text{ExCharHomCl}_{(<\omega)}^{\mathbb{L}}$ *is undecidable for* $\mathbb{TGD}$ *sentences whenever* $\mathbb{L}$ *is a logic able to express* $\text{cq}(\mathfrak{F})$ *and having a decidable model checking problem.*

After establishing these negative results, we turn to the other end of the spectrum to identify cases where the homomorphism closure is guaranteed to be characterizable in $\mathbb{FO}_=^2$ and thus – as a consequence of the homomorphism preservation theorems – even in $\exists^* \mathbb{FO}_=^{\pm}$. For $\exists^* \forall^* \mathbb{FO}_=$, the Bernays-Schönfinkel class, this is an easy consequence of the existence of "small" induced submodels.

Proposition 40. *Every* $\exists^* \forall^* \mathbb{FO}_=$ *sentence* Φ *satisfies* $[\![\Phi]\!]^{\rightarrow} = [\![\Phi]\!]_{<\omega}^{\rightarrow}$ *and one can effectively compute a* $\exists^* \mathbb{FO}_=^{\pm}$ *sentence* Ψ *with* $[\![\Phi]\!]^{\rightarrow} = [\![\Psi]\!]$.

Using descriptive complexity theory [17], this implies a very low data complexity of checking homclosure membership.

Corollary 41. *For any fixed* $\exists^* \forall^* \mathbb{FO}_=$ *sentence* Φ, *checking membership in* $[\![\Phi]\!]_{(<\omega)}^{\rightarrow}$ *is in* AC^0 *in the size of the structure.*

Now, after establishing the extremal cases (of characterizability failing altogether vs. it succeeding already within $\exists^* \mathbb{FO}_=^{\pm}$) we investigate the middle ground in between: the fragments $\exists^* \forall \forall \exists^* \mathbb{FO}$, $\mathbb{FO}_=^2$, $\mathbb{TGF}$, $\mathbb{GFO}_=$, and $\mathbb{GNFO}_=$. By Theorem 7, we know that $[\![\Phi]\!]^{\rightarrow<\omega} = [\![\Phi]\!]_{<\omega}^{\rightarrow<\omega}$ for every sentence Φ in any of these logics. Therefore, any sentence characterizing $[\![\Phi]\!]^{\rightarrow}$ will also finitely characterize $[\![\Phi]\!]_{<\omega}^{\rightarrow<\omega}$. Hence, we will next describe how to characterize $[\![\Phi]\!]^{\rightarrow}$ for the considered fragments. The task of characterizing $[\![\Psi]\!]_{<\omega}^{\rightarrow}$ is slightly different and will be discussed thereafter.

It is helpful to study projective homclosure characterizations.

Definition 42. *For logics* $\mathbb{L}, \mathbb{L}' \subseteq \mathbb{SO}$, *we say* $\mathbb{L}'$ *homcaptures* $\mathbb{L}$ *if for every* τ *and* $\mathbb{L}$ τ-sentence Φ *there exist some* σ *and a* $\mathbb{L}'$ $\tau \uplus \sigma$-sentence Ψ *such that* $[\![\Phi]\!]^{\rightarrow} = [\![\Psi]\!]|_{\tau}$.

Note that whenever a logic $\mathbb{L}'$ homcaptures a logic $\mathbb{L}$, it follows that for any $\mathbb{L}$ sentence Φ, we find an $\exists \mathbb{SO}(\mathbb{L}')$ sentence characterizing Φ's homclosure, where $\exists \mathbb{SO}(\mathbb{L}')$ describes the fragment of $\mathbb{SO}$ sentences led by second-order existential quantifiers followed by an $\mathbb{L}'$ sentence.

We will next give a description of a generic type-based approach toward homcapturing results, which will be applicable to all mentioned logics with slight variations in the details.

Fixing a finite signature τ, we let n-*types* denote maximal consistent sets t of literals (i.e., negated and unnegated atoms) using all constants from τ and a finite collection of variables $v_1, \ldots, v_n$ (denoted $\boldsymbol{v}$) disjoint from the variables used in the considered sentence. For convenience, we identify the (necessarily finite) set t with $\bigwedge \mathsf{t}$. We say an n-type t is *realized* by an n-tuple $\boldsymbol{a} = (a_1, \ldots, a_n) \in A^n$ of domain elements of a τ-structure $\mathfrak{A}$, if $(\mathfrak{A}, \{v_i \mapsto a_i\}) \models \mathsf{t}$. We say t is *realized* (k times) in $\mathfrak{A}$ if it is realized by (k distinct) tuples of $\mathfrak{A}$. Note that every tuple $\boldsymbol{a}$ realizes one unique type, which entirely characterizes the substructure of $\mathfrak{A}$ induced by all the elements of $\boldsymbol{a}$ together with all elements denoted by constants. Our approach makes use of the fact that for several logics, many questions regarding a structure can be answered, if one knows what types of a certain kind are realized therein. On a generic level, we assume there is a function that, given a τ-sentence Φ yields a finite set $\mathcal{E}$ of *eligible types* (possibly containing n-types of varying n), which are sufficient for our purposes. Then, given a τ-structure $\mathfrak{A}$, we let its *type summary* $\mathcal{E}(\mathfrak{A})$ denote the pair $(\mathcal{E}_+, \mathcal{E}_!)$ where $\mathcal{E}_+$ ($\mathcal{E}_!$) consists of those types from $\mathcal{E}$ realized (exactly once) in $\mathfrak{A}$.[7] Furthermore, the *model summary* $(\!|\Phi|\!)$ of a sentence Φ is defined by $\{\mathcal{E}(\mathfrak{A}) \mid \mathfrak{A} \in [\![\Phi]\!]\}$, which is a finite set by assumption.

We will extend our signature by a set σ of *type predicates*, containing an n-ary predicate symbol $\mathtt{Tp}_\mathsf{t}$ for every eligible n-type $\mathsf{t} \in \mathcal{E}$. Given a τ-structure $\mathfrak{A}$ and a set $\mathcal{E}$ of eligible types over τ, the $\mathcal{E}$-*adornment* of $\mathfrak{A}$, denoted $\mathfrak{A} \cdot \mathcal{E}$ is the $\tau \uplus \sigma$-structure expanding $\mathfrak{A}$ such that $\mathtt{Tp}_\mathsf{t}^{\mathfrak{A} \cdot \mathcal{E}} = \{\boldsymbol{a} \mid \boldsymbol{a} \text{ realizes } \mathsf{t} \text{ in } \mathfrak{A}\}$. In a way, $\mathcal{E}$-adornments materialize the information regarding relevant types in a model. While this information is obviously redundant in the model itself, we will make good use of it when characterizing a model's homomorphic targets.

Given a τ-sentence Φ (which we presume to be in some normal form with bounded quantifier alternation), we will construct a $\tau \uplus \sigma$-sentence Ψ that (projectively) identifies those $\mathfrak{B}$ into which some model $\mathfrak{A}$ of Φ has a homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$, using the following idea: assuming such an h exists, then it will also serve as homomorphism from $\mathfrak{A} \cdot \mathcal{E}$ to $\mathfrak{B} \cdot h(\mathfrak{A} \cdot \mathcal{E}|_\sigma)$ denoting the structure $\mathfrak{B}'$ expanding $\mathfrak{B}$ by letting $\mathtt{Tp}_\mathsf{t}^{\mathfrak{B}'} = \{(h(a_1), \ldots, h(a_n)) \mid (a_1, \ldots, a_n) \in \mathtt{Tp}_\mathsf{t}^{\mathfrak{A} \cdot \mathcal{E}}\}$, that is, where all type information materialized in $\mathfrak{A}$ is transferred to $\mathfrak{B}$ along h. In fact, Ψ is then built in a way that its models are precisely such $\mathfrak{B}'$. Thereby, one integral part of Ψ is a conjunct Ψ_{hom} ensuring that types are mapped to $\mathfrak{B}$ in a way compatible with the homomorphism condition, i.e., that at least all positive literals postulated by the type are indeed present:

$$\Psi_{\text{hom}} = \bigwedge_{\mathsf{t} \in \mathcal{E}} \forall \boldsymbol{v}.\big(\mathtt{Tp}_\mathsf{t}(\boldsymbol{v}) \Rightarrow \bigwedge \{\alpha \in \mathsf{t} \mid \alpha \text{ positive lit.}\}\big) \quad (25)$$

[7] In fact, $\mathcal{E}_!$ is only needed for the $\mathbb{FO}_{\overline{=}}^2$ case in order to deal with the phenomenon of so-called "kings", i.e., 1-types forced to be unique in models. For the other logics considered, $\mathcal{E}_+$ suffices.

Moreover, we require that the collection of "transferred type atoms" found in $\mathfrak{B}'$ corresponds to the type summary of some model. To this end, we include in Ψ the case distinction

$$\bigvee_{(\mathcal{E}_+, \mathcal{E}_!) \in (\!|\Phi|\!)} \Psi_{(\mathcal{E}_+, \mathcal{E}_!)}, \quad (26)$$

where each $\Psi_{(\mathcal{E}_+, \mathcal{E}_!)}$ details existence and relationships between the mapped types. Some of the conditions specified in $\Psi_{(\mathcal{E}_+, \mathcal{E}_!)}$ are generically derived from $(\mathcal{E}_+, \mathcal{E}_!)$, others are obtained by translating constraints from Φ into "type language", expressed by type atoms. By construction, the sentence Ψ thus obtained holds in some type-adorned structure $\mathfrak{A} \cdot \mathcal{E}$ if and only if $\mathfrak{A} \in [\![\Phi]\!]$. Also, its satisfaction is preserved under "type-faithful" homomorphisms $h \colon \mathfrak{A} \cdot \mathcal{E} \to \mathfrak{B}'$ as described above, thus ensuring $[\![\Phi]\!]^{\to} \subseteq [\![\Psi]\!]|_\tau$. The other direction $[\![\Phi]\!]^{\to} \supseteq [\![\Psi]\!]|_\tau$ is shown by taking a model $\mathfrak{B}' \in [\![\Psi]\!]$, removing all information about τ, and "unfolding" or "unraveling" it such that, in the resulting structure $\mathfrak{A}'$, the diverse $\mathtt{Tp}_\mathsf{t}$ instances do not provide contradicting information about the τ-connections between their arguments. This allows for reconstructing the interpretations of the original signature elements τ from the $\mathtt{Tp}_\mathsf{t}$ instances. After forgetting the latter, we have arrived at our wanted $\mathfrak{A} \in [\![\Phi]\!]$.

Applying the described strategy yields the following theorem.

Theorem 43. *1)* $\mathbb{FO}_{\overline{=}}^2$ *homcaptures itself.*

2) $\mathbb{TGF}$ *homcaptures itself and* $\exists^* \forall \forall \exists^* \mathbb{FO}$.

3) $\mathbb{GFO}_=$ *homcaptures itself and* $\mathbb{GNFO}_=$.

As an example for all these logics, note that the homclosure of Φ_∞ from Definition 36, is projectively characterized by

$$\exists x.\big(\mathtt{Lv}(x)\big) \wedge \forall x.\big(\mathtt{Lv}(x) \Rightarrow \exists y.(\mathtt{P}(x,y) \wedge \mathtt{Lv}(y))\big). \quad (27)$$

In view of our earlier discussion, this ensures that all these logics admit homclosure characterization in $\exists \mathbb{SO}$, and that the data complexity of homclosure membership is in NP [9].

Corollary 44. *For any fixed* $\mathbb{FO}_{\overline{=}}^2$, $\mathbb{TGF}$, $\exists^* \forall \forall \exists^* \mathbb{FO}$, $\mathbb{GFO}_=$ *or* $\mathbb{GNFO}_=$ *sentence* Φ, *checking* $\mathfrak{A} \in [\![\Phi]\!]^{\to}$ *is in* NP.

We next complement these NP-membership results by providing matching lower bounds even for much weaker logics.

Definition 45. *Given an instance of 3SAT* $\mathcal{S} = \{C_1, \ldots, C_n\}$ *consisting of* n *clauses each containing 3 literals over propositional variables* $p_1, \ldots, p_m$, *we define the structure* $\mathfrak{A}_\mathcal{S}$ *by*

$$A_\mathcal{S} = \{a_i \mid 1 \le i \le n+1\} \cup \{b_i, b_i' \mid 1 \le i \le m\} \quad (28)$$
$$\mathtt{First}^{\mathfrak{A}_\mathcal{S}} = \{a_1\} \quad (29)$$
$$\mathtt{CLst}^{\mathfrak{A}_\mathcal{S}} = \{a_i \mid 1 \le i \le n\} \quad (30)$$
$$\mathtt{Nil}^{\mathfrak{A}_\mathcal{S}} = \{a_{n+1}\} \quad (31)$$
$$\mathtt{Next}^{\mathfrak{A}_\mathcal{S}} = \{(a_i, a_{i+1}) \mid 1 \le i \le n\} \quad (32)$$
$$\mathtt{Lit}^{\mathfrak{A}_\mathcal{S}} = \{(a_i, b_j) \mid p_j \in C_i\} \cup \{(a_i, b_j') \mid \neg p_j \in C_i\} \quad (33)$$
$$\mathtt{Sel}^{\mathfrak{A}_\mathcal{S}} = \{b_i, b_i' \mid 1 \le i \le m\} \quad (34)$$
$$\mathtt{Cmp}^{\mathfrak{A}_\mathcal{S}} = \{(b_i, b_j), (b_i', b_j'), (b_i, b_k'), (b_i', b_k) \mid i \ne k\} \quad (35)$$

We let $\Phi_{3\mathrm{SAT}}$ be the sentence containing the conjunction of the following sentences (leading universal quantifier omitted):

$$\exists x.\mathrm{First}(x) \tag{36}$$

$$\mathrm{First}(x) \Rightarrow \mathrm{CLst}(x) \tag{37}$$

$$\mathrm{CLst}(x) \Rightarrow \exists y.\mathrm{Next}(x,y) \wedge \big(\mathrm{CLst}(y) \vee \mathrm{Nil}(y)\big) \tag{38}$$

$$\mathrm{CLst}(x) \Rightarrow \exists y.\mathrm{Lit}(x,y) \wedge \mathrm{Sel}(y) \tag{39}$$

$$\mathrm{Sel}(x) \wedge \mathrm{Sel}(y) \Rightarrow \mathrm{Cmp}(x,y). \tag{40}$$

Obviously, $\Phi_{3\mathrm{SAT}}$ is contained in constant-free, equality-free $\mathbb{FO}^2$ (and therefore in $\mathbb{FO}^2_=$ and $\mathbb{TGF}$). It can be rewritten into constant-free, equality-free $\forall\forall\exists\exists\mathbb{FO}$ (and therefore into $\exists^*\forall\forall\exists^*\mathbb{FO}$). Moreover, for any set $\mathcal{S}$ of 3-clauses, satisfiability of the 3SAT instance $\bigwedge_{\{\ell_1,\ell_2,\ell_3\}\in\mathcal{S}} \ell_1 \vee \ell_2 \vee \ell_3$ coincides with $\mathfrak{A}_{\mathcal{S}} \in [\![\Phi_{3\mathrm{SAT}}]\!]^{\rightarrow}_{(<\omega)}$. Hence we obtain the following statement, providing matching lower bounds.

Proposition 46. *Checking homclosure membership is* NP-*hard for constant-free, equality-free $\mathbb{FO}^2$ and $\forall\forall\exists\exists\mathbb{FO}$ with predicate arity ≤ 2.*

Note that this settles the case of the data complexity of homclosure membership for any logic between $\mathbb{FO}^2$ / $\forall\forall\exists\exists\mathbb{FO}$ on the lower and $\mathbb{FO}^2_=$ / $\mathbb{TGF}$ / $\exists^*\forall\forall\exists^*\mathbb{FO}$ on the upper end, showing that one cannot hope for characterizability in a logic having a lower data complexity of model checking. It does not provide a lower bound for $\mathbb{GFO}_=$ or $\mathbb{GNFO}_=$, though. As it turns out, we can in fact do better for these...

Given a $\mathbb{GNFO}_=$ sentence Φ, let Ψ be the $\mathbb{GFO}_=$ sentence projectively characterizing $[\![\Phi]\!]^{\rightarrow}$ constructed as described before. Then we obtain $[\![\Phi]\!]^{\rightarrow} = [\![\exists\sigma.\Psi]\!]$ and also observe

$$\exists\sigma.\Psi = \exists\sigma.\Big(\Psi_{\mathrm{hom}} \wedge \bigvee_{(\mathcal{E}_+,\mathcal{E}_!)\in(\![\Phi]\!)} \Psi_{(\mathcal{E}_+,\mathcal{E}_!)}\Big) \tag{41}$$

$$\equiv \bigvee_{(\mathcal{E}_+,\mathcal{E}_!)\in(\![\Phi]\!)} \exists\sigma.\big(\Psi_{(\mathcal{E}_+,\mathcal{E}_!)} \wedge \Psi_{\mathrm{hom}}\big). \tag{42}$$

Thoroughly inspecting $\exists\sigma.(\Psi_{(\mathcal{E}_+,\mathcal{E}_!)} \wedge \Psi_{\mathrm{hom}})$, we find that it can be equivalently rewritten into the following $\mathbb{FO}_=$ sentence with least fixed points over simultaneous inductive definitions for "complement predicates" $\overline{\mathrm{Tp}}_t$ from a signature $\overline{\sigma}$:

$$\bigwedge_{t\in\mathcal{E}_+} \exists\boldsymbol{x}.\neg\Big[\mathbf{lfp}_{\overline{\mathrm{Tp}}_t}\big\{\overline{\mathrm{Tp}}_{t'}(\boldsymbol{y}) \leftarrow \xi_{\overline{\mathrm{Tp}}_{t'}}[\boldsymbol{y}] \mid t'\in\mathcal{E}\big\}\Big](\boldsymbol{x}), \tag{43}$$

where the $\xi_{\overline{\mathrm{Tp}}_{t'}}$ are $\mathbb{FO}_= \ \tau\uplus\overline{\sigma}$-formulae with only positive occurrences of type atoms, obtained from aggregating the bodies of the contrapositions of the implications in $\Psi_{(\mathcal{E}_+,\mathcal{E}_!)}$.[8] Simultaneous induction can in turn be expressed using plain least-fixed-point logic [14], therefore Ψ as a whole can be

---

[8] Conceptually, this equivalent transformation can be explained as follows: for any two models $\mathfrak{B}_1, \mathfrak{B}_2$ of $\Psi_{(\mathcal{E}_+,\mathcal{E}_!)}$ with $\mathfrak{B}_1|_\tau = \mathfrak{B}_2|_\tau$, the structure $\mathfrak{B}_1\cdot\mathfrak{B}_2$ obtained as the σ-expansion of $\mathfrak{B}_1|_\tau$ by $\mathrm{Tp}_t^{\mathfrak{B}_1\cdot\mathfrak{B}_2} = \mathrm{Tp}_t^{\mathfrak{B}_1}\cup\mathrm{Tp}_t^{\mathfrak{B}_2}$ for all eligible t is also a model. Hence, for every $(\mathcal{E}_+,\mathcal{E}_!)$-compatible τ-structure $\mathfrak{B}$, there is a "maximal σ-expansion" $\mathfrak{B}^*$ satisfying $\Psi_{(\mathcal{E}_+,\mathcal{E}_!)}$, which can be deterministically obtained in a type-elimination like process, via a least fixed-point computation over predicates $\overline{\mathrm{Tp}}_t$ that express exactly the absence of Tp_t. This spares us the "guessing" of type relations via $\exists\sigma$.

expressed in first-order least-fixed-point logic denoted $\mathbb{FO}^{\mathbf{lfp}}_=$ (sometimes also referred to as LFP or FO+LFP). Using the fact that evaluating $\mathbb{FO}^{\mathbf{lfp}}_=$ sentences over structures can be done in polynomial time [16, 30], we obtain an improved result regarding the data complexity of homclosure membership.

Theorem 47. *Homclosures of $\mathbb{GNFO}_=$ and $\mathbb{GFO}_=$ sentences can be characterized in $\mathbb{FO}^{\mathbf{lfp}}_=$. Thus, for a fixed $\mathbb{GNFO}_=$ or $\mathbb{GFO}_=$ sentence Φ, checking $\mathfrak{A}\in[\![\Phi]\!]^{\rightarrow}$ is* P-*complete.*

As a minimalistic example, consider the $\mathbb{GFO}_=$ sentence Φ_∞ from Definition 36, whose homclosure cannot be characterized in $\mathbb{FO}_=$ (see Proposition 37), but by the $\mathbb{FO}^{\mathbf{lfp}}_=$ sentence

$$\exists x.\neg\Big[\mathbf{lfp}_{\mathrm{Cds}}\big\{\mathrm{Cds}(y) \leftarrow \forall z.\big(\mathrm{P}(y,z) \Rightarrow \mathrm{Cds}(z)\big)\big\}\Big](x), \tag{44}$$

where the fixed point predicate Cds identifies "cul-de-sac" elements in the given $\{\mathrm{P}\}$-structure.

We finally turn to the case of finding a characterization of $[\![\Phi]\!]^{\rightarrow}_{<\omega}$, i.e., of those – finite as well as infinite – structures receiving a homomorphism from a *finite* model of Φ. We start with a negative result for a very simple sentence, following from compactness of $\mathbb{FO}_=$.

Proposition 48. $[\![\Phi_\infty]\!]^{\rightarrow}_{<\omega}$ *cannot be characterized in* $\exists\mathbb{SO}$.

However, it is easy to see that an immaterial extension of $\exists\mathbb{SO}$ suffices to remedy this situation: let $\exists\mathbb{SO}^{\#}$ denote $\exists\mathbb{SO}$ extended by a special second-order quantifier $\exists^{<\omega}$ which requires that the quantified-over predicate corresponds to a finite relation – an idea borrowed from weak monadic second-order logic. We note that $\exists^{<\omega}\mathrm{U}$ is also expressible in plain $\mathbb{SO}$.

Theorem 49. *Let Φ be some $\mathbb{SO}$ sentence and let Ψ be such that $[\![\Psi]\!]_{<\omega} = [\![\Phi]\!]^{\rightarrow,<\omega}_{<\omega}$. Then $[\![\Phi]\!]^{\rightarrow}_{<\omega} = [\![\exists^{<\omega}\mathrm{U}.\Psi^{\mathrm{rel}(\mathrm{U})}]\!]$.*

Together with Theorem 7 and the previously established characterization results, this allows us to close this case.

Corollary 50. *For all Φ in any of $\exists^*\forall\forall\exists^*\mathbb{FO}$, $\mathbb{TGF}$, $\mathbb{FO}^2_=$, and $\mathbb{GNFO}_=$, $[\![\Phi]\!]^{\rightarrow}_{<\omega}$ can be characterized in $\exists\mathbb{SO}^{\#}$.*

VII. NORMAL FORMS FOR HOMCLOSED FRAGMENTS

This section is devoted to identifying "homclosed normal forms" of certain logics. Given a logic $\mathbb{L}$, a homclosed normal form $\mathbb{HL}$ of $\mathbb{L}$ is a fragment which satisfies

- $\mathbb{HL} \subseteq \mathbb{L}$,
- every $\Phi \in \mathbb{HL}$ is homclosed,
- for each homclosed $\Phi \in \mathbb{L}$ exists a $\Phi' \in \mathbb{HL}$ with $\Phi \equiv \Phi'$,
- membership in $\mathbb{HL}$ is decidable.

It is, of course, desirable to strengthen the last requirement to say that membership can be decided "easily", e.g., in polytime. Likewise, we might require the existence of a (preferably computationally inexpensive) algorithm to compute Φ' from a given Φ, or of a (not too high) upper bound regarding the size of Φ' with respect to Φ.

Normal Forms for $\mathbb{FO}_=$ and Fragments. The homomorphism preservation theorem provides a normal form for full $\mathbb{FO}_=$, namely $\exists^*\mathbb{FO}_=^+$; the equivalent $\exists^*\mathbb{FO}_=^+$ sentences may even be chosen to have equal quantifier rank. Rossman's theorem shows that the same normal form can be used if we consider finite models only; more recently, Rossman even established that any homclosed first-order formula of quantifier rank k is equivalent to an existential positive formula of quantifier rank $k^{O(1)}$ [23]. Note that computing the normal form can be quite costly: there is a potentially non-elementary blow-up for the length of a shortest equivalent existential positive formula [22, Theorem 6.1].

Obviously, this characterization of a normal form for $\mathbb{FO}_=$ carries over to any $\mathbb{FO}_=$ fragment that fully encompasses $\exists^*\mathbb{FO}_=^+$. For Bernays-Schönfinkel fragments with bounded number of existential quantifiers, we state the following observation.

Proposition 51. *For every homclosed $\exists^n\forall^*\mathbb{FO}_=$ sentence there exists an equivalent $\exists^n\mathbb{FO}_=^+$ sentence.*

For other fragments, the case is less immediate. Whenever HomClosed is decidable for $\mathbb{L}$, we can use this (usually expensive) test to directly define $\mathbb{HL}$, thus obtaining $\mathbb{HTGF}$, $\mathbb{HFO}_=^2$, and $\mathbb{H}\forall\forall\exists\exists\mathbb{FO}$. For $\mathbb{FO}_=$ fragments not covered by these cases (such as $\forall\forall\exists\mathbb{FO}$), the question remains open.

Normal Form for Homclosed $\mathbb{SO}$. We next provide a syntactic normal form for $\mathbb{SO}$ sentences whose class of models is homclosed. This normal form arises as a combination of a normal form for superstructure-closed classes and a normal form for classes closed under surjective homomorphisms. Our proof relies on the interplay of syntactic manipulations of $\mathbb{SO}$ sentences on the one hand and "operators" on classes of structures on the other. We focus on the following two operations that can be applied to classes $\mathcal{C}$ of τ-structures:

$$\mathcal{C}^{\hookrightarrow_{(<\omega)}} = \{\mathfrak{A} \text{ (finite)} \mid \mathfrak{B} \hookrightarrow \mathfrak{A} \text{ for some } \mathfrak{B} \in \mathcal{C}\}, \quad (45)$$
$$\mathcal{C}^{\mathrm{sH}_{(<\omega)}} = \{\mathfrak{A} \text{ (finite)} \mid \{\mathfrak{A}\}^{\twoheadrightarrow} \subseteq \mathcal{C}\}. \quad (46)$$

With these notions, we can establish the following characterizations of classes of τ-structures.

Lemma 52. *For any class $\mathcal{C}$ of (finite) τ-structures holds:*

1) *$\mathcal{C}^{\hookrightarrow_{(<\omega)}}$ is closed under (finite) superstructures. Moreover, if $\mathcal{C}$ is closed under (finite) superstructures, then $\mathcal{C}^{\hookrightarrow_{(<\omega)}} = \mathcal{C}$.*

2) *$\mathcal{C}^{\mathrm{sH}_{(<\omega)}}$ is closed under (finite) surjective homomorphisms. Moreover, if $\mathcal{C}$ is closed under (finite) surjective homomorphisms, then $\mathcal{C}^{\mathrm{sH}_{(<\omega)}} = \mathcal{C}$.*

3) *$\left(\mathcal{C}^{\mathrm{sH}_{(<\omega)}}\right)^{\hookrightarrow_{(<\omega)}}$ is (finitely) homclosed. Moreover, every (finitely) homclosed $\mathcal{C}$ satisfies $\left(\mathcal{C}^{\mathrm{sH}_{(<\omega)}}\right)^{\hookrightarrow_{(<\omega)}} = \mathcal{C}$.*

Based on this, we can introduce a normal form for superstructure-closed $\mathbb{SO}$.

Definition 53. *Let τ be a signature and let Ψ be an $\mathbb{SO}$ τ-sentence. Then the $\mathbb{SO}$ τ-sentence Ψ^{sup} is defined as $\exists\dot{\mathrm{U}}.\Psi^{\mathrm{rel}(\dot{\mathrm{U}})}$, where $\dot{\mathrm{U}}$ is a fresh unary predicate.*

Proposition 54. *Let Ψ be an $\mathbb{SO}$ τ-sentence. Then*

$$\llbracket\Psi\rrbracket^{\hookrightarrow_{(<\omega)}}_{(<\omega)} = \llbracket\Psi^{\mathrm{sup}}\rrbracket_{(<\omega)}. \quad (47)$$

Proposition 54 and Lemma 52 give rise to the normal form for superstructure-closed $\mathbb{SO}$.

Corollary 55. *For any $\mathbb{SO}$ τ-sentence Ψ, $\llbracket\Psi^{\mathrm{sup}}\rrbracket_{(<\omega)}$ is closed under (finite) superstructures. Moreover, if $\llbracket\Psi\rrbracket_{(<\omega)}$ is closed under (finite) superstructures, then $\llbracket\Psi^{\mathrm{sup}}\rrbracket_{(<\omega)} = \llbracket\Psi\rrbracket_{(<\omega)}$.*

We proceed by establishing a normal form for $\mathbb{SO}$ sentences closed under surjective homomorphisms. For a signature τ and variables z, z', we let $\eta^\tau(z, z')$ denote the formula

$$\bigwedge_{\substack{\mathrm{P}\in\tau \\ k=\mathrm{ar}(\mathrm{P})}} \bigwedge_{\substack{1\leq i\leq k \\ \boldsymbol{x}=(x_1,\ldots,x_{i-1}) \\ \boldsymbol{y}=(x_{i+1},\ldots,x_k)}} \forall\boldsymbol{xy}.\mathrm{P}(\boldsymbol{x}, z, \boldsymbol{y}) \Leftrightarrow \mathrm{P}(\boldsymbol{x}, z', \boldsymbol{y}), \quad (48)$$

and, for a unary $\mathrm{U} \notin \tau$, we define the $\tau\uplus\{\mathrm{U}\}$-sentence Θ_{U}^τ by

$$\forall x\exists y.(\mathrm{U}(y) \wedge \eta^\tau(x, y)). \quad (49)$$

Definition 56. *Let τ and τ' be disjoint signatures such that τ' consists of copies P' of all predicates $\mathrm{P} \in \tau$. Let Ψ be an $\mathbb{SO}$ τ-sentence. Then the $\mathbb{SO}$ τ-sentence Ψ^{shom} is defined by*

$$\forall\mathrm{U}\forall\tau'.\left(\left(\Theta_{\mathrm{U}}^\tau \wedge \bigwedge_{\mathrm{P}'\in\tau'} \forall\boldsymbol{x}.\big(\mathrm{P}(\boldsymbol{x}) \Rightarrow \mathrm{P}'(\boldsymbol{x})\big)\right) \Rightarrow \Psi_{\tau\mapsto\tau'}^{\mathrm{rel}(\mathrm{U})}\right), \quad (50)$$

where U is a fresh unary predicate and $\Psi_{\tau\mapsto\tau'}^{\mathrm{rel}(\mathrm{U})}$ denotes $\Psi^{\mathrm{rel}(\mathrm{U})}$ with every $\mathrm{P} \in \tau$ replaced by its copy $\mathrm{P}' \in \tau'$.

Proposition 57. *Let Ψ be an $\mathbb{SO}$ τ-sentence. Then*

$$\llbracket\Psi\rrbracket^{\mathrm{sH}_{(<\omega)}}_{(<\omega)} = \llbracket\Psi^{\mathrm{shom}}\rrbracket_{(<\omega)}. \quad (51)$$

This result, together with Proposition 54 and Lemma 52, yields the normal form for surjective-homomorphism-closed $\mathbb{SO}$.

Corollary 58. *For any $\mathbb{SO}$ τ-sentence Ψ, $\llbracket\Psi^{\mathrm{shom}}\rrbracket_{(<\omega)}$ is closed under surjective (finite) homomorphisms. Moreover, if $\llbracket\Psi\rrbracket_{(<\omega)}$ is closed under surjective (finite) homomorphisms, then $\llbracket\Psi^{\mathrm{shom}}\rrbracket_{(<\omega)} = \llbracket\Psi\rrbracket_{(<\omega)}$.*

We are now ready to combine the established results toward the desired normal form for homclosed $\mathbb{SO}$. Invoking Corollary 55, Corollary 58, and Lemma 52, we obtain the following statement, motivating the subsequent definition.

Theorem 59. *An $\mathbb{SO}$-definable class of (finite) τ-structures is of the form*
$$\llbracket(\Psi^{\mathrm{shom}})^{\mathrm{sup}}\rrbracket_{(<\omega)} \quad (52)$$
for an $\mathbb{SO}$ τ-sentence Ψ if and only if it is closed under (finite) homomorphisms.

Definition 60. *We define $\mathbb{HSO}$ as the set of $\mathbb{SO}$ sentences of the form $(\Psi^{\mathrm{shom}})^{\mathrm{sup}}$ for any $\mathbb{SO}$ sentence Ψ.*

It is clear that membership in $\mathbb{HSO}$ is decidable. We can now formulate the main result of this section, as a direct consequence from Theorem 59 and the fact that $(\Psi^{\mathrm{shom}})^{\mathrm{sup}}$ is polytime-computable from Ψ.

TABLE I

OVERVIEW OF NEWLY ESTABLISHED RESULTS (5 LAST COLUMNS, EXCEPT NORMAL FORM FOR $\mathbb{FO}^2_{\underline{=}}$) AS WELL AS REFERENCES USED (EASY INSIGHTS ARE LEFT WITHOUT REFERENCE). ALL COMPLEXITIES ARE TIME COMPLEXITIES. CLOSURE OF $\mathbb{GNFO}_{\underline{=}}$ UNDER $\neg$ ONLY HOLDS FOR SENTENCES.

logic name	SAT fin/arb	finite model property (size)	closure $\neg$	closure $\wedge$	InHomCl comb.	InHomCl data	HomClosed fin/arb	homclosure characterizable in logic	normal form fragment
$\mathbb{FO}_{\underline{=}}$	und. [8, 29]	no	yes	yes	und.	und.	und.	none	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$
$\mathbb{DTGD}$	trivial	yes (1)	no	yes	und.	und.	und.	none	UCQ
$\mathbb{MDTGD}$	trivial	yes (1)	no	no	und.	und.	und.	none	CQ$\vee$CQ
$\mathbb{TGD}$	trivial	yes (1)	no	yes	und.	und.	NP	none	CQ
$\mathbb{TGF}$	N2Exp [25]	yes (2Exp) [20]	yes	yes	N2Exp	NP	coN2Exp	$\exists\mathbb{SO}(\mathbb{TGF})$	$\mathbb{HTGF}$
$\mathbb{FO}^2_{\underline{=}}$	NExp [12]	yes (Exp) [12]	yes	yes	NExp	NP	coNExp	$\exists\mathbb{SO}(\mathbb{FO}^2_{\underline{=}})$	$\mathbb{HFO}^2_{\underline{=}}$
$\mathbb{GNFO}_{\underline{=}}$	2Exp [1]	yes (2Exp) [1]	yes	yes	2Exp	P	2Exp	$\mathbb{FO}^{\mathbf{lfp}}$ / $\exists\mathbb{SO}(\mathbb{GFO}_{\underline{=}})$	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$
$\mathbb{GFO}_{\underline{=}}$	2Exp [11, 28]	yes (2Exp) [11, 2]	yes	yes	2Exp	P	2Exp	$\mathbb{FO}^{\mathbf{lfp}}_{\underline{=}}$ / $\exists\mathbb{SO}(\mathbb{GFO}_{\underline{=}})$	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$
$\forall\forall\forall\exists\mathbb{FO}$	und. [27]	no	no	no	und.	und.	und.	none	?
$\exists^*\forall\forall\exists^*\mathbb{FO}$	NExp [21]	yes (2Exp) [26]	no	yes	NExp	NP	und.	$\exists\mathbb{SO}(\mathbb{TGF})$	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$
$\forall\forall\exists\exists\mathbb{FO}$	NExp	yes (2Exp)	no	no	NExp	NP	coNExp	$\exists\mathbb{SO}(\mathbb{TGF})$	$\mathbb{H}\forall\forall\exists\exists\mathbb{FO}$
$\exists^*\forall^*\mathbb{FO}_{\underline{=}}$	NExp [21]	yes (C+Ex)	no	yes	NExp	AC^0	und.	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$
$\forall^*\mathbb{FO}_{\underline{=}}$	NExp	yes max(C,1)	no	yes	NExp	AC^0	coNExp	$\exists\mathbb{FO}^+$	$\varnothing\mathbb{FO}^{\pm}_{\underline{=}}$
$\exists\exists\exists\forall\mathbb{FO}$	NP	yes (C+3)	no	no	NP	AC^0	und.	$\exists\exists\exists\mathbb{FO}^+$	$\exists\exists\exists\mathbb{FO}^+$
$\exists\exists\forall\forall\mathbb{FO}$	NP	yes (C+2)	no	no	NP	AC^0	coNExp	$\exists\exists\mathbb{FO}^+$	$\exists\exists\mathbb{FO}^+$
$\exists^*\mathbb{FO}_{\underline{=}}$	NP	yes (C+Ex)	no	yes	NP	AC^0	coNExp	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$
$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$	const.	yes (C+Ex)	no	yes	NP	AC^0	trivial	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$	$\exists^*\mathbb{FO}^{\pm}_{\underline{=}}$
$\mathbb{SO}$	und.	no	yes	yes	und.	und.	und.	none	$\mathbb{HSO}$

Corollary 61. *Every (finitely) homclosed $\mathbb{SO}$ sentence Φ is equivalent (on finite structures) to a $\mathbb{HSO}$ sentence Ψ which can be computed from Φ in polynomial time.*

VIII. CONCLUSION

Inspired by the homomorphism preservation theorem and motivated by routinely encountering – clearly fundamental yet seemingly largely neglected – questions regarding homclosures of logically characterized model classes, we undertook a principled analysis of four basic questions related to that matter and clarified them for a wide range of theoretically and practically relevant logical formalisms, both in the finite-model setting and the arbitrary-model one. Table I summarizes the results obtained.

Next to several newly introduced generic techniques, which might prove useful well beyond the specific formalisms considered here, our most noteworthy achievements are probably the establishment of a computationally well-behaved normal form for homclosed $\mathbb{SO}$ sentences, and the insight that checking homclosure membership for the very expressive guarded negation fragment of first-order logic ($\mathbb{GNFO}_{\underline{=}}$) has only polynomial-time data complexity.

Plenty of open problems remain. Obviously other, different logical formalisms could be investigated regarding the considered questions. Not all of the established techniques immediately lend themselves to coping with "counting logics", such as two-variable logic with counting quantifiers. More generally, logics without the finite model property might turn out to be harder to handle.

The investigation into the prefix classes has not yet resulted in a complete characterization of decidability and complexity of the homclosedness problem. A first inspection seems to indicate that corresponding results would hinge on a more fine-grained analysis of fragments defined by the conjunction over sentences from different prefix classes.

As far as normal forms are concerned, for first-order prefix classes with bounded number of existential quantifiers, preceded by universal quantification, the question remains generically unsolved. For $\mathbb{TGF}$ and $\mathbb{FO}^2_{\underline{=}}$, normal forms have been shown to exist, but recognizing them is costly and more syntactic solutions would be preferred.

An interesting open problem is whether there exists a syntactic fragment of *existential second-order logic* such that the classes of finite structures that are expressible in the fragment are precisely the homomorphism-closed classes in NP. Similarly, we ask whether there exists a logic (in the sense of Gurevich [13]) that captures precisely the homclosed classes of finite structures that are in the complexity class P.

Another interesting avenue arises from the area of constraint satisfaction problems (CSPs). It seems advisable to systematically explore which (complements of) CSPs can be described as the homclosure of sentences in one of the well-behaved logics $\mathbb{GFO}_{\underline{=}}$ and $\mathbb{GNFO}_{\underline{=}}$ – guaranteeing solvability in P, by Theorem 47. For a start, note that Theorem 43 implies that the homclosures of $\mathbb{GNFO}_{\underline{=}}$ and $\mathbb{GFO}_{\underline{=}}$ sentences are expressible in guarded existential second-order logic; this even holds projectively. Yet, by a recent result of a subset of the authors [6], if the complement of a CSP can be expressed in this form, then the CSP is of the form CSP($\mathfrak{B}$) for an ω-categorical structure $\mathfrak{B}$. This implies that the so-called universal-algebraic approach can be used to study complexity and expressivity questions CSPs of said type [4].

References

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

[2] V. Bárány, G. Gottlob, and M. Otto. Querying the guarded fragment. *Log. Methods Comput. Sci.*, 10(2), 2014.

[3] R. Berger. *The undecidability of the domino problem*. Number 66. American Mathematical Soc., 1966.

[4] M. Bodirsky. *Complexity of Infinite-Domain Constraint Satisfaction*. Cambridge University Press, 2021. to appear in the LNL Series.

[5] M. Bodirsky, T. Feller, S. Knäuer, and S. Rudolph. On logics and homomorphism closure. *arXiv*, (CoRR abs/2104.11955), 2021.

[6] M. Bodirsky, S. Knäuer, and S. Rudolph. Datalog-expressibility for monadic and guarded second-order logic. In *Proceedings of the 48th International Colloquium on Automata, Languages, and Programming (ICALP)*, 2021. To appear, preprint available at https://arxiv.org/abs/2010.05677.

[7] E. Börger, E. Grädel, and Y. Gurevich. *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer, 1997.

[8] A. Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2):345–363, 1936.

[9] R. Fagin. Generalized first-order spectra, and polynomial-time recognizable sets. In R. Karp, editor, *Complexity of Computation, SIAM-AMS Proceedings*, volume 7, pages 43–73, 1974.

[10] M. L. Furst, J. B. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. *Math. Syst. Theory*, 17(1):13–27, 1984.

[11] E. Grädel. On the restraining power of guards. *J. Symb. Log.*, 64(4):1719–1742, 1999.

[12] E. Grädel, P. Kolaitis, and M. Y. Vardi. On the decision problem for two-variable first-order logic. *Bulletin of Symbolic Logic*, 3(1):53–69, 1997.

[13] M. Grohe. The quest for a logic capturing PTIME. In *Proceedings of the 23rd Annual IEEE Symposium on Logic in Computer Science (LICS)*, pages 267–271. IEEE Computer Society, 2008.

[14] Y. Gurevich and S. Shelah. Fixed-point extensions of first-order logic. *Ann. Pure Appl. Log.*, 32:265–280, 1986.

[15] Y. S. Gurevich and I. O. Koryakov. Remarks on Berger's paper on the domino problem. *Siberian Mathematical Journal*, 13(2):319–321, 1972.

[16] N. Immerman. Relational queries computable in polynomial time (extended abstract). In H. R. Lewis, B. B. Simons, W. A. Burkhard, and L. H. Landweber, editors, *Proceedings of the 14th Annual ACM Symposium on Theory of Computing (STOC)*, pages 147–152. ACM, 1982.

[17] N. Immerman. *Descriptive complexity*. Graduate texts in computer science. Springer, 1999.

[18] N. D. Jones. Space-bounded reducibility among combinatorial problems. *J. Comput. Syst. Sci.*, 11(1):68–85, 1975.

[19] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller, J. W. Thatcher, and J. D. Bohlinger, editors, *Proceedings of a symposium on the Complexity of Computer Computations*, pages 85–103, Boston, MA, 1972. Springer US.

[20] E. Kieroński and S. Rudolph. Finite model theory of the triguarded fragment and related logics. In L. Libkin, editor, *Proceedings of the 36th Annual Symposium on Logic in Computer Science (LICS)*. IEEE, 2021. In press.

[21] H. Lewis. Complexity results for classes of quantificational formulas. *J. Comput. Syst. Sci.*, 21:317–353, 1980.

[22] B. Rossman. Homomorphism preservation theorems. *Journal of the ACM*, 55(3), 2008.

[23] B. Rossman. An improved homomorphism preservation theorem from lower bounds in circuit complexity. In C. H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 67 of *LIPIcs*, pages 27:1–27:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.

[24] S. Rudolph and M. Krötzsch. Flag & check: data access with monadically defined queries. In R. Hull and W. Fan, editors, *Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, pages 151–162. ACM, 2013.

[25] S. Rudolph and M. Simkus. The triguarded fragment of first-order logic. In G. Barthe, G. Sutcliffe, and M. Veanes, editors, *Proceedings of the 22nd International Conference on Logic for Programming, Artificial Intelligence and Reasoning (LPAR)*, volume 57 of *EPiC Series in Computing*, pages 604–619. EasyChair, 2018.

[26] K. Schütte. Untersuchungen zum Entscheidungsproblem der mathematischen Logik. *Math. Annalen*, 109(4):572–603, 1934.

[27] J. Surányi. *Reduktionstheorie des Entscheidungsproblems im Prädikatenkalkül der ersten Stufe*. Ungarische Akademie der Wissenschaften, Budapest, 1959.

[28] B. ten Cate and M. Franceschet. Guarded fragments with constants. *Journal of Logic, Language and Information*, 14(3):281–288, 2005.

[29] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265, 1937.

[30] M. Y. Vardi. The complexity of relational query languages (extended abstract). In H. R. Lewis, B. B. Simons, W. A. Burkhard, and L. H. Landweber, editors, *Proceedings of the 14th Annual ACM Symposium on Theory of Computing (STOC)*, pages 137–146. ACM, 1982.

[31] H. Wang. Dominoes and the AEA case of the decision problem. In *Proceedings of the Symposium on Mathematical Theory of Automata*, pages 23–55. Polytechnic Press Brooklyn, NY", 1963.

[32] H. Wang. Dominoes and the AEA case of the decision problem. In *Computation, Logic, Philosophy: A Collection of Essays*, pages 218–245. Springer Netherlands, 1990.

A. On the Different Notions of Homclosure

At the first glance, the different notions $[\![\Phi]\!]^{\rightarrow}$, $[\![\Phi]\!]^{\rightarrow_{<\omega}}$, $[\![\Phi]\!]^{\rightarrow}_{<\omega}$, and $[\![\Phi]\!]^{\rightarrow_{<\omega}}_{<\omega}$ might seem abundant and give rise to the question, if some of these may coincide. We show that this is not the case.

From $[\![\Phi]\!]_{<\omega} \subseteq [\![\Phi]\!]$ and $\mathcal{C}^{\rightarrow_{<\omega}} \subseteq \mathcal{C}^{\rightarrow}$, one can directly infer the following inclusions:

- $[\![\Phi]\!]^{\rightarrow_{<\omega}}_{<\omega} \subseteq [\![\Phi]\!]^{\rightarrow}_{<\omega}$
- $[\![\Phi]\!]^{\rightarrow_{<\omega}}_{<\omega} \subseteq [\![\Phi]\!]^{\rightarrow_{<\omega}}$
- $[\![\Phi]\!]^{\rightarrow_{<\omega}} \subseteq [\![\Phi]\!]^{\rightarrow}$
- $[\![\Phi]\!]^{\rightarrow}_{<\omega} \subseteq [\![\Phi]\!]^{\rightarrow}$
- $[\![\Phi]\!]^{\rightarrow_{<\omega}}_{<\omega} \subseteq [\![\Phi]\!]^{\rightarrow}$ (by transitivity)

We proceed to show that all these inclusions are strict and that $[\![\Phi]\!]^{\rightarrow}_{<\omega}$ and $[\![\Phi]\!]^{\rightarrow_{<\omega}}$ are indeed incomparable, even for one fixed Φ.

Let Φ be the conjunction over the following sentences:

- $\forall x. \big(x \neq \mathtt{b} \Rightarrow \exists y. \mathrm{P}(x, y) \big)$
- $\forall x. \big(x \neq \mathtt{a} \Leftrightarrow \exists y. \mathrm{P}(y, x) \big)$
- $\forall xyz. \big(\mathrm{P}(x, y) \wedge \mathrm{P}(x, z) \Rightarrow y = z \big)$
- $\forall xyz. \big(\mathrm{P}(x, z) \wedge \mathrm{P}(y, z) \Rightarrow x = y \big)$

We now define the following $\{\mathtt{a}, \mathtt{b}, \mathrm{P}\}$-structures:

- $\mathfrak{A}_{k\text{-path}} = (\{0, \dots, k\}, 0, k, succ)$
- $\mathfrak{A}_{k\text{-path}^+} = (\mathbb{N}, 0, k, succ)$
- $\mathfrak{A}_{2\text{loops}} = (\{0, 1\}, 0, 1, \{(0, 0), (1, 1)\})$
- $\mathfrak{A}_{\infty\text{-gap}} = (\mathbb{Z} \setminus \{0\}, 1, -1, succ)$

We obtain:

	$[\![\Phi]\!]^{\rightarrow_{<\omega}}_{<\omega}$	$[\![\Phi]\!]^{\rightarrow}_{<\omega}$	$[\![\Phi]\!]^{\rightarrow_{<\omega}}$	$[\![\Phi]\!]^{\rightarrow}$
$\mathfrak{A}_{k\text{-path}}$	$\in$	$\in$	$\in$	$\in$
$\mathfrak{A}_{k\text{-path}^+}$	$\notin$	$\in$	$\notin$	$\in$
$\mathfrak{A}_{2\text{loops}}$	$\notin$	$\notin$	$\in$	$\in$
$\mathfrak{A}_{\infty\text{-gap}}$	$\notin$	$\notin$	$\notin$	$\in$

B. Tilings

We prove in this section the undecidability of the deterministic margin-constraint tiling problems. The proof is a slight modification of the first undecidability proof for origin-constraint tiling problems in [31] (see also [32]). We include it here for the convenience of the reader. The words "domino" and "tile" are used fully synonymously in the following. A tile system is the same as a domino system and a set of tiles is the same as a set of dominoes.

Lemma 2. *The following problems are both undecidable even for deterministic margin-constrained domino systems $\mathcal{D}$:*

1) Given $\mathcal{D}$, does there exist a $\mathcal{D}$-tiling?

2) Given $\mathcal{D}$, does there exist an ultimately periodic $\mathcal{D}$-tiling?

Proof. In order to prove Item 1 we follow the construction of Wang in [32].

He showed undecidability of the origin-constrained tiling problem by a reduction from the halting problem of deterministic Turing machines. For a given Turing machine, Wang defined a set of tiles such that the ith computation step of the machine is completely given by a tiling of the ith row of the $\mathbb{N} \times \mathbb{N}$ grid. Therefore, a tiling of the $\mathbb{N} \times \mathbb{N}$ grid can only be found if the Turing machine runs forever.

We show in the following how this can be done with a margin-constrained system of dominoes. Our construction depends, in addition to Wang's idea, on two modifications of the classical set of tiles. First we shift the Turing machine tape content in each new computation step one position to the right. To this end, we define a set of dominoes $\mathcal{D}_1$ of new tiles which occur only in the beginning of each row; their number increases by one from row to row. The second idea is that we store the tape content not once but three times. Each placed tile corresponds to a position on the tape and contains the tape content at this position, but furthermore it also contains the tape content of its left and right neighbors. This idea leads to the definition of the tile system $\mathcal{D}_2$.

For a binary relation R on a set D we denote by R also the map $R \colon D \to 2^D$ defined by $R(x) := \{d \in D \mid (x, d) \in R\}$. It will always be clear from the context which of the two objects we mean. Furthermore, if the set $R(x)$ contains only one element we drop the set brackets, i.e., we write $R(x) = d$ for $d \in D$.

Let $\mathcal{M} = \langle \{q_0, \dots, q_n, q_f\}, \{0, 1\}, \{0, 1\}, \delta, q_0, q_f \rangle$ be a deterministic Turing machine where $\{q_0, \dots, q_n, q_f\}$ is the set of states, $\{0, 1\}$ is the input and tape alphabet, and δ is a partial transition function $\{q_0, \dots, q_n\} \times \{0, 1\} \to \{q_0, \dots, q_n, q_f\} \times \{0, 1\} \times \{L, R\}$.

Let $D_1 := \{o, l, b, c, c', d\}$ be a set of dominoes such that o is the origin tile. We define the relations H_1 and V_1 on D_1 as follows:

- $H_1(o) := b$, $V_1(o) := l$,
- $H_1(b) := b$, $V_1(b) := \{c', d\}$,
- $V_1(l) := l$, $H_1(l) := \{c, c'\}$,
- $H_1(c) := \{c, c'\}$, $V_1(c) := c$,
- $H_1(c') := d$, $V_1(c') := c$,
- $V_1(d) := c'$.

In the next step we define tiles that contain the information of a Turing machine tape position. Consider the set $T := \{T_i^p \mid p \in \{q_0, \dots, q_n, q_f, e\}, i \in \{0, 1\}\}$. We define the set of tiles D_2 as $(T \uplus \{d\}) \times T \times T$. We denote the three coordinates of a tile $x \in D_2$ with x_l, x_m and x_r. We want to ensure in the following, that a tile $x \in D_2$ that is placed at the grid corresponds to a tape position of the Turing machine after a certain number of computation steps. The tape content is encoded in the main tape information $x_m = T_i^p$. The index i tells us which symbol is written on the tape at that certain position. If the head of the machine is at this position we get the state information by the superscript p. Otherwise $p = e$, which means the head is somewhere else. The coordinates x_l

and x_r give exactly this kind of information for the left and right neighbors of the tape position.

We give now the relation H_2 and V_2 for a tile $x \in D_2$.

$$H_2(x) := \{y \in D_2 \mid x_m = y_l \wedge x_r = y_m\} \quad (53)$$

For V_2 we make the following case distinction on the different types of the form $x = (T_r^u, T_s^v, T_t^w) \in D_2$.

1) $u = q$ for $q \in \{q_1, \ldots, q_n\}$ and $\delta(q, r) = (q', i, R)$:

$$V_2(x) := \{(y_l, T_i^e, T_s^{q'}) \in D_2 \mid y_l \in T \cup \{d\}\}. \quad (54)$$

2) $v = q$ for $q \in \{q_1, \ldots, q_n\}$ and $\delta(q, s) = (q', i, R)$:

$$V_2(x) := \{(y_l, T_r^u, T_i^e) \in D_2 \mid y_l \in T \cup \{d\}\}. \quad (55)$$

3) $v = q$ for $q \in \{q_1, \ldots, q_n\}$ and $\delta(q, s) = (q', i, L)$:

$$V_2(x) := \{(y_l, T_s^{q'}, T_i^e) \in D_2 \mid y_l \in T \cup \{d\}\}. \quad (56)$$

4) $w = q$ for $q \in \{q_1, \ldots, q_n\}$ and $\delta(q, s) = (q', i, L)$:

$$V_2(x) := \{(y_l, x_l, T_s^{q'}) \in D_2 \mid y_l \in T \cup \{d\}\}. \quad (57)$$

5) In all the remaining cases we define

$$V_2(x) := \{y \in D_2 \mid x_m = y_r\}. \quad (58)$$

Now consider the set of tiles $D := D_1 \cup D_2$. We define the relations H_3 and V_3 as follows. For the tiles in D_2 that are of the form $x = (d, T_s^v, T_t^w)$, where $v, w \in \{q_0, \ldots, q_n, e\}$ holds, we set $V_3(x) := \{d\}$. Furthermore we define $H_3(d) := (d \times T^2)$.

In the last step of our construction we have to specify the second row of the grid. This row should correspond to the initial tape configuration of the Turing machine. Therefore we define a relation V_4 such that

$$V(b) := \{(d, T_0^{q_0}, T_0^e), (T_0^{q_0}, T_0^e, T_0^e), (T_0^e, T_0^e, T_0^e)\}. \quad (59)$$

We claim that the domino system $\mathcal{D}$ on the set D with the relations $B := \{b, o\}$, $L := \{l, o\}$, $H := H_1 \cup H_2 \cup H_3$ and $V := V_1 \cup V_2 \cup V_3 \cup V_4$ is a deterministic margin constraint domino system which tiles the $\mathbb{N} \times \mathbb{N}$ grid if and only if the Turing machine $\mathcal{M}$ does not halt.

In order to prove this we give the following description of the tile system $\mathcal{D}$.

<u>Claim 1:</u> $\mathcal{M}$ runs at least k steps if and only if $\mathcal{D}$ tiles at least the first $k+1$ rows. In this case the $i+1$th row, with $1 \leq i \leq k$ has the following description from left to right:

- the margin symbol l,
- $i - 1$ times the symbol c,
- one time the symbol c',
- one time the symbol d,
- an infinite sequence of tiles from D_2 that store the Turing machine configuration after computation step $i - 1$.

It is clear that this claim holds for $k = 1$, by our encoding of the initial machine configuration in the second row (see definition of V_4). Now assume the statement holds for $k-1$. It is easy to see from the definition of H_1 and V_1 that the c-block is enlarged by one c from one row to the next. Note that our definition of V_2 ensures the last item in the claim. The main tape information of a tile gets the left neighbor information of its upper neighbor. This is due to the shifting of the tape. Furthermore one can check that the state transition and head position change in each computation step are also encoded in the definition of V_2 (see items 1 to 4). Note that this implies also that a tile where the main information consists of a final state does not have an upper neighbor. This is since there exists no transition in a deterministic Turing machine from a final state to another state. This proves Claim 1.

The domino system $\mathcal{D}$ satisfies clearly the items 1 to 3 in Definition 1. Let d_1 and d_2 be tiles. It is clear that item 4 is satisfied whenever $d_1 \in D_1 \setminus \{d\}$ or $d_2 \in D_1$ holds. Suppose therefore $d_1 \in D_2 \cup \{d\}$ and $d_2 \in D_2$. Note that for a tile $x \in H(d_1)$ the coordinates x_l and x_m are uniquely determined by d_1. The coordinate x_l corresponds as a tape position to $(d_2)_m$. The tape information at this position after the next computation step is determined by its two neighbors $(d_2)_l$ and $(d_2)_r$. The uniqueness of this transition can be checked in the definition of V_2.

This proves that $\mathcal{D}$ is deterministic. It follows from Claim 1 that this system tiles the grid whenever the machine $\mathcal{M}$ does not halt on the empty input. On the other hand, if the machine halts, then there exists a maximal k such that the first k rows can be tiled. This proves the first item of the lemma.

Item 2 is also proved by a reduction from the halting problem for Turing machines. We give a reduction such that a Turing machine halts on the empty tape whenever the corresponding tile system admits an ultimately periodic tiling of the grid. On the other hand if the Turing machine runs forever, the defined set of tiles does not allow for an ultimately periodic tiling.

Note that our reduction from Item 1 satisfies already the second implication. If the Turing machine $\mathcal{M}$ runs forever the tile system $\mathcal{D}$ tiles the grid by a tiling t. By the definition of $\mathcal{D}$ and Claim 1 above the ith row starts with exactly $i - 1$ times the symbol c after one l. This means that for every $\ell_{\text{init}}, \ell_{\text{period}} \in \mathbb{N}$ there exists $k \in \mathbb{N}$ such that $t(k, \ell_{\text{init}}) \neq c$ and $t(k, \ell_{\text{init}} + \ell_{\text{period}}) = c$. Therefore the tiling is not ultimately periodic. Moreover, since the used margin-constrained tile systems is deterministic, this is the unique tiling of the grid by this set of tiles, which proves that $\mathcal{D}$ does not admit a ultimately periodic tiling.

We modify the tiling from Item 1 in the following such that $\mathcal{D}$ tiles the grid ultimately periodically whenever the Turing machine $\mathcal{M}$ halts. If the Turing machine halts it is in the final state. Also, we can assume without loss of generality that the reading head will then be at the leftmost tape position. Let f and e be new tiles and let D' be the union of the set of

dominoes from Item 1 with $\{f, e\}$. We define new relations V_5 and H_4 as follows.

1) Let $x \in D_2$. If x is of the form $x = (d, T_s^{q_f}, T_t^w)$ where q_f is the final state we define $V_5(x) := f$.
Otherwise $V_5(x) := e$.

2) $V_5(f) := f$ and $V_5(e) := e$.

3) $H_4(c) := f$, $H_4(c') := f$, $H_4(f) := e$ and $H_4(e) := e$.

One can check that the new tile system $\mathcal{D}'$ on the set of tiles D' with the relations $B' := B$, $L' := L$, $H' := H \cup H_4$ and $V' := V \cup V_5$ is still a deterministic margin-constrained domino system: Consider tiles d_1 and d_2 and assume first that $d_1 \in D$ and $d_2 \in D$ hold, then the only possibilities for d_1 that have to be checked are $d_1 = c$ and $d_1 = c'$. The definition of V_5 involves only old elements from D_2 (Item 1 above). Therefore d_2 can be assumed to be from D_2. It is easy to see that in this cases $|H'(d_1) \cap V'(d_2)| \leq 1$ holds. The remaining cases are $d_1 \in \{e, f\}$ or $d_2 \in \{e, f\}$. In these cases $|H'(d_1) \cap V'(d_2)| \leq 1$ holds trivially since e and f have unique upper and right neighbors.

We characterize now how a tiling with the system $\mathcal{D}'$ looks like if the Turing machine halts.

Claim 2: If $\mathcal{M}$ runs k steps and then halts, then $\mathcal{D}$ tiles the grid such that for every $i \geq k+3$ the ith row has the following description from left to right

- The margin symbol l,
- k times the symbol c,
- one time the symbol f,
- a infinite sequence of tiles e.

Note that we have by Claim 1 a description of the kth row. Since the Turing machine halts after k steps the first symbol is this row after the symbol d is of the kind $x = (d, T_s^{q_f}, T_t^w)$ where q_f is the final state. Therefore the unique tiling of the $k + 2$th row is as follows:

- the margin symbol l,
- k times the tile c,
- one time the tile c',
- one time the tile f, and
- a infinite sequence of tiles e.

From this description it is easy to compute the unique tilings of the $k + 3$th and $k + 4$th row. Since both are identical and by Item 4 in Definition 1 it follows that all further rows admit also tilings that are identical to the $k + 3$th one. This proves Claim 2.

The tiling described in Claim 2 is ultimately periodic. For the vertical period this follows directly from Claim 2. For the horizontal period node that a Turing machine can in k computation step reach at most $k + 1$ different tape positions. This implies that in the rows 2 to $k + 2$ the tiles right of the $2(k + 1)$th column encode the empty tape. Since Claim 2 ensures the horizontal period for the rows greater than $k + 1$

the tiling is ultimately periodic.

Note also that our modifications to achieve this do not affect the property of a (or the) tiling being not ultimately periodic if the Turing machine runs forever. This finishes the proof of the lemma. $\qquad \square$

C. Proofs of Tools Section

Lemma 3. *Let Φ be any $\mathbb{FO}_=$ sentence. Then $\mathfrak{A}^\lambda \models \Phi$ exactly if $\mathfrak{A}_\lambda \models \mathrm{tr}^n(\Phi)$.*

Proof. Note that the negation normal form for a formula Φ is equivalent to Φ itself. Then, by the definition of tr^n it is enough to prove the statement for sentences Φ in NNF. In order to do this, we prove the following statement. For every $\mathbb{FO}_=$ formula $\Phi(\boldsymbol{x})$ in NNF, every n-labelled constant-sole structure $(\mathfrak{A}, \lambda)$, and every variable assignment $\beta \colon V \to \mathfrak{A}^\lambda$ it holds that

$$(\mathfrak{A}^\lambda, \beta) \models \Phi(\boldsymbol{x}) \quad \text{iff} \quad (\mathfrak{A}_\lambda, \beta_1) \models \mathrm{tr}^n_{\beta_2}(\Phi(\boldsymbol{x})), \quad (60)$$

where $\beta_1 \colon V \to A$ and $\beta_2 \colon V \to \{1, \dots, n\}$ are defined such that $\beta(x) = (\beta_1(x), \beta_2(x))$ for all $x \in V$. We denote by β_2^+ the extensions of the assignment β_2 to all constants c such that $\beta_2^+(c) = 1$ (cf. definition of tr^n_{as}). Furthermore β_1^+ is the extension of β_1 to all constants c such that $\beta_1^+(c) = c^{\mathfrak{A}_\lambda}$. The map β^+ is now defined by $\beta^+(c) := (\beta_1^+(c), \beta_2^+(c))$. Note that this is compatible with the definition of the constants in the constant-sole structure $\mathfrak{A}^\lambda$ since $c^{\mathfrak{A}^\lambda} = (c^{\mathfrak{A}}, 1)$ holds.

We prove the statement by structural induction over the set of formulae. For literals of the form $\mathrm{P}(\boldsymbol{t})$ or $\neg \mathrm{P}(\boldsymbol{t})$ the statement follows immediately from the definition of tr^n_{as} since the projection map from $\mathfrak{A}^\lambda$ to $\mathfrak{A}_\lambda|_\tau$ is a strong surjective homomorphism. Therefore the only cases that we have to check are formulae of the form $t_1 = t_2$ and $t_1 \neq t_2$ for terms t_1 and t_2.

Suppose that β is any variable assignment such that $(\mathfrak{A}^\lambda, \beta) \models (t_1 = t_2)$. Then $\beta_1^+(t_1) = \beta_1^+(t_2)$ and $\beta_2^+(t_1) = \beta_2^+(t_2)$. This implies together with (8) that $(\mathfrak{A}_\lambda, \beta_1) \models \mathrm{tr}^n_{\beta_2}(t_1 = t_2)$ holds. For the other direction observe that if $(\mathfrak{A}_\lambda, \beta_1) \models \mathrm{tr}^n_{\beta_2}(t_1 = t_2)$ holds, then $\mathrm{tr}^n_{\beta_2}(t_1 = t_2)$ is exactly the formula $t_1 = t_2$ and for the assignment β it holds that $\beta_2^+(t_1) = \beta_2^+(t_2)$ by (5). Furthermore we get in this case that $\beta_1^+(t_1) = \beta_1^+(t_2)$ must hold and therefore we get $(\mathfrak{A}^\lambda, \beta) \models (t_1 = t_2)$. This concludes the second direction.

For the case of a literal of the form $t_1 \neq t_2$ suppose that β is a variable assignment such that $(\mathfrak{A}^\lambda, \beta) \models (t_1 \neq t_2)$ holds. This implies that $\beta_1^+(t_1) \neq \beta_1^+(t_2)$ or $\beta_2^+(t_1) \neq \beta_2^+(t_2)$ holds. For the first possibility it follows by (9) that $(\mathfrak{A}_\lambda, \beta_1) \models \mathrm{tr}^n_{\beta_2}(t_1 \neq t_2)$ holds. In the case that $\beta_2^+(t_1) \neq \beta_2^+(t_2)$ holds, we get by (9) that $\mathrm{tr}^n_{\beta_2}(t_1 \neq t_2)$ is precisely $\top$ and therefore again the implication is valid. For the other direction assume that $(\mathfrak{A}_\lambda, \beta_1) \models \mathrm{tr}^n_{\beta_2}(t_1 \neq t_2)$ holds. Then either $\mathrm{tr}^n_{\beta_2}(t_1 \neq t_2)$ is equal to $\top$ or it is equal to $t_1 \neq t_2$ and $\beta_1^+(t_1) \neq \beta_1^+(t_2)$. In both cases at least one of β_i satisfies $\beta_i^+(t_1) \neq \beta_i^+(t_2)$ and therefore $\beta^+(t_1) \neq \beta^+(t_2)$ holds. This proves the claim.

For the inductive step it is straightforward to see that statement holds for all formulae of the form $\Phi(\boldsymbol{x}) \wedge \Psi(\boldsymbol{x})$ and $\Phi(\boldsymbol{x}) \vee \Psi(\boldsymbol{x})$ whenever it holds for $\Phi(\boldsymbol{x})$ and $\Psi(\boldsymbol{x})$.

It therefore remains to prove the statement for formulae of the form $\exists x.\Phi(x, \boldsymbol{y})$ and $\forall x.\Phi(x, \boldsymbol{y})$. Let β be an arbitrary variable assignment such that $(\mathfrak{A}^\lambda, \beta) \models \exists x.\Phi(x, \boldsymbol{y})$ holds. This is true if and only if there exists a variable assignment $\beta' := \beta \cup \{x \mapsto (a, j)\}$, where (a, i) is a domain element of $\mathfrak{A}^\lambda$ such that $(\mathfrak{A}^\lambda, \beta') \models \Phi(x, \boldsymbol{y})$ holds. By the inductive assumption we get that this is the case if and only if $(\mathfrak{A}_\lambda, \beta'_1) \models \mathrm{tr}^n_{\beta'_2}(\Phi(x, \boldsymbol{y}))$ holds. Since $[\lambda(a) \geq j]$ holds clearly in $\mathfrak{A}_\lambda$ this is equivalent to the validity of $(\mathfrak{A}_\lambda, \beta'_1) \models [\lambda(x) \geq j] \wedge \mathrm{tr}^n_{\beta'_2}(\Phi(x, \boldsymbol{y}))$ for $\beta' = \beta \cup \{x \mapsto (a, j)\}$. Note that such a β can be found if and only if $(\mathfrak{A}_\lambda, \beta_1) \models \exists x.[\lambda(x) \geq j] \wedge \mathrm{tr}^n_{\beta_2}(\Phi(x, \boldsymbol{y}))$ holds for some j. By Definition (12) this means that $(\mathfrak{A}_\lambda, \beta_1) \models \mathrm{tr}^n_{\beta_2}(\exists x.\Phi(x, \boldsymbol{y}))$ holds.

The statement for formulae of the form $\forall x.\Phi(x, \boldsymbol{y})$ can be shown by an analogous argument. $\square$

D. Proofs of Homclosure Membership Section

Lemma 62. *Let Φ be a τ-sentence and let $\mathfrak{A}$ be a finite τ-structure as in Definition 5. Then the following holds:*

1) *$\Phi^{\mathrm{ext}}_{\mathfrak{A}}$ and $\Phi^{\mathrm{int}}_{\mathfrak{A}}$ are of polynomial size wrt. $|\mathfrak{A}|$ and the size of Φ.*

2) *$\Omega_{\mathfrak{A}}$ projectively characterizes the class of (finite) τ-structures having a homomorphism into $\mathfrak{A}$.*

3) *$\Phi^{\mathrm{ext}}_{\mathfrak{A}}$ projectively characterizes the class of (finite) models of Φ having a homomorphism into $\mathfrak{A}$.*

Proof. The first statement follows directly from the definitions of $\Phi^{\mathrm{ext}}_{\mathfrak{A}}$ and $\Phi^{\mathrm{int}}_{\mathfrak{A}}$.

For (2), let $\mathfrak{B}$ be a $\tau \cup \rho$-structure that satisfies $\Omega_{\mathfrak{A}}$. Let λ be the n-labeling of $\mathfrak{B}$ given by the ρ-reduct of $\mathfrak{B}$. The map λ is a homomorphism from the τ-reduct $\mathfrak{B}'$ of $\mathfrak{B}$ to $\mathfrak{A}$. To see this assume that $\boldsymbol{b} \in \mathrm{P}^{\mathfrak{B}'}$ holds for a tuple $\boldsymbol{b}$ from B. By the definition of $\Omega_{\mathfrak{A}}$ we get that $\boldsymbol{b} \in (\chi^{\mathrm{P}}_{\mathfrak{A}})^{\mathfrak{B}}$ holds. This implies that $\lambda(\boldsymbol{b}) \in \mathrm{P}^{\mathfrak{A}}$ holds and therefore λ is a homomorphism. For the other inclusion assume the $h \colon \mathfrak{B} \to \mathfrak{A}$ is a homomorphism. It is easy to check that the implicit representation $\mathfrak{B}_h$ of the n-labeled structure $(\mathfrak{B}, h)$ satisfies $\Omega_{\mathfrak{A}}$ and is an expansion of $\mathfrak{B}$. This proves the claim.

Claim (3) follows immediately from the definition of $\Phi^{\mathrm{ext}}_{\mathfrak{A}}$ and the previous claim. $\square$

Lemma 63. *Let $\mathfrak{A}$ and ρ as in Definition 5 and let $\mathfrak{B}$ be an arbitrary τ-structure such that $h \colon \mathfrak{B} \to \mathfrak{A}$ is a homomorphism. Then it holds for every τ-formula Φ and all assignments β that*

$$(\mathfrak{B}, \beta) \models \Phi(\boldsymbol{t}) \quad \text{if and only if} \quad (\mathfrak{B}_h, \beta) \models \Phi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}),$$

where $\mathfrak{B}_h$ is the implicit representation of the n-labeld structure $(\mathfrak{B}, h)$.

Proof. We prove the statement by structural induction over the set of formulae. For the induction beginning let $\Phi(\boldsymbol{t})$ be an atomic formula $\mathrm{P}(\boldsymbol{t})$ and let β be a term assignment such that $(\mathfrak{B}, \beta) \models \Phi(\boldsymbol{t})$ holds. Since h is a homomorphism it follows that $(\mathfrak{A}, h \circ \beta) \models \mathrm{P}(\boldsymbol{t})$ and therefore also $(\mathfrak{B}_h, \beta) \models \chi^{\mathrm{P}}_{\mathfrak{A}}(\boldsymbol{t})$ holds. The second direction of the equivalence follows directly from the definition of $\chi^{\mathrm{P}}_{\mathfrak{A}}$ and since $\mathfrak{B}_h$ is an expansion of $\mathfrak{B}$.

For the induction step let $\Phi(\boldsymbol{t})$ be of the form $\neg \Psi(\boldsymbol{t})$ such that the statement already holds for Ψ. Note that then clearly $\Phi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t})$ is equal to $\neg \Psi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t})$ and we get:

$$
\begin{aligned}
(\mathfrak{B}, \beta) \models \Phi(\boldsymbol{t}) \quad &\text{iff} \quad (\mathfrak{B}, \beta) \not\models \Psi(\boldsymbol{t}) \\
&\text{iff} \quad (\mathfrak{B}_h, \beta) \not\models \Psi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}) \\
&\text{iff} \quad (\mathfrak{B}_h, \beta) \models \Phi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}).
\end{aligned}
$$

Assume that $\Phi(\boldsymbol{t})$ is of the from $\Psi(\boldsymbol{t}) \vee \Gamma(\boldsymbol{t})$ for formulae $\Psi(\boldsymbol{t})$ and $\Gamma(\boldsymbol{t})$ for which the statement holds. Since $\Phi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t})$ is by its definition of the form $\Psi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}) \vee \Gamma^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t})$ we get

$$
\begin{aligned}
(\mathfrak{B}, \beta) &\models \Phi(\boldsymbol{t}) \\
\text{iff} \quad &(\mathfrak{B}, \beta) \models \Psi(\boldsymbol{t}) \vee \Gamma(\boldsymbol{t}) \\
\text{iff} \quad &(\mathfrak{B}, \beta) \models \Psi(\boldsymbol{t}) \ \text{ or } \ (\mathfrak{B}, \beta) \models \Gamma(\boldsymbol{t}) \\
\text{iff} \quad &(\mathfrak{B}_h, \beta) \models \Psi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}) \ \text{ or } \ (\mathfrak{B}_h, \beta) \models \Gamma^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}) \\
\text{iff} \quad &(\mathfrak{B}_h, \beta) \models \Psi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}) \vee \Gamma^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}) \\
\text{iff} \quad &(\mathfrak{B}_h, \beta) \models \Phi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}).
\end{aligned}
$$

As the last case assume that $\Phi(\boldsymbol{t})$ is of the from $\exists t_1 \Psi(\boldsymbol{t})$ for formulae $\Psi(\boldsymbol{t})$ for which the statement holds. Then $\Phi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t})$ is by its definition of the form $\exists t_1 \Psi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t})$. We get the following:

$$
\begin{aligned}
(\mathfrak{B}, \beta) &\models \Phi(\boldsymbol{t}) \\
\text{iff} \ \ &\exists b \in B \text{ such that } (\mathfrak{B}, \beta \cup \{t_1 \mapsto b\}) \models \Psi(\boldsymbol{t}) \\
\text{iff} \ \ &\exists b \in B \text{ such that } (\mathfrak{B}_h, \beta \cup \{t_1 \mapsto b\}) \models \Psi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}) \\
\text{iff} \ \ &(\mathfrak{B}_h, \beta) \models \Phi^{\mathrm{int}}_{\mathfrak{A}}(\boldsymbol{t}).
\end{aligned}
$$

This concludes the proof of the lemma. $\square$

Lemma 6. *Let Φ be an arbitrary τ-sentence and let $\mathfrak{A}$ be a finite τ-structure. The following are equivalent:*

1) Φ has a (finite) model admitting a homomorphism into $\mathfrak{A}$.

2) $\Phi^{\mathrm{int}}_{\mathfrak{A}}$ is (finitely) satisfiable.

3) $\Phi^{\mathrm{ext}}_{\mathfrak{A}}$ is (finitely) satisfiable.

Proof. By Lemma 62 Item 3, the sentence $\Phi^{\mathrm{ext}}_{\mathfrak{A}}$ projectively characterizes the class of (finite) models of Φ having a homomorphism into $\mathfrak{A}$. Therefore the equivalence of 1) and 3) follows immediately.

The implication from 1) to 2) follows from Lemma 63, since for every (finite) model $\mathfrak{B}$ of Φ admitting a homomorphism h into $\mathfrak{A}$, the structure $\mathfrak{B}_\lambda$ with $\lambda := h$ is a (finite) model of $\Phi^{\mathrm{int}}_{\mathfrak{A}}$.

For the implication from 2) to 1) assume that $\mathfrak{B}$ is a model of $\Phi^{\mathrm{int}}_{\mathfrak{A}}$ and let λ be the n-labeling of $\mathfrak{B}$ given by the ρ-reduct of $\mathfrak{B}$. Let $\mathfrak{C}$ be the τ-structure obtained by setting $C := B$,

$c^{\mathfrak{C}} := c^{\mathfrak{B}}$ for all constants, and $P^{\mathfrak{C}} := (\chi_{\mathfrak{A}}^{P})^{\mathfrak{B}}$. It follows from the definition of $\chi_{\mathfrak{A}}^{P}(t)$ that λ is a homomorphism from $\mathfrak{C}$ to $\mathfrak{A}$. Furthermore the definition of $\mathfrak{C}$ implies that for all $P \in \tau$ and all $\boldsymbol{b}$ from B it holds

$$\boldsymbol{b} \in (\chi_{\mathfrak{A}}^{P})^{\mathfrak{C}_{\lambda}} \quad \text{if and only if} \quad \boldsymbol{b} \in (\chi_{\mathfrak{A}}^{P})^{\mathfrak{B}}.$$

Since $\mathfrak{B}$ satisfies $\Phi_{\mathfrak{A}}^{\mathrm{int}}$ it follows that $\mathfrak{C}_{\lambda}$ does as well. Now we apply Lemma 63 and get that $\mathfrak{C}$ satisfies Φ which proves that 1) holds. $\qquad\square$

Theorem 7. *Let* $\mathbb{L} \subseteq \mathbb{SO}$ *and* $\mathbb{L}^{\mathrm{ext}} := \{\Phi_{\mathfrak{A}}^{\mathrm{ext}} \mid \Phi \in \mathbb{L}, \mathfrak{A} \text{ finite}\}$ *as well as* $\mathbb{L}^{\mathrm{int}} := \{\Phi_{\mathfrak{A}}^{\mathrm{int}} \mid \Phi \in \mathbb{L}, \mathfrak{A} \text{ finite}\}$. *If* $\mathbb{L}^{\mathrm{ext}}$ *or* $\mathbb{L}^{\mathrm{int}}$ *has the FMP, then* $[\![\Psi]\!]^{\to_{<\omega}} = [\![\Psi]\!]_{<\omega}^{\to_{<\omega}}$ *for every* $\Psi \in \mathbb{L}$.

Proof. In order to show $[\![\Psi]\!]^{\to_{<\omega}} \subseteq [\![\Psi]\!]_{<\omega}^{\to_{<\omega}}$ let $\mathfrak{B} \in [\![\Psi]\!]^{\to_{<\omega}}$. Then $\Psi_{\mathfrak{B}}^{\mathrm{ext}}$ and $\Psi_{\mathfrak{B}}^{\mathrm{int}}$ have by Lemma 6 models. By the assumption $\mathbb{L}^{\mathrm{ext}}$ or $\mathbb{L}^{\mathrm{int}}$ have the FMP which means that $\Phi_{\mathfrak{B}}^{\mathrm{ext}}$ or $\Phi_{\mathfrak{B}}^{\mathrm{int}}$ are finitely satisfiable. Again by Lemma 6, we get that Φ has a finite model admitting a homomorphism into $\mathfrak{B}$ which shows that $\mathfrak{B} \in [\![\Psi]\!]_{<\omega}^{\to_{<\omega}}$ holds. $\qquad\square$

Theorem 8. $\mathrm{InHomCl}_{(\mathrm{fin})}$ *is complete for*

1) 2ExpTime *for* $\mathbb{GNFO}_{=}$,
2) N2ExpTime *for* $\mathbb{TGF}$,
3) NExpTime *for* $\mathbb{FO}_{=}^{2}$, $\exists^{*}\forall\forall\exists^{*}\mathbb{FO}$, $\exists^{*}\forall^{*}\mathbb{FO}_{=}$, $\forall^{*}\mathbb{FO}_{=}$,
4) NP *for* $\exists^{*}\mathbb{FO}_{=}$ *and* $\exists^{*}\mathbb{FO}_{=}^{+}$.

Proof. First consider $\exists^{*}\mathbb{FO}_{=}^{+}$, noting that every such sentence is already homclosed. Thus $\mathrm{InHomCl}_{(\mathrm{fin})}$ is equivalent to checking modelhood for the given $\mathfrak{A}$. This can be done in nondeterministic time by guessing an instantiation of the existentially quantified variables and checking satisfaction. NP-hardness follows from a reduction from the graph homomorphism problem (even for disjunction-free sentences). We now consider the other fragments. Extrinsic coloring for $\mathbb{GNFO}_{=}$ and $\mathbb{TGF}$ and intrinsic coloring for the others establishes a polynomial reduction to the corresponding satisfiability problems via Lemma 6. Hardness follows by a reduction from satisfiability via Proposition 4 in all cases (see Table I for references for the complexity of the satisfiability problem in each case). $\qquad\square$

Definition 9. *We let* $\mathfrak{A}_{\mathbb{N}\times\mathbb{N}}$ *denote the* $\mathbb{N} \times \mathbb{N}$ *grid, i.e., the* $\{\mathtt{H}, \mathtt{V}\}$-*structure with domain* $A_{\mathbb{N}\times\mathbb{N}} = \mathbb{N} \times \mathbb{N}$ *that satisfies* $\mathtt{H}^{\mathfrak{A}_{\mathbb{N}\times\mathbb{N}}} = \{((k,\ell),(k+1,\ell)) \mid k,\ell \in \mathbb{N}\}$ *as well as* $\mathtt{V}^{\mathfrak{A}_{\mathbb{N}\times\mathbb{N}}} = \{((k,\ell),(k,\ell+1)) \mid k,\ell \in \mathbb{N}\}$.

Let Φ_{grid} *denote the* $\mathbb{TGD}$ *sentence*

$$\left(\forall x \exists y.\mathtt{H}(x,y)\right) \wedge \left(\forall x \exists y.\mathtt{V}(x,y)\right) \wedge \\ \left(\forall xyzv.\mathtt{H}(x,y) \wedge \mathtt{V}(x,z) \wedge \mathtt{V}(y,v) \Rightarrow \mathtt{H}(z,v)\right). \tag{16}$$

For a given domino system $\mathcal{D} = (D, D, D, H, V)$, *we define the finite* $\{\mathtt{H}, \mathtt{V}\}$-*structure* $\mathfrak{A}_{\mathcal{D}}$ *as follows:* $A_{\mathcal{D}} = D$ *as well as* $\mathtt{H}^{\mathfrak{A}_{\mathcal{D}}} = H$ *and* $\mathtt{V}^{\mathfrak{A}_{\mathcal{D}}} = V$.

Lemma 64. *1) Every model* $\mathfrak{B}$ *of* Φ_{grid} *is a* $\{\mathtt{H}, \mathtt{V}\}$-*structures such that for every domain element* $b \in B$ *there is a homomorphism* $h : \mathfrak{A}_{\mathbb{N}\times\mathbb{N}} \to \mathfrak{B}$ *with* $h((0,0)) = b$.

2) *Every model of* Φ_{grid} *has a homomorphism into a structure that is not a model of* Φ_{grid}.

3) *Every homomorphism from the* $\mathbb{N} \times \mathbb{N}$ *grid into* $\mathfrak{A}_{\mathcal{D}}$ *corresponds to a* $\mathcal{D}$-*tiling.*

4) *Every homomorphism from some model of* Φ_{grid} *into* $\mathfrak{A}_{\mathcal{D}}$ *gives rise to a* $\mathcal{D}$-*tiling.*

5) *Every homomorphism from some finite model of* Φ_{grid} *into* $\mathfrak{A}_{\mathcal{D}}$ *gives rise to an ultimately periodic* $\mathcal{D}$-*tiling.*

6) *Both membership in* $[\![\Phi_{\mathrm{grid}}]\!]^{\to}$ *and* $[\![\Phi_{\mathrm{grid}}]\!]_{<\omega}^{\to}$ *are undecidable problems.*

Proof. For (1), note that an easy inductive argument under the usage of the extension properties defined in Φ_{grid} proves the claim.

For (2), simply choose $\mathfrak{F} \uplus \mathfrak{I}$ as target structure.

(3) follows immediately from the definition of a tiling and from the definition of the structure $\mathfrak{A}_{\mathcal{D}}$.

For (4), let $\mathfrak{B}$ be a model of Φ_{grid} that has a homomorphism to $\mathfrak{A}_{\mathcal{D}}$. By (1) we get that $\mathfrak{A}_{\mathbb{N}\times\mathbb{N}}$ also has a homomorphism to $\mathfrak{B}$ and therefore also to $\mathfrak{A}_{\mathcal{D}}$. (3) implies that this gives rise to a $\mathcal{D}$-tiling.

For (5), assume that $\mathfrak{B}$ is a finite model of Φ_{grid} with a homomorphism t into $\mathfrak{A}_{\mathcal{D}}$. Let $v : B \to B$ and $h : B \to B$ to functions such that $(x, h(x)) \in \mathtt{H}^{\mathfrak{B}}$ and $(x, v(x)) \in \mathtt{V}^{\mathfrak{B}}$ hold. It is clear that such functions exist since $\mathfrak{B}$ satisfies Φ_{grid}. We define a map f from $\mathbb{N} \times \mathbb{N}$ to B as follows. For an arbitrary element $o \in B$ we set $f(0,0) := o$. In the next step we define inductively the map on the elements $W_0 := \{(x,0) \mid x \in \mathbb{N}\}$. This is done by $f(i+1, 0) := h(f(i,0))$. Since h is a function this is well defined. Now we can assume that the map is defined on the elements $W_i := \{(x,i) \mid x \in \mathbb{N}\}$ and we extend it to $W_{i+1} := \{(x, i+1) \mid x \in \mathbb{N}\}$ by $f(j, i+1) := v(f(j, i+1))$. Note that by the use of the function h it is clear that after a finite initial sequence of lenght at most $|B|$, the mapping f is periodic on the first row W_0 with respect to the assigned elements from B. Since the definition of f at the nth column of the grid only depends on the value $f(n,0)$ the columns are in the same way periodic as W_0 is. Furthermore the graph of the map v contains finitely many cycles of finite length. Let q be the greatest common multiple of these lengths. It is clear that for every element $x \in B$ that lies in a cycle applying q times the function v on x results in the element x itself. This shows that f also has a repetition on the rows in the grid in its definition after a possible initial sequence of at most $|B|$ elements.

What we observed about the function f implies immediately that if $t \circ f$ defines a tiling of the grid by tiles of $\mathcal{D}$ then this tiling is ultimately periodic.

In order to prove that $t \circ f$ defines a tiling it is enough according

to (3) to show that f is a homomorphism from $\mathfrak{A}_{\mathbb{N}\times\mathbb{N}}$ to $\mathfrak{B}$.

Consider the substructure $\mathfrak{W}_0$ of $\mathfrak{A}_{\mathbb{N}\times\mathbb{N}}$ induced by W_0.

The map from $\mathfrak{W}_0$ to $\mathfrak{B}$ induced by f is a homomorphism, since $(f(i,0), f(i+1,0)) = (f(i,0), h(f(i,0))) \in \mathsf{H}^{\mathfrak{B}}$ holds for all $((i,0),(i+1,0)) \in \mathsf{H}^{\mathfrak{W}_0}$.

Now we prove by induction that f is a homomorphism on all structure $\mathfrak{W}_{i+1}$ which are induced by the sets W_{i+1}. It follows immediately by the choice of v that all tuples of the form $((j,i),(j,i+1)) \in \mathsf{V}^{\mathfrak{W}_{i+1}}$ for arbitrary $j \in \mathbb{N}$ are preserved by f. It remain to show the preservation of tuples of the form $((j,i+1),(j+1,i+1)) \in \mathsf{H}^{\mathfrak{W}_{i+1}}$. Consider the elements $f(j,i+1), f(j+1,i+1), f(j,i)$ and $f(j+1,j)$ in B. By the induction hypothesis and the observation from before we get that $\mathsf{H}(f(j,i), f(j+1,i))$, $\mathsf{V}(f(j,i), f(j,i+1))$ and $\mathsf{V}(f(j+1,i), f(j+1,i+1))$ hold in $\mathfrak{B}$. Since $\mathfrak{B}$ satisfies the sentence Φ_{grid} it follows that also $\mathsf{H}(f(j,i+1), f(j+1,i+1))$ holds in $\mathfrak{B}$. This implies that f induces a homomorphism from $\mathfrak{W}_{i+1}$ to $\mathfrak{B}$ and therefore concludes the proof that f is a homomorphism.

For (6), we prove the statements by reductions from the tiling problem and the ultimately periodic tiling problem. The undecidability of these problems was shown in [3] and in [15], respectively.

The reductions map a instance of the tiling problem, that is a set of tiles $\mathcal{D}$ to the structure $\mathfrak{A}_{\mathcal{D}}$.

To see that this is a correct reduction to the membership problem of $[\![\Phi_{\mathrm{grid}}]\!]^{\to}$ note that $\mathcal{D}$ tiles by (3) the grid if and only if $\mathfrak{A}_{\mathbb{N}\times\mathbb{N}}$ has a homomorphism to $\mathfrak{A}_{\mathcal{D}}$. By (1) and since $\mathfrak{A}_{\mathbb{N}\times\mathbb{N}}$ is a model of Φ_{grid} this holds if and only if $\mathfrak{A}_{\mathcal{D}} \in [\![\Phi_{\mathrm{grid}}]\!]^{\to}$. This proves the undecidability of the membership problem for $[\![\Phi_{\mathrm{grid}}]\!]^{\to}$.

For the undecidability of the membership problem for $[\![\Phi_{\mathrm{grid}}]\!]^{\to}_{<\omega}$ we use the same mapping from sets of tiles to structures. Assume that a set of tiles $\mathcal{D}$ admits an ultimately periodic tiling t of the grid. Let $k_{\mathrm{init}}, k_{\mathrm{period}}, \ell_{\mathrm{init}}, \ell_{\mathrm{period}}$ from the definition of ultimately periodic tiling. Let $\mathfrak{D}'$ be the finite substructure of $\mathfrak{A}_{\mathbb{N}\times\mathbb{N}}$ that is induced by elements of the from (x,y) where $x \leq k_{\mathrm{init}} + k_{\mathrm{period}}$ and $y \leq l_{\mathrm{init}} + l_{\mathrm{period}}$ hold. We add new tuples to the relations $\mathsf{H}^{\mathfrak{D}'}$ and $\mathsf{V}^{\mathfrak{D}'}$ from $\mathfrak{D}'$ and define in this way a new structure $\mathfrak{D}$. For all elements of the form $(k_{\mathrm{init}} + k_{\mathrm{period}}, y)$ we add the tuple $((k_{\mathrm{init}} + k_{\mathrm{period}}, y),(k_{\mathrm{init}}+1,y))$ to the relation $\mathsf{H}^{\mathfrak{D}'}$. Analogous we add for all elements of the form $(x, l_{\mathrm{init}} + l_{\mathrm{period}})$ the tuple $((x, l_{\mathrm{init}}+l_{\mathrm{period}}),(x,l_{\mathrm{init}}+1))$ to the relation $\mathsf{V}^{\mathfrak{D}'}$. It is easy to see that the so defined structure $\mathfrak{D}$ satisfies Φ_{grid}. Since the tiling t was ulitmately periodic the restriction of the tiling to the elements of $\mathfrak{D}$ defines a homomorphism from $\mathfrak{D}$ to $\mathfrak{A}_{\mathcal{D}}$. Since $\mathfrak{D}$ is clearly finite this implies that the structure $\mathfrak{A}_{\mathcal{D}}$ is in $[\![\Phi_{\mathrm{grid}}]\!]^{\to}_{<\omega}$.

For the other direction of the reduction proof assume that $\mathfrak{A}_{\mathcal{D}}$ is in $[\![\Phi_{\mathrm{grid}}]\!]^{\to}_{<\omega}$. This implies by (5) immediately that $\mathcal{D}$ admits an ultimately periodic tiling of the grid. $\square$

Theorem 10. $\mathrm{InHomCl}_{(\mathrm{fin})}$ *is undecidable for* $\mathbb{TGD}$ *(and hence for* $\mathbb{MDTGD}$ *and* $\mathbb{DTGD}$*).*

Proof. Immediate from Item 6 of Lemma 64. $\square$

E. Proofs of Homclosedness Section

Lemma 11. *Let* $\mathbb{L}, \mathbb{L}' \subseteq \mathbb{SO}$ *and* $f\colon \mathbb{L}' \to \mathbb{L}$ *such that*
- *(finite) unsatisfiability in* $\mathbb{L}'$ *is* C*-hard and*
- f *is* C'*-computable and maps* τ*-sentences to* $\tau \cup \{\mathsf{U}\} \cup \tau'$*-sentences with* $[\![f(\Phi)]\!]_{(<\omega)}|_{\tau\cup\{\mathsf{U}\}} = \{(\mathfrak{A},\emptyset) \mid \mathfrak{A} \in [\![\Phi]\!]_{(<\omega)}\}$.

Then $\mathrm{HomClosed}_{(\mathrm{fin})}$ *for* $\mathbb{L}$ *is* C*-hard under* C'*-reductions.*

Proof. We C'-reduce (finite) unsatisfiability in $\mathbb{L}'$ to $\mathrm{HomClosed}_{(\mathrm{fin})}$ for $\mathbb{L}$ as follows: given Φ, compute $f(\Phi)$.

If the set of (finite) models of $f(\Phi)$ is closed under (finite-target) homomorphisms then $[\![f(\Phi)]\!]_{(<\omega)}|_{\tau\cup\{\mathsf{U}\}}$ is closed under (finite-target) homomorphisms and therefore an empty set. This implies that Φ is (finitely) unsatisfiable. On the other hand if Φ is (finitely) unsatisfiable then $[\![f(\Phi)]\!]_{(<\omega)}|_{\tau\cup\{\mathsf{U}\}}$ as well as $[\![f(\Phi)]\!]_{(<\omega)}$ are closed under (finite-target) homomorphisms since they are empty. $\square$

Lemma 12. *Let* $\mathbb{L}, \mathbb{L}' \subseteq \mathbb{SO}$ *and* $f\colon \mathbb{L}' \to \mathbb{L}$ *such that*
- *(finite) unsatisfiability in* $\mathbb{L}'$ *is* C*-hard and*
- f *is* C'*-computable and maps* τ*-sentences to* $\tau\cup\{\mathsf{U}\}$*-senten-ces with* $[\![f(\Phi)]\!]_{(<\omega)} = \{(\mathfrak{A},\tilde{A}) \mid \mathfrak{A} \in [\![\neg\Phi]\!]_{(<\omega)}$ *or* $\tilde{A} \neq A\}$.

Then $\mathrm{HomClosed}_{(\mathrm{fin})}$ *for* $\mathbb{L}$ *is* C*-hard under* C'*-reductions.*

Proof. We C'-reduce (finite) unsatisfiability in $\mathbb{L}'$ to $\mathrm{HomClosed}_{(\mathrm{fin})}$ for $\mathbb{L}$ as follows: given Φ, compute $f(\Phi)$.

If Φ is (finitely) unsatisfiable, then $f(\Phi)$ is (finitely) valid and therefore its (finite) models are closed under (finite target) homomorphisms.

If Φ is (finitely) satisfiable witnessed by some (finite) model $\mathfrak{A}$, this means that $(\mathfrak{A}, A)$ is not a model of $f(\Phi)$. Let now $\mathfrak{B} = \mathfrak{A} \cdot \{\breve{\mathsf{U}}\}$, which clearly is a (finite) model of $f(\Phi)$. Moreover, the identity function is a homomorphism from $\mathfrak{B}$ to $(\mathfrak{A}, A)$, which shows that $f(\Phi)$ is not (finitely) homomorphism closed.

It follows that checking (finite) homomorphism-closedness of $f(\Phi)$ is equivalent to checking (finite) unsatisfiability of Φ. $\square$

Proposition 14. *For any two* τ*-structures* $\mathfrak{A}$ *and* $\mathfrak{B}$*, any homo-morphism* $h\colon \mathfrak{A} \to \mathfrak{B}$ *can be represented as the composition* $h_2 \circ h_1$ *of two homomorphisms where*
- $h_1\colon \mathfrak{A} \hookrightarrow \mathfrak{C}$ *is injective and*
- $h_2\colon \mathfrak{C} \twoheadrightarrow \mathfrak{B}$ *is surjective and strong.*

The same holds when consideration is restricted exclusively to (homomorphisms between) finite structures.

Proof. Given $\mathfrak{A}$, $\mathfrak{B}$, and h, we obtain $\mathfrak{C}$, h_1, and h_2 as follows:

- $C = A \uplus B$
- $h_1\colon A \hookrightarrow C$ is set to id_A

- $h_2 \colon C \Rrightarrow B$ is obtained via $h_2 = h \uplus id_B$
- $\mathsf{c}^{\mathfrak{C}} = \mathsf{c}^{\mathfrak{A}}$ for every constant $\mathsf{c} \in \tau$
- $\mathsf{P}^{\mathfrak{C}} = \{(c_1, \ldots, c_k) \mid (h_2(c_1), \ldots, h_2(c_k)) \in \mathsf{P}^{\mathfrak{B}}\}$ for every k-ary predicate $\mathsf{P} \in \tau$. $\qquad\square$

Lemma 18. *Let $\mathfrak{B}$ and $\mathfrak{A}$ be τ-structures, let $h \colon \mathfrak{B} \Rrightarrow \mathfrak{A}$ be a strong surjective homomorphism. Let n be a natural number. Then there exists a function $\lambda \colon A \to \{1, \ldots, n\}$ such that*
- *there is a strong surjective homomorphism $h' \colon \mathfrak{B} \Rrightarrow \mathfrak{A}^\lambda$ satisfying $h = \pi \circ h'$, and*
- *$\mathfrak{B}$ and $\mathfrak{A}^\lambda$ satisfy the same $\mathbb{FO}_=$ sentences of quantifier rank n.*

Proof. Let λ be defined by $a \mapsto \min(n, |h^{-1}(a)|)$ where $h^{-1}(a)$, as usual, denotes the set $\{b \in B \mid h(b) = a\}$. Now define $h' \colon B \to A^\lambda$ such that the elements of $h^{-1}(a)$ are mapped to $\{(a, 1), \ldots, (a, \lambda(a))\}$ in a surjective way, which is possible by construction. By definition of $\mathfrak{A}^\lambda$, the map h' is a strong surjective homomorphism and clearly satisfies $h = \pi \circ h'$. The equivalence of $\mathfrak{B}$ and $\mathfrak{A}^\lambda$ regarding $\mathbb{FO}_=$ sentences of quantifier rank n can be shown as usual via an n-step Ehrenfeucht-Fraïssé game, where, upon spoiler picking e, duplicator picks a fresh element from $h^{-1}(\pi(e))$ or $\pi^{-1}(h(e))$, respectively. $\qquad\square$

Theorem 32. $\mathrm{HomClosed}_{(\mathrm{fin})}$ *for*
1) $\mathbb{GNFO}_=$ *is* 2ExpTime-*complete,*
2) $\mathbb{TGF}$ *is* coN2ExpTime-*complete,*
3) *any of* $\mathbb{FO}^2_=$, $\forall^* \mathbb{FO}_=$, $\exists^* \mathbb{FO}_=$, $\forall\forall\exists\exists\mathbb{FO}$, *and* $\exists\exists\forall\forall\mathbb{FO}$ *is* coNExpTime-*complete.*

Proof. For all cases, membership can be shown for the finite cases directly via Theorem 31 by a polynomial reduction to a fragment with the appropriate unsatisfiability problem. The general cases can be reduced to the finite ones via Theorem 24.

Hardness follows for $\mathbb{GNFO}_=$, $\mathbb{TGF}$, $\mathbb{FO}^2_=$, and $\forall^* \mathbb{FO}_=$ as well as $\forall\forall\exists\exists\mathbb{FO}$ via Lemma 11 from the corresponding known unsatisfiability complexities. For $\exists^* \mathbb{FO}_=$ and $\exists\exists\forall\forall\mathbb{FO}$, it follows via Lemma 12 from the unsatisfiability complexities of $\forall^* \mathbb{FO}_=$ and $\forall\forall\exists\exists\mathbb{FO}$, respectively. $\qquad\square$

Theorem 33. $\mathrm{HomClosed}_{(\mathrm{fin})}$ *for* $\mathbb{TGD}$ *is* NP-*complete.*

Proof. For NP-hardness, we reduce the 3-colorability problem to $\mathrm{HomClosed}_{(\mathrm{fin})}$ as follows: given a finite graph expressed as $\{E\}$-structure $\mathfrak{G}$, three-colorability of $\mathfrak{G}$ is equivalent to the existence of a homomorphism from $\mathfrak{G}$ to $\mathfrak{K}_3$ (the complete but loop-free 3-element graph). However, this is the case exactly if the $\mathbb{TGD}$ sentence $\Phi = \mathrm{cq}(\mathfrak{K}_3) \Rightarrow \mathrm{cq}(\mathfrak{G})$ is a tautology. This, in turn coincides with $\llbracket \Phi \rrbracket$ being closed under homomorphisms (as, for the negative case, we find that $\mathfrak{K}_3 \notin \llbracket \Phi \rrbracket$ receives a homomorphism from $\mathfrak{I} \in \llbracket \Phi \rrbracket$).

In order to show NP membership, we provide a corresponding nondeterministic polytime algorithm that answers the question. From now on, we consider a $\mathbb{TGD}$ sentence Φ as the finite set of conjuncts, each representing a single TGD (simply called "rule" in the following). We first

apply a PTime preprocessing as follows: One by one, go through all the rules $\forall \boldsymbol{x}.\varphi(\boldsymbol{x}) \Rightarrow \exists \boldsymbol{y}.\psi(\boldsymbol{x}, \boldsymbol{y})$ of Φ and do the following: Let $\sim$ be the smallest equivalence relation over all the atoms from ψ such that $\alpha \sim \alpha'$ whenever α and α' share an $\exists$-quantified variable from $\boldsymbol{y}$. For every $\sim$-equivalence class, form the conjunction over its atoms and $\exists$-quantify over all variables from $\boldsymbol{y}$ occurring therein. Let $\exists \boldsymbol{y_1}.\psi_1, \ldots, \exists \boldsymbol{y_n}.\psi_n$ be the formulae thus obtained. Now, replace $\forall \boldsymbol{x}.\varphi(\boldsymbol{x}) \Rightarrow \exists \boldsymbol{y}.\psi(\boldsymbol{x}, \boldsymbol{y})$ by the rules $\forall \boldsymbol{x}.\varphi(\boldsymbol{x}) \Rightarrow \exists \boldsymbol{y_1}.\psi_1(\boldsymbol{x}, \boldsymbol{y_1}), \ldots, \forall \boldsymbol{x}.\varphi(\boldsymbol{x}) \Rightarrow \exists \boldsymbol{y_n}.\psi_n(\boldsymbol{x}, \boldsymbol{y_n})$. Finally, the new rule set is deterministically partitioned into two sets: those where all variables present in ψ's atoms are $\exists$-quantified will be called *disconnected*, all other rules will be called *connected*. Obviously all this preprocessing can be done in deterministic polynomial time and produces a partitioned rule set of polynomial size in the original one. Moreover, the obtained rule set is semantically equivalent to the original one.

We proceed do describe a polysize certificate witnessing homomorphism-closedness of a preprocessed rule set. Our certificate contains the following:

- a *relevance partitioning* of the rule set into *self-redundant* and *self-irredundant* rules, where all connected rules (but possibly also some disconnected ones) are categorized as self-redundant. It is straight forward to check validity of such a partition in PTime.
- a *redundancy witness* for each of the self-redundant rules: a substitution $h \colon \boldsymbol{y} \to \boldsymbol{x}$ sending every atom inside ψ to an atom inside φ. Obviously redundancy witnesses can be verified in PTime.
- a *universal model derivation*: a forward-chaining derivation sequence starting from $\mathfrak{I}$ using only self-irredundant rules and applying each rule at most once, including the homomorphisms used for the rule applications. Let $\mathfrak{M}$ denote the universal model constructed in this derivation. Again such a guess is of polynomial size and allows for PTime verification.
- a *discharge witness* for each self-irredundant rule: a mapping $h \colon \boldsymbol{y} \to M$ sending every atom $\mathsf{R}(y_1, \ldots, y_n)$ inside ψ to a "semantic atom" inside $\mathfrak{M}$, i.e., $(h(y_1), \ldots, h(y_n)) \in \mathsf{R}^{\mathfrak{M}}$. Discharge witnesses can be verified in PTime.

Summing up, the algorithm's guess consists of one relevance partitioning, one redundancy witness for every self-redundant rule, one universal model derivation, and one discharge witness per self-irredundant rule. Verification for each can be done in PTime.

We now argue that the algorithm is correct (i.e., that the existence of a certificate as above coincides with the TGD set's homomorphism-closedness):

i) First, consider the case where one of the connected rules is not self-redundant (i.e., no appropriate witness can be provided). For such a rule $\forall \boldsymbol{x}.\varphi(\boldsymbol{x}) \Rightarrow \exists \boldsymbol{y}.\psi(\boldsymbol{x}, \boldsymbol{y})$, let $\mathfrak{A}_\varphi$ be such that $\mathrm{cq}(\mathfrak{A}_\varphi) = \exists \boldsymbol{x}.\varphi(\boldsymbol{x})$. Then, $\mathfrak{F}$ is a model, but

$\mathfrak{F} \uplus \mathfrak{A}_\varphi$ is not – due to the special form of ψ that we achieved in the preprocessing. This is despite the existence of a homomorphism from $\mathfrak{F}$. Thus the rule set is not homclosed, as required. In the following, we consider the case where all connected rules are indeed (self-)redundant.

ii) Obviously, self-redundant rules are tautologies and hence have no influence on the set of models, so the question if the rule set is homomorphism-closed reduces to the question if its subset of self-irredundant rules is. By now we know that, in the considered rule set, every such rule $\forall \boldsymbol{x}.\varphi(\boldsymbol{x}) \Rightarrow \exists \boldsymbol{y}.\psi(\boldsymbol{x}, \boldsymbol{y})$ must be disconnected, i.e., be such that no variables from $\boldsymbol{x}$ occur in ψ, therefore any such rule can be written as $(\exists \boldsymbol{x}.\varphi(\boldsymbol{x})) \Rightarrow (\exists \boldsymbol{y}.\psi(\boldsymbol{y}))$. Henceforth, we assume our considered rule set contains the self-irredundant in this syntactic form. From the structure of the rules follows that there exists a finite universal model (i.e. one which has a homomorphism into every other model) $\mathfrak{M}$ which can be obtained in a forward-chaining derivation where every rule is applied at most once. This information is provided by the certificate's universal model derivation.

iii) We proceed to consider two cases.

a) Assume the non-existence of a discharge witness of $(\exists \boldsymbol{x}.\varphi(\boldsymbol{x})) \Rightarrow (\exists \boldsymbol{y}.\psi(\boldsymbol{y}))$. This means that among the remaining rules, there is one $(\exists \boldsymbol{x}.\varphi(\boldsymbol{x})) \Rightarrow (\exists \boldsymbol{y}.\psi(\boldsymbol{y}))$ for which $\mathfrak{M} \not\models \exists \boldsymbol{y}.\psi(\boldsymbol{y})$. Yet in that case $\mathfrak{M} \uplus \mathfrak{A}_\varphi$ is not a model (mark that the rule is not self-redundant by assumption), despite the existence of the trivial homomorphism from $\mathfrak{M}$ – consequently the rule set is not homclosed.

b) The existence of all discharge witnesses proves that each remaining rule's head is satisfied in $\mathfrak{M}$. Then, the same holds for every structure $\mathfrak{M}'$ into which a homomorphism from $\mathfrak{M}$ exists, which trivially entails that all remaining rules are satisfied in any such $\mathfrak{M}'$ and hence the rule set is homclosed. $\qquad \square$

F. Proofs of Characterizing Homclosures Section

Proposition 37. $[\![\Phi_\infty]\!]^{\rightarrow}$ *cannot be characterized within* $\mathbb{FO}_=$.

Proof. We note that $\mathfrak{A} \in [\![\Phi_\infty]\!]^{\rightarrow}$ exactly if $\mathfrak{A}$ contains an infinite P-path. Hence, if $\mathfrak{A}$ is finite, $\mathfrak{A} \in [\![\Phi_\infty]\!]^{\rightarrow}$ exactly if $\mathfrak{A}$ contains a directed P-cycle. Checking for the existence of directed cycles is NLOGSPACE-complete [18] whereas $\mathbb{FO}_=$ sentences can only describe problems in AC^0 [17], which is strictly weaker than NLOGSPACE [10]. $\qquad \square$

Theorem 38. $\mathrm{ExCharHomCl}^{\mathbb{L}}_{(<\omega)}$ *is undecidable for* $\mathbb{FO}_=$ *sentences whenever* $\mathbb{L}$ *is a non-degenerate logic with a decidable model checking problem.*

Proof. We show the claim by a reduction from (finite) satisfiability. Let Ψ be an arbitrary FO formula.

Consider Ψ' defined as $\Psi^{\mathrm{rel}(\mathtt{U})} \wedge \Phi_{\mathrm{grid}}$ chosen in a way that the signature $\tau_1 \uplus \{\mathtt{U}\}$ of $\Psi^{\mathrm{rel}(\mathtt{U})}$ and the signature τ_2 of Φ_{grid} are

disjoint. If Ψ is (finitely) unsatisfiable, then so is Ψ'. In that case $[\![\Psi']\!]^{\rightarrow(<\omega)}_{(<\omega)}$ is empty and can be characterized by $\Phi_{\boldsymbol{false}}$. If Ψ is (finitely) satisfiable, then any sentence $\tilde{\Psi}$ (finitely) characterizing $[\![\Psi']\!]^{\rightarrow(<\omega)}_{(<\omega)}$ can be turned into a procedure for deciding if for a domino system $\mathcal{D}$ a(n ultimately periodic) $\mathcal{D}$-tiling exists: this is the case exactly if the structure $\mathfrak{A}_{\mathcal{D}} \cdot \widehat{\tau_1} \cdot \widehat{\{\mathtt{U}\}}$ is a model of $\tilde{\Psi}$.

It follows that the (finite) homomorphism closure of $[\![\Psi']\!]_{(<\omega)}$ is characterizable (in any non-degenerate logic $\mathbb{L}$) exactly if Ψ is unsatisfiable. Checking the latter is undecidable. $\qquad \square$

Theorem 39. $\mathrm{ExCharHomCl}^{\mathbb{L}}_{(<\omega)}$ *is undecidable for* $\mathbb{TGD}$ *sentences whenever* $\mathbb{L}$ *is a logic able to express* $\mathrm{cq}(\mathfrak{F})$ *and having a decidable model checking problem.*

Proof. We show the claim by a reduction from the existence of a(n ultimately periodic) tilling by the deterministic domino set $\mathcal{D}$.

Let the TGD sentence $\Phi_{\mathcal{D}\text{-tiling}}$ with signature τ be defined as in Definition 34. Let

$$\Phi'_{\mathcal{D}\text{-tiling}} = \Phi_{\mathcal{D}\text{-tiling}} \wedge \Big((\exists x.\mathtt{T}_\emptyset(x)) \Rightarrow \mathrm{cq}(\mathfrak{F}) \Big). \qquad (61)$$

If no (ultimately periodic) $\mathcal{D}$-tiling exists, then $\mathrm{cq}(\mathfrak{F})$ characterizes $[\![\Phi'_{\mathcal{D}\text{-tiling}}]\!]^{\rightarrow(<\omega)}_{(<\omega)}$.

If, on the other hand, some (ultimately periodic) $\mathcal{D}$-tiling exists, then any sentence Ψ (finitely) characterizing $[\![\Phi'_{\mathcal{D}\text{-tiling}}]\!]^{\rightarrow(<\omega)}_{(<\omega)}$ can be turned into a procedure for deciding if for some other domino system $\mathcal{D}'$ a(n ultimately periodic) $\mathcal{D}'$-tiling exists: this is the case exactly if the structure $\mathfrak{A}_{\mathcal{D}'} \cdot \widehat{\sigma}$ with $\sigma = \tau \setminus \{\mathtt{H}, \mathtt{V}\}$ is a model of Ψ. This holds since all $\{\mathtt{H}, \mathtt{V}\}$-reducts of models of $\Phi'_{\mathcal{D}\text{-tiling}}$ satisfy Φ_{grid} and furthermore if an ultimately periodic $\mathcal{D}'$-tiling exists we can find a model of $\Phi'_{\mathcal{D}\text{-tiling}}$ such that the period of the grid induced by its $\{\mathtt{H}, \mathtt{V}\}$-reduct is a multiple of the period of the $\mathcal{D}'$-tiling. This is possible since a domino set that tiles the grid with an ultimately periodic tiling with period length k also tiles it with an ultimately periodic tiling with period length $k \cdot s$ for all $s \in \mathbb{N}$.

It follows that the (finite) homomorphism closure of $[\![\Phi'_{\mathcal{D}\text{-tiling}}]\!]_{(<\omega)}$ is characterizable (in any decent logic $\mathbb{L}$ in the above sense) exactly if some (ultimately periodic) $\mathcal{D}$-tiling exists. Checking the latter is undecidable. $\qquad \square$

Proposition 40. *Every* $\exists^*\forall^*\mathbb{FO}_=$ *sentence* Φ *satisfies* $[\![\Phi]\!]^{\rightarrow} = [\![\Phi]\!]^{\rightarrow}_{<\omega}$ *and one can effectively compute a* $\exists^*\mathbb{FO}^\pm_=$ *sentence* Ψ *with* $[\![\Phi]\!]^{\rightarrow} = [\![\Psi]\!]$.

Proof. Assume Φ is over the signature τ. Let Φ^{sk} be the $\forall^*\mathbb{FO}_=$ sentence obtained from Φ via skolemization of all the existentially quantified variables into constants. Clearly then $[\![\Phi]\!]_{(<\omega)} = [\![\Phi^{\mathrm{sk}}]\!]_{(<\omega)}|_\tau$. Since Φ^{sk} is a $\forall^*\mathbb{FO}_=$ sentence, any model $\mathfrak{A}$ gives rise to a "small" submodel by taking the substructure $\mathfrak{A}'$ induced by the set $A' \subseteq A$ containing all elements denoted by constants (if the signature has no constants, pick one arbitrary element). Moreover, the identity

on A' is a homomorphism from $\mathfrak{A}'$ to $\mathfrak{A}$. Let $\mathcal{A}$ be the set of small submodels thus obtained, for which therefore holds $\mathcal{A}^{\rightarrow} = [\![\Phi^{\mathrm{sk}}]\!]^{\rightarrow}$ and hence also $\mathcal{A}|_{\tau}^{\rightarrow} = [\![\Phi]\!]^{\rightarrow}$. Due to the bound on the domain size (number of constants), $\mathcal{A}$ (and thus $\mathcal{A}|_{\tau}$) can contain only finitely many different structures (up to isomorphism) and is effectively computable via enumeration and model checking. Hence $\mathcal{A}|_{\tau}^{\rightarrow}$ can be characterized by the $\exists^*\mathbb{FO}_=$ sentence $\Psi = \bigvee_{\mathfrak{A} \in \mathcal{A}|_{\tau}} \mathrm{cq}(\mathfrak{A})$. $\qquad\square$

Theorem 43. *1)* $\mathbb{FO}_=^2$ *homcaptures itself.*

2) $\mathbb{TGF}$ *homcaptures itself and* $\exists^*\forall\forall\exists^*\mathbb{FO}$*.*

3) $\mathbb{GFO}_=$ *homcaptures itself and* $\mathbb{GNFO}_=$*.*

We split the proof into separate statements and prove them individually:

- Proposition 69 deals with homcapturing of $\mathbb{FO}_=^2$
- Proposition 84 deals with homcapturing of $\mathbb{GFO}_=$
- Proposition 86 deals with homcapturing of $\mathbb{GNFO}_=$
- Proposition 87 deals with homcapturing of $\mathbb{TGF}$
- Proposition 88 deals with homcapturing of $\exists^*\forall\forall\exists^*\mathbb{FO}$

Before, however, we will show a useful result, which allows us to – without loss of generality – restrict our attention to (projective) normal forms of the considered logics (as they are typically used in results related to satisfiability).

Lemma 65 (projective normal form sufficiency)**.** *For some* $\tau \subseteq \tau' \subseteq \tau''$, *let*

- Φ *be a* τ*-sentence,*
- Φ' *be a* τ'*-sentence projectively characterizing* Φ*, i.e.,* $[\![\Phi']\!]|_{\tau} = [\![\Phi]\!]$*,*
- Ψ *be a* τ''*-sentence projectively characterizing the hom-closure of* Φ'*, i.e.,* $[\![\Phi']\!]^{\rightarrow} = [\![\Psi]\!]|_{\tau'}$*.*

Then Ψ *also projectively characterizes the homclosure of* Φ*, i.e.,* $[\![\Phi]\!]^{\rightarrow} = [\![\Psi]\!]|_{\tau}$ *holds.*

Proof. Since $\tau \subseteq \tau'$, we obtain

$$[\![\Psi]\!]|_{\tau} = [\![\Psi]\!]|_{\tau'}|_{\tau} = [\![\Phi']\!]^{\rightarrow}|_{\tau} \overset{*}{=} [\![\Phi']\!]|_{\tau}^{\rightarrow} = [\![\Phi]\!]^{\rightarrow}. \quad (62)$$

For the equality denoted by $\overset{*}{=}$, we show the two inclusions separately:

"$\subseteq$": Let $\mathfrak{B} \in [\![\Phi']\!]^{\rightarrow}|_{\tau}$. Then there exists a τ'-structure $\mathfrak{B}' \in [\![\Phi']\!]^{\rightarrow}$ with $\mathfrak{B}'|_{\tau} = \mathfrak{B}$. Hence there exist a τ'-structure $\mathfrak{A}' \in [\![\Phi']\!]$ and a homomorphism $h\colon \mathfrak{A}' \to \mathfrak{B}'$. But then h also serves as a homomorphism from $\mathfrak{A} = \mathfrak{A}'|_{\tau}$ to $\mathfrak{B}'|_{\tau} = \mathfrak{B}$. By $\mathfrak{A} \in [\![\Phi']\!]|_{\tau}$ we then obtain $\mathfrak{B} \in [\![\Phi']\!]|_{\tau}^{\rightarrow}$.

"$\supseteq$": Let $\mathfrak{B} \in [\![\Phi']\!]|_{\tau}^{\rightarrow}$. Then there exist a τ'-structure $\mathfrak{A}' \in [\![\Phi']\!]$ and a homomorphism $h\colon \mathfrak{A}'|_{\tau} \to \mathfrak{B}$. Let now $\mathfrak{B}' = \mathfrak{B} \cdot \widehat{\tau'\backslash\tau}$. Clearly, then h serves as homomorphism from $\mathfrak{A}'$ to $\mathfrak{B}'$, thus $\mathfrak{B}' \in [\![\Phi']\!]^{\rightarrow}$. Yet, since $\mathfrak{B}'|_{\tau} = \mathfrak{B}$, we obtain $\mathfrak{B} \in [\![\Phi']\!]^{\rightarrow}|_{\tau}$. $\qquad\square$

As the next three results (namely Proposition 69, Proposition 84, and Proposition 87) will make use of some essentially

equal parts in the homcapturing sentence we introduce these "generic" parts beforehand.

Definition 66. *Let* Φ *be a* $\mathbb{FO}$ *sentence over some fixed signature* τ *and let* C *denote the set of all constant symbols from* τ*. Additionally, we fix some map* width *mapping each* τ*-formula to some natural number, its so called* width*. We specify this map when the need arises.*

A type t *is a maximal consistent set of* τ*-literals (i.e., negated and unnegated atoms where the relation symbol is from* τ*) over terms from* $\boldsymbol{v} \cup C$*, where* $\boldsymbol{v}$ *is a dedicated, finite, linearly ordered set of* type variables *of cardinality* $\geq$ width(Φ)*. When necessary we might explicitly mention the signature at consideration, i.e. we say* τ*-type. To each type* t *we associate a* $\boldsymbol{v}$*-subsequence* $\boldsymbol{v}_{\mathsf{t}}$ *enumerating the variables of* t*. We say* t *is an* n*-type, or equivalently is a type of* order n*, if* $\boldsymbol{v}_{\mathsf{t}}$ *is a sequence of exactly* n *variables. We define* order *to be a map from the set of types to the natural numbers, associating to each type its order, i.e., an* n*-type* t *has* order$(\mathsf{t}) = n$*. For a structure* $\mathfrak{A}$ *(with domain* A*) we say that an* n*-type* t *is* realized *by an* n*-tuple* $\boldsymbol{a} = (a_1, \ldots, a_n) \in A^n$*, if for the componentwise evaluation map* $\nu\colon \boldsymbol{v}_{\mathsf{t}} \mapsto \boldsymbol{a}, (\boldsymbol{v}_{\mathsf{t}})_i \mapsto a_i$ *the relation* $(\mathfrak{A}, \nu) \models \mathsf{t}$ *holds. We say a type* t *of order* n *is* realized *(*k *times) in* $\mathfrak{A}$ *if it is realized by (*k *distinct)* n*-tuples from* $\mathfrak{A}$*.*

Given an atom δ *using exactly the variables from* $\boldsymbol{x}$ *(and possibly additional constants), we say a type* t *is* guarded by δ *if there exists a variable substitution* $\nu\colon \boldsymbol{x} \to \boldsymbol{v}_{\mathsf{t}} \cup C$ *with* $\boldsymbol{v}_{\mathsf{t}} \subseteq \{\nu(x) \mid x \in \boldsymbol{x}\}$ *such that* $\nu(\delta) \in \mathsf{t}$*, where the application of a substitution* ν *to an expression replaces all occurrences of every variable* x *in* ν*'s domain by* $\nu(x)$*. Then* ν *will be called* guard witness*. A type* t *is* guarded*, if it is guarded by some atom. A type* t *is called* rigid *if it does not contain literals of the form* $v = v'$ *(for distinct variables* v, v' *from* $\boldsymbol{v}_{\mathsf{t}}$*) or* $\mathsf{c} = v$ *or* $v = \mathsf{c}$ *for* $\mathsf{c} \in C$*. For* $\boldsymbol{v}' \subseteq \boldsymbol{v}_{\mathsf{t}}$*, we obtain the* $\boldsymbol{v}'$*-subtype of* t*, denoted* $\mathsf{t}|_{\boldsymbol{v}'}$*, by removing from* t *all literals mentioning any variable from* $\boldsymbol{v}_{\mathsf{t}} \setminus \boldsymbol{v}'$*.*

By $\mathcal{E}$ *we will denote a fixed finite set of rigid types and call it the set of* eligible types*. For a* τ*-structure* $\mathfrak{A}$ *we denote by* $\mathcal{E}(\mathfrak{A})$ *its type summary, which is a pair* $(\mathcal{E}_+, \mathcal{E}_!)$ *of subsets of* $\mathcal{E}$ *such that* $\mathcal{E}_+$ *(*$\mathcal{E}_!$*) contains every type realized (exactly once) in* $\mathfrak{A}$*. Furthermore, the* model summary $(\!|\Phi|\!)$ *of a sentence* Φ *is defined by* $\{\mathcal{E}(\mathfrak{A}) \mid \mathfrak{A} \in [\![\Phi]\!]\}$*.*

For a set $\mathcal{E}$ *of eligible types we extend the signature* τ *by a set* σ *of* type predicates*:* σ *contains a new* n*-ary predicate* $\mathtt{Tp}_{\mathsf{t}}$ *for every eligible* n*-type* $\mathsf{t} \in \mathcal{E}$*. Let* $\mathfrak{A}$ *be a* τ*-structure. The* $\mathcal{E}$*-adornment of* $\mathfrak{A}$*, denoted by* $\mathfrak{A} \cdot \mathcal{E}$ *is the* $\tau \uplus \sigma$*-structure expanding* $\mathfrak{A}$ *such that* $\mathtt{Tp}_{\mathsf{t}}^{\mathfrak{A} \cdot \mathcal{E}} = \{\boldsymbol{a} \mid \boldsymbol{a} \text{ realizes } \mathsf{t} \text{ in } \mathfrak{A}\}$*.*

If $\mathfrak{A}, \mathfrak{B}$ *are* τ*-structures such that there is a homomorphism* $h\colon \mathfrak{A} \to \mathfrak{B}$ *and we have a set of eligible types* $\mathcal{E}$*, then we denote by* $\mathfrak{B} \cdot h(\mathfrak{A} \cdot \mathcal{E}|_{\sigma})$ *the* $\tau \uplus \sigma$*-structure* $\mathfrak{B}'$ *expanding* $\mathfrak{B}$ *with* $\mathtt{Tp}_{\mathsf{t}}^{\mathfrak{B}'} = \{h(\boldsymbol{a}) \mid \boldsymbol{a} \in \mathtt{Tp}_{\mathsf{t}}^{\mathfrak{A} \cdot \mathcal{E}}\}$*.*

We define

$$\Psi_{\mathrm{hom}} = \bigwedge_{\mathfrak{t}\in\mathcal{E}} \forall \boldsymbol{v}_{\mathfrak{t}}.\big(\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Rightarrow \bigwedge\{\alpha\in\mathfrak{t} \mid \alpha \text{ positive lit.}\}\big). \quad (63)$$

For every $(\mathcal{E}_+,\mathcal{E}_!) \in (\![\Phi]\!)$ let $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ be the conjunction of the following formulae:

$$\exists \boldsymbol{v}_{\mathfrak{t}}.\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \qquad\qquad\qquad\quad \text{for } \mathfrak{t}\in\mathcal{E}_+ \quad (64)$$
$$\forall \boldsymbol{v}_{\mathfrak{t}}.\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Rightarrow \bot \qquad\qquad\quad \text{for } \mathfrak{t}\in\mathcal{E}\setminus\mathcal{E}_+ \quad (65)$$
$$\forall \boldsymbol{v}_{\mathfrak{t}}.\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Rightarrow \mathrm{Tp}_{\mathfrak{t}'}(\boldsymbol{v}_{\mathfrak{t}'}) \quad \text{for } \mathfrak{t},\mathfrak{t}'\in\mathcal{E}_+, \mathfrak{t}'=\mathfrak{t}|_{\boldsymbol{v}_{\mathfrak{t}'}} \quad (66)$$
$$\forall \boldsymbol{x}.\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{x}) \Rightarrow \mathrm{Tp}_{\kappa\mathfrak{t}}(\eta_{\boldsymbol{v}}^{\kappa\boldsymbol{v}_{\mathfrak{t}}}(\boldsymbol{x})) \quad \text{for } \mathfrak{t}\in\mathcal{E}, \kappa:\boldsymbol{v}\hookrightarrow\boldsymbol{v} \quad (67)$$

where

- *the permutation κ on the set $\boldsymbol{v}$ is extended to variable tuples in the usual way and also used as substitution on logical expressions,*

- *for a (repetition-free) tuple $\boldsymbol{v}'$ of variables from $\boldsymbol{v}$, we let $\eta_{\boldsymbol{v}}^{\boldsymbol{v}'}$ denote the sorting function rearranging the entries of $|\boldsymbol{v}'|$-tuples in a way that $\eta_{\boldsymbol{v}}^{\boldsymbol{v}'}(\boldsymbol{v}')$ becomes a subsequence of $\boldsymbol{v}$.*

Several remarks to this definition:

1) Note that each signature τ is finite (hence also the set of constant symbols C from τ) and the set of dedicated type variables $\boldsymbol{v}$ is finite as well. Hence there are only finitely many distinct literals with terms from $\boldsymbol{v}\cup C$. This implies the finiteness of every type $\mathfrak{t}$. For convenience we may identify $\mathfrak{t}$ (which is a set of literals) with the conjunction over all the literals contained in it, i.e. $\bigwedge\mathfrak{t}$ (which then is a formula).

2) Given a τ-structure $\mathfrak{A}$ and a tuple $\boldsymbol{a}=(a_1,\ldots,a_n)$ (such that $\mathrm{length}(\boldsymbol{a})\leq\mathrm{length}(\boldsymbol{v})$), $\boldsymbol{a}$ realizes one, up to permutation of type variables, unique type $\mathfrak{t}$ characterizing the induced substructure of $\mathfrak{A}$ by the elements from $\boldsymbol{a}$ and the interpretations of every constant symbol. To be more precise, let $\boldsymbol{v}_{\mathfrak{t}}=(v_1,\ldots,v_n)$ be the initial segment of $\boldsymbol{v}$ of length $\mathrm{length}(\boldsymbol{a})$ and $\mu:\boldsymbol{v}_{\mathfrak{t}}\to\boldsymbol{a}, v_i\mapsto a_i$ the componentwise evaluation of $\boldsymbol{v}_{\mathfrak{t}}$ by $\boldsymbol{a}$. We extend μ to $\boldsymbol{v}_{\mathfrak{t}}\cup C$ by defining $\mu(\mathtt{c})=\mathtt{c}^{\mathfrak{A}}$ for every $\mathtt{c}\in C$. For each R and $\boldsymbol{t}=(t_1,\ldots,t_{\mathrm{ar}(\mathrm{R})})\subseteq\boldsymbol{v}_{\mathfrak{t}}\cup C$ the literal
   - $\mathrm{R}(\boldsymbol{t})$ is in $\mathfrak{t}$, iff $\mu(\boldsymbol{t})\in\mathrm{R}^{\mathfrak{A}}$ and
   - $\neg\mathrm{R}(\boldsymbol{t})$ is in $\mathfrak{t}$, iff $\mu(\boldsymbol{t})\notin\mathrm{R}^{\mathfrak{A}}$.

   This set is consistent as (by $\mathfrak{A}$ being a structure) for $\boldsymbol{t}$ we cannot have $\mu(\boldsymbol{t})\in\mathrm{R}^{\mathfrak{A}}$ and $\mu(\boldsymbol{t})\notin\mathrm{R}^{\mathfrak{A}}$ simultaneously. Furthermore it is maximal since either $\mu(\boldsymbol{t})\in\mathrm{R}^{\mathfrak{A}}$ or $\mu(\boldsymbol{t})\notin\mathrm{R}^{\mathfrak{A}}$. To show uniqueness (up to variable perrmutations), assume there were a second type $\mathfrak{t}^*$ also realized by $\boldsymbol{a}$ such that for all $\kappa:\boldsymbol{v}\hookrightarrow\boldsymbol{v}$ the inequality $\kappa\mathfrak{t}^*\neq\mathfrak{t}$ holds. Without loss of generality we may assume that $\boldsymbol{v}_{\mathfrak{t}}=\boldsymbol{v}_{\mathfrak{t}^*}=(v_1,\ldots,v_n)$ and κ mapping each $v\in\boldsymbol{v}_{\mathfrak{t}}$ to itself (and hence $\kappa\mathfrak{t}^*=\mathfrak{t}^*$). As $\mathfrak{t}\neq\mathfrak{t}^*$ there is some $\mathrm{R}\in\tau$ and $\boldsymbol{t}(t_1,\ldots,t_{\mathrm{ar}(\mathrm{R})})\subseteq\boldsymbol{v}_{\mathfrak{t}}\cup C$ such that $\mathrm{R}(\boldsymbol{t})\in\mathfrak{t}$ and $\mathrm{R}(\boldsymbol{t})\notin\mathfrak{t}^*$. By maximal consistency of types this

means $\neg\mathrm{R}(\boldsymbol{t})\in\mathfrak{t}^*$. But as $\boldsymbol{a}$ realizes both types we obtain $\mu(\boldsymbol{t})\in\mathrm{R}^{\mathfrak{A}}$ and $\mu(\boldsymbol{t})\notin\mathrm{R}^{\mathfrak{A}}$, a contradiction.

3) A note on eligible types: Depending on the logical fragment under scrutiny, we will make the notion of eligibility more specific. In fact, for Φ a sentence from some fragment of $\mathbb{FO}$, the finiteness will be ensured by considering just types whose order is bounded by $\mathrm{width}(\Phi)$.

4) As a corollary to sets $\mathcal{E}$ being finite, we immediately obtain that for an $\mathbb{FO}$ sentence Φ (and a set of eligible types $\mathcal{E}$) its model summary $(\![\Phi]\!)$ is finite.

Lemma 67. *Let τ be a signature and Φ a $\mathbb{FO}$ τ-sentence. Let $\mathcal{E}$ be a finite set of rigid types.*

1) *For every $\mathfrak{t},\mathfrak{t}' \in \mathcal{E}$ and every $\mathfrak{A} \in [\![\Phi]\!]$ with $\mathcal{E}(\mathfrak{A}) = (\mathcal{E}_+,\mathcal{E}_!) \in (\![\Phi]\!)$ with $\mathfrak{t}'=\mathfrak{t}|_{\boldsymbol{v}_{\mathfrak{t}'}}$ and $\mathfrak{t}\in\mathcal{E}_+$, then also $\mathfrak{t}' \in \mathcal{E}_+$. More specifically: Every componentwise assignment $\mu:\boldsymbol{v}_{\mathfrak{t}}\to\boldsymbol{a}$ of the variables from $\mathfrak{t}$ to a tuple $\boldsymbol{a}$ from $\mathfrak{A}$ of length $\mathrm{order}(\mathfrak{t})$ witnessing the realization of $\mathfrak{t}$ by $\boldsymbol{a}$ in $\mathfrak{A}$ gives rise to a subtuple $\boldsymbol{a}'$ of $\boldsymbol{a}$, namely $\mu(\boldsymbol{v}_{\mathfrak{t}'})$, such that $\boldsymbol{a}'=\mu|_{\boldsymbol{v}_{\mathfrak{t}'}}$ witnesses the realization of $\mathfrak{t}'$ by $\boldsymbol{a}'$ in $\mathfrak{A}$.*

2) *If $\mathcal{E}$ is closed under taking variable permutations, i.e., for every $\mathfrak{t}\in\mathcal{E}$ and $\kappa:\boldsymbol{v}\hookrightarrow\boldsymbol{v}$ also $\kappa\mathfrak{t}\in\mathcal{E}$, then so is $\mathcal{E}_+$ for every $(\mathcal{E}_+,\mathcal{E}_!)\in(\![\Phi]\!)$.*

Note that Item 1 essentially asserts for every $(\mathcal{E}_+,\mathcal{E}_!) \in (\![\Phi]\!)$ a kind of limited closure under taking subtypes in $\mathcal{E}_+$.

Proof. Let Φ be a $\mathbb{FO}$ τ-sentence and $\mathcal{E}$ a set of eligible types. We first show Item 1. Let $\mathfrak{t},\mathfrak{t}' \in \mathcal{E}$ with $\mathfrak{t}'=\mathfrak{t}|_{\boldsymbol{v}_{\mathfrak{t}'}}$, $\mathfrak{A}\in[\![\Phi]\!]$, $\mathcal{E}(\mathfrak{A})=(\mathcal{E}_+,\mathcal{E}_!)\in(\![\Phi]\!)$, and $\mathfrak{t}\in\mathcal{E}_+$. Hence $\mathfrak{t}$ is realized in $\mathfrak{A}$ by some tuple $\boldsymbol{a}=(a_1,\ldots,a_n)\in A^n$ where $n=\mathrm{order}(\mathfrak{t})$. Let $\mu:\boldsymbol{v}_{\mathfrak{t}}\to\boldsymbol{a}$ be the componentwise variable evaluation witnessing this realization, i.e. $(\mathfrak{A},\mu)\vDash\mathfrak{t}$. Now we set $\iota:\boldsymbol{v}_{\mathfrak{t}'}\to\boldsymbol{v}_{\mathfrak{t}}, v\mapsto v$. Then the map $\mu\circ\iota$ witnesses the realization of $\mathfrak{t}'=\mathfrak{t}|_{\boldsymbol{v}_{\mathfrak{t}'}}$ in $\mathfrak{A}$ by the subtuple $\mu(\boldsymbol{v}_{\mathfrak{t}'})$ (obtained by componentwise application of μ) of $\boldsymbol{a}$. As $\mathfrak{t}'\in\mathcal{E}$ we conclude $\mathfrak{t}'\in\mathcal{E}_+$.

Now take $\mathcal{E}$ to be closed under taking permutations of variables, let $\mathfrak{A}\in[\![\Phi]\!]$, and $\mathcal{E}(\mathfrak{A})=(\mathcal{E}_+,\mathcal{E}_!)\in(\![\Phi]\!)$. Furthermore take some type $\mathfrak{t}\in\mathcal{E}_+$ and some permutation $\kappa:\boldsymbol{v}\hookrightarrow\boldsymbol{v}$. Since $\mathfrak{t}\in\mathcal{E}_+$, it is realized by some tuple $\boldsymbol{a}=(a_1,\ldots,a_n)\in A^n$ where $n=\mathrm{order}(\mathfrak{t})$. We denote by $\mu:\boldsymbol{v}_{\mathfrak{t}}\to\boldsymbol{a}$ the componentwise variable assignment witnessing the realization of $\mathfrak{t}$ by $\boldsymbol{a}$ in $\mathfrak{A}$, i.e. $(\mathfrak{A},\mu)\vDash\mathfrak{t}$. Then the tuple $\eta_{\boldsymbol{v}}^{\kappa\boldsymbol{v}_{\mathfrak{t}}}(\boldsymbol{a})$ realizes $\kappa\mathfrak{t}$, witnessed by $\mu':\kappa\boldsymbol{v}_{\mathfrak{t}}\to\eta_{\boldsymbol{v}}^{\kappa\boldsymbol{v}_{\mathfrak{t}}}(\boldsymbol{a}), (\eta_{\boldsymbol{v}}^{\kappa\boldsymbol{v}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}))_i\mapsto(\eta_{\boldsymbol{v}}^{\kappa\boldsymbol{v}_{\mathfrak{t}}}(\boldsymbol{a}))_i$. As $\mathcal{E}$ is closed under taking permutations of variables, we conclude that $\kappa\mathfrak{t}\in\mathcal{E}_+$. This establishes Item 2. $\qquad\square$

Lemma 68. *Let Φ be a $\mathbb{FO}$ τ-sentence, $\mathcal{E}$ be a set of eligible types (i.e., a finite set of rigid type) closed under variable permutations, and σ be the set of fresh type predicates $\mathrm{Tp}_{\mathfrak{t}}$ for all $\mathfrak{t}\in\mathcal{E}$. Then for $\mathfrak{A}\in[\![\Phi]\!]$ the $\mathcal{E}$-adornment $\mathfrak{A}\cdot\mathcal{E}$ satisfies $\Psi_{\mathrm{gen},\mathcal{E}(\mathfrak{A})}$. Furthermore, for every $\mathfrak{B}$ being a homomorphic*

codomain of $\mathfrak{A}$ via a homomorphism $h\colon \mathfrak{A} \to \mathfrak{B}$ the $\tau \uplus \sigma$-structure $\mathfrak{B} \cdot h(\mathfrak{A} \cdot \mathcal{E}|_\sigma)$ equally satisfies $\Psi_{\mathrm{gen},\mathcal{E}(\mathfrak{A})}$.

Proof. Let Φ be a $\mathbb{FO}$ τ-sentence, $\mathfrak{A} \in \llbracket \Phi \rrbracket$ and let $\mathcal{E}$ be the set of eligible types and $\mathcal{E}(\mathfrak{A}) = (\mathcal{E}_+, \mathcal{E}_!)$ denote the type summary of $\mathfrak{A}$. We show that the $\mathcal{E}$-adornment $\mathfrak{A} \cdot \mathcal{E}$ of $\mathfrak{A}$ satisfies $\Psi_{\mathrm{gen},\mathcal{E}(\mathfrak{A})}$. More specifically:

For all $\mathsf{t} \in \mathcal{E}$ the interpretation $\mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$ contains by definition some tuple $\boldsymbol{a}$ from $\mathfrak{A}\cdot\mathcal{E}$ if and only if $\boldsymbol{a}$ realizes the associated type t in $\mathfrak{A}$. As every $\mathsf{t} \in \mathcal{E}_+$ is realized in $\mathfrak{A}$ we conclude that $\mathfrak{A} \cdot \mathcal{E}$ satisfies Formulae 64. As none of the types $\mathsf{t} \in \mathcal{E} \setminus \mathcal{E}_+$ is being realized in $\mathfrak{A}$, we see by the same token that the interpretations $\mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$ have to be empty. Hence we immediately obtain that $\mathfrak{A} \cdot \mathcal{E}$ satisfies Formulae 65.

Now take $\mathsf{t}, \mathsf{t}' \in \mathcal{E}$ such that $\mathsf{t}' = \mathsf{t}|_{v_{t'}}$ and let $\boldsymbol{a}$ be some tuple from $\mathfrak{A} \cdot \mathcal{E}$ such that $\boldsymbol{a} \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$. By definition $\boldsymbol{a}$ realizes t in $\mathfrak{A}$, let $\mu\colon \boldsymbol{v}_\mathsf{t} \to \boldsymbol{a}$ be the componentwise evaluation witnessing this. By Lemma 67 t' is realized in $\mathfrak{A}$ (i.e. $\mathsf{t}' \in \mathcal{E}_+$) by the subtuple $\mu(\boldsymbol{v}_{t'})$ (μ applied componentwise) of $\boldsymbol{a}$ realizing t'. Hence it is in $\mathrm{Tp}_{\mathsf{t}'}^{\mathfrak{A}\cdot\mathcal{E}}$ by definition. Finally, observing that μ assigns the tuple $\boldsymbol{a}$ componentwise to $\boldsymbol{v}_\mathsf{t}$, yields the satisfaction of Formulae 66 by $\mathfrak{A} \cdot \mathcal{E}$.

To show the satisfaction of the last kind of Formulae (i.e. Formulae 67), let $\mathsf{t} \in \mathcal{E}$, $\kappa\colon \boldsymbol{v} \hookrightarrow \boldsymbol{v}$ a permutation of the variables and $\boldsymbol{a}$ a tuple from $\mathfrak{A} \cdot \mathcal{E}$ such that $\boldsymbol{a} \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$. Note that, since $\mathcal{E}$ is closed under taking permutations of (type) variables, $\kappa\mathsf{t} \in \mathcal{E}$. As for every $\mathsf{t} \in \mathcal{E} \setminus \mathcal{E}_+$ the interpretation $\mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$ is empty, it suffices to discuss the case when $\mathsf{t} \in \mathcal{E}_+$. Hence $\boldsymbol{a}$ realizes t in $\mathfrak{A}$. It is plain that $\eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_\mathsf{t}}(\boldsymbol{a})$ realizes $\kappa\mathsf{t}$ in $\mathfrak{A}$, hence $\kappa\mathsf{t} \in \mathcal{E}_+$. Thus we obtain $\eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_\mathsf{t}}(\boldsymbol{a}) \in \mathrm{Tp}_{\kappa\mathsf{t}}^{\mathfrak{A}\cdot\mathcal{E}}$, showing that $\mathfrak{A} \cdot \mathcal{E}$ satisfies Formulae 67.

We now turn to the second part of the lemma. Let $\mathfrak{B}$ be a τ-structure such that there is a homomorphism $h\colon \mathfrak{A} \to \mathfrak{B}$. We show that the structure $\mathfrak{B} \cdot h(\mathfrak{A} \cdot \mathcal{E}|_\sigma)$ (which we will denote by $\mathfrak{B}'$) also satisfies $\Psi_{\mathrm{gen},\mathcal{E}(\mathfrak{A})}$. Additionally, $h\colon \mathfrak{A} \to \mathfrak{B}$ is also a homomorphism if lifted to $h'\colon \mathfrak{A} \cdot \mathcal{E} \to \mathfrak{B}'$. As h is already a homomorphism from $\mathfrak{A}$ to $\mathfrak{B}$ we only have to check whether for every $\mathrm{Tp}_\mathsf{t} \in \sigma$ and $\boldsymbol{a} \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$ the image of $\boldsymbol{a}$ under h' satisfies $h'(\boldsymbol{a}) \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{B}'}$. But by definition of $\mathfrak{B}'$ the interpretation of these type predicates is already defined to be the homomorphic image of the respective interpretations in $\mathfrak{A} \cdot \mathcal{E}$. Hence $h'(\boldsymbol{a}) \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{B}'}$ yielding that h' is also a homomorphism.

Applying this knowledge we immediately obtain that $\mathfrak{B}'$ satisfies Formulae 64: As $\mathfrak{A} \cdot \mathcal{E}$ satisfies 64 we find for every $\mathsf{t} \in \mathcal{E}_+$ a tuple $\boldsymbol{a}$ such that $\boldsymbol{a} \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$. As h' is a homomorphism we get $h'(\boldsymbol{a}) \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{B}'}$. To show satisfaction of Formulae 65 we note first, that by definition of $\mathfrak{B}'$ we interpreted the type predicates exactly as the homomorphic image of their interpretation in $\mathfrak{A}\cdot\mathcal{E}$. Hence for every $\mathsf{t} \in \mathcal{E}\setminus\mathcal{E}_+$ we obtain $\mathrm{Tp}_\mathsf{t}^{\mathfrak{B}'} = \emptyset$ as $\mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$ was already empty. This yields Formulae 65.

Now let $\mathsf{t}, \mathsf{t}' \in \mathcal{E}$ such that $\mathsf{t}' = \mathsf{t}|_{\boldsymbol{v}_{t'}}$ and let $\boldsymbol{b}$ be a tuple from $\mathfrak{B}'$ such that $\boldsymbol{b} \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{B}'}$. As $\boldsymbol{b}$ is in $\mathrm{Tp}_\mathsf{t}^{\mathfrak{B}'}$ we know by definition of $\mathfrak{B}'$ that $\boldsymbol{b}$ is the homomorphic image of some

tuple from $\mathfrak{A} \cdot \mathcal{E}$, more specifically we find a tuple $\boldsymbol{a}$ in $\mathfrak{A} \cdot \mathcal{E}$ such that $\boldsymbol{a} \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$ and $h'(\boldsymbol{a}) = \boldsymbol{b}$. Let $\mu\colon \boldsymbol{v}_\mathsf{t} \to \boldsymbol{a}$ be the componentwise evaluation witnessing the satisfaction of Tp_t by $\boldsymbol{a}$ in $\mathfrak{A}\cdot\mathcal{E}$. By Lemma 67 the subtuple $\boldsymbol{a}^* = \mu(\boldsymbol{v}_{t'})$ (where μ is applied componentwise) of $\boldsymbol{a}$ satisfies $\mathrm{Tp}_{\mathsf{t}'}$ in $\mathfrak{A} \cdot \mathcal{E}$. By applying h' to $\boldsymbol{a}^*$ we obtain a subtuple $\boldsymbol{b}^* = h'(\boldsymbol{a}^*)$ of $\boldsymbol{b}$ which satisfies $\boldsymbol{b}^* \in \mathrm{Tp}_{\mathsf{t}'}^{\mathfrak{B}'}$. This ensures the satisfaction of Formulae 66 by $\mathfrak{B}'$.

Finally, let $\mathsf{t} \in \mathcal{E}$, $\kappa\colon \boldsymbol{v} \hookrightarrow \boldsymbol{v}$ a permutation of the variables and $\boldsymbol{b}$ a tuple from $\mathfrak{B}'$ such that $\boldsymbol{b} \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{B}'}$. Since by definition of $\mathfrak{B}'$ the interpretation $\mathrm{Tp}_\mathsf{t}^{\mathfrak{B}'}$ is non-empty if and only if $\mathsf{t} \in \mathcal{E}_+$ it suffices to assume just that. From $\boldsymbol{b} \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{B}'}$ we obtain via the definition of $\mathfrak{B}'$ that there is a tuple $\boldsymbol{a}$ in $\mathfrak{A} \cdot \mathcal{E}$ such that $\boldsymbol{a} \in \mathrm{Tp}_\mathsf{t}^{\mathfrak{A}\cdot\mathcal{E}}$ and $h'(\boldsymbol{a}) = \boldsymbol{b}$. As Formulae 67 hold in $\mathfrak{A} \cdot \mathcal{E}$ we can conlude that $\eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_\mathsf{t}}(\boldsymbol{a}) \in \mathrm{Tp}_{\kappa\mathsf{t}}^{\mathfrak{A}\cdot\mathcal{E}}$. As the application of maps that are defined on elements extends to tuples componentwise, $\eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_\mathsf{t}}$ and h' commute, hence

$$h'(\eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_\mathsf{t}}(\boldsymbol{a})) = \eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_\mathsf{t}}(h'(\boldsymbol{a})) = \eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_\mathsf{t}}(\boldsymbol{b}) \qquad (68)$$

and since h' is a homomorphism we obtain immediately $\eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_\mathsf{t}}(\boldsymbol{b}) \in \mathrm{Tp}_{\kappa\mathsf{t}}^{\mathfrak{B}'}$. This proves that $\mathfrak{B}'$ satisfies Formulae 67 and concludes the whole proof of the lemma. $\qquad\square$

Proposition 69. *For every $\mathbb{FO}_{=}^2$ τ-sentence Φ, there exists a $\mathbb{FO}_{=}^2$ sentence Ψ such that $\llbracket \Phi \rrbracket^{\to} = \llbracket \Psi \rrbracket|_\tau$.*

Proof. Thanks to Lemma 65, we can w.l.o.g. assume Φ to be in Scott's normal form, i.e.

$$\Phi = \forall xy.\varphi(x,y) \wedge \bigwedge_i \forall x \exists y.\psi_i(x,y), \qquad (69)$$

where φ and all ψ_i are quantifier-free $\mathbb{FO}_{=}^2$ τ-formulae.

Let $\boldsymbol{v} = \{v_1, v_2\}$ be the set of dedicated type variables we will use in this proof. We now define the set $\mathcal{E}$ of eligible types as all rigid 1-types and 2-types (using variables from $\boldsymbol{v}$) wrt. τ. Note that $\mathcal{E}$ is closed under variable permutations. As there are only two permutations on $\{v_1, v_2\}$ (the identity and the permutation κ mapping v_1 to v_2 and v_2 to v_1) we will denote for every $\mathsf{t} \in \mathcal{E}$ by t^- its flipped version $\kappa\mathsf{t}$. Note that this leaves constants untouched.

Let now

$$\Psi = \Psi_{\mathrm{hom}} \wedge \bigvee_{(\mathcal{E}_+, \mathcal{E}_!) \in (\![\Phi]\!)} \Psi_{(\mathcal{E}_+, \mathcal{E}_!)}, \qquad (70)$$

where $\Psi_{(\mathcal{E}_+, \mathcal{E}_!)}$ is obtained as the conjunction over $\Psi_{\mathrm{gen},(\mathcal{E}_+, \mathcal{E}_!)}$ and the following sentences (letting $\mathcal{E}_{+,1}$ and $\mathcal{E}_{+,2}$ denote the 1- and 2-types in $\mathcal{E}_+$, respectively, and letting $\mathcal{E}_{!,1}$ denote the 1-types in $\mathcal{E}_!$):

$$\forall xy.\mathrm{Tp}_\mathsf{t}(x) \wedge \mathrm{Tp}_\mathsf{t}(y) \Rightarrow x = y \qquad \text{for } \mathsf{t} \in \mathcal{E}_{!,1} \quad (71)$$

$$\forall xy.\mathrm{Tp}_{\mathsf{t}'}(x) \wedge \mathrm{Tp}_{\mathsf{t}''}(y) \Rightarrow \bigvee_{\substack{\mathsf{t} \in \mathcal{E}_{+,2} \\ \mathsf{t}|_{v_1} = \mathsf{t}', \, \mathsf{t}|_{v_2} = \mathsf{t}''}} \mathrm{Tp}_\mathsf{t}(x,y) \quad \text{for } \mathsf{t}', \mathsf{t}'' \in \mathcal{E}_{+,1} \, (72)$$

$$\forall x \exists y. \left(\bigvee_{\mathsf{t} \in \mathcal{E}_{+,1}} \mathrm{Tp}_\mathsf{t}(x) \Rightarrow \bigvee_{\substack{\mathsf{t}' \in \mathcal{E}_{+,2} \\ \mathsf{t}' \models \psi_i(v_1, v_2)}} \mathrm{Tp}_{\mathsf{t}'}(x,y) \right) \quad \psi_i \text{ from } \Phi \quad (73)$$

Clearly, Ψ is a $\mathbb{FO}^2_{\doteq}$ $\tau \uplus \sigma$-sentence. Moreover, we obtain $\llbracket \Phi \rrbracket^{\rightarrow} = \llbracket \Psi \rrbracket|_{\tau}$:

"$\subseteq$": Let $\mathfrak{A} \in \llbracket \Phi \rrbracket$ and let $h \colon \mathfrak{A} \to \mathfrak{B}$. Obtain $\mathfrak{A}' = \mathfrak{A} \cdot \mathcal{E}$ and $\mathfrak{B}' = \mathfrak{B} \cdot h(\mathfrak{A} \cdot \mathcal{E}|_{\sigma})$ as described before. Pick $(\mathcal{E}_+, \mathcal{E}_!) = \mathcal{E}(\mathfrak{A})$. Then $\mathfrak{A}' \models \Psi_{(\mathcal{E}_+, \mathcal{E}_!)}$ because of Φ-modelhood and $\mathfrak{A}' \models \Psi_{\text{hom}}$ by construction. Furthermore, by applying Lemma 68 $\mathfrak{A}'$ as well as $\mathfrak{B}'$ satisfy $\Psi_{\text{gen},(\mathcal{E}_+, \mathcal{E}_!)}$. Next, Formulae 71 holds by definition of $\mathfrak{A}'$ and the fact that $\mathcal{E}_{!,1}$ contains the 1-types realized exactly once. Hence every type predicate associated to such a 1-type can only be satisfied by exactly one element from $\mathfrak{A}'$. By definition of $\mathfrak{B}'$ Formulae 71 also hold there: Since every type predicate from σ is interpreted in $\mathfrak{B}'$ as the homomorphic image of the interpretation of the very same type predicate in $\mathfrak{A}'$, and those are singleton sets, we obtain the claimed satisfaction. Now let b_1, b_2 be elements from $\mathfrak{B}'$ such that there are types $\mathfrak{t}', \mathfrak{t}'' \in \mathcal{E}_{+,1}$ with $b_1 \in \text{Tp}_{\mathfrak{t}'}^{\mathfrak{B}'}$ and $b_1 \in \text{Tp}_{\mathfrak{t}''}^{\mathfrak{B}'}$. By definition there are a_1 and a_2 in $\mathfrak{A}'$ with $a_1 \in \text{Tp}_{\mathfrak{t}'}^{\mathfrak{A}'}$ and $a_1 \in \text{Tp}_{\mathfrak{t}''}^{\mathfrak{A}'}$ such that $h(a_1) = b_1$ and $h(a_2) = b_2$. Without loss of generality we now assume $\boldsymbol{v}_{\mathfrak{t}'} = \{v_1\}$ and $\boldsymbol{v}_{\mathfrak{t}''} = \{v_2\}$. If this were not the case, using the closure under variable permutation we could enforce this condition. Note that it will be only important for the next step in a technical way, to extract a 2-type that is rigid. Because now we obtain by definition of $\mathfrak{A}'$ (and the uniqueness of the realized types of elements/tuples) that a_1 realizes $\mathfrak{t}'$, a_2 realizes $\mathfrak{t}''$ and (a_1, a_2) realizes some type $\mathfrak{t}$ that naturally satisfies $\mathfrak{t}|_{v_1} = \mathfrak{t}'$ and $\mathfrak{t}|_{v_2} = \mathfrak{t}''$. Hence $(a_1, a_2) \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{A}'}$ and applying h we obtain $(b_1, b_2) = h(a_1, a_2) \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{B}'}$. This argument yields 72. Finally we discuss that $\mathfrak{B}'$ satisfies 73. Suppose b be in $\mathfrak{B}'$ and let $\mathfrak{t} \in \mathcal{E}_{+,1}$ be some type such that $b \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{B}'}$. Hence there is a in $\mathfrak{A}'$ such that $h(a) = b_1$ and $a \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{A}'}$. As $\mathfrak{A}'$ as an expansion has the same domain as $\mathfrak{A}$ and $\mathfrak{A} \models \Phi$ we obtain an element a_i such that (a, a_i) satisfies $\psi_i(x, y)$ for all $\forall \exists$-conjuncts in Φ. The type $\mathfrak{t}(v_1, v_2)$ realized by (a, a_i) is in $\mathcal{E}_{+,2}$ and we have $\mathfrak{t}' \models \psi_i(v_1, v_2)$. Hence $(a, a_i) \in \text{Tp}_{\mathfrak{t}'}^{\mathfrak{A}'}$. Applying h we first obtain a $b_i = h(a_i)$ such that also $(b, b_i) \in \text{Tp}_{\mathfrak{t}'}^{\mathfrak{B}'}$ holds. By this argument 73 holds in $\mathfrak{B}'$. Consequently $\mathfrak{B}'$ must be a model of Ψ, hence $\mathfrak{B} \in \llbracket \Psi \rrbracket|_{\tau}$.

"$\supseteq$": Let $\mathfrak{B}'$ be a model of Ψ and let $\Psi_{(\mathcal{E}_+, \mathcal{E}_!)}$ be the disjunct made true in the rear part of Ψ. The crucial observation about the structure $\mathfrak{B}'$ is that in general, elements a or a tuple (a, b) from B' may satisfy more than one type predicate. Note that if this would not be the case, we'd easily find a weak substructure $\mathfrak{A}'$ of $\mathfrak{B}$ which is projectively in $\llbracket \Phi \rrbracket$. This holds since the type predicates inherit the satisfiability of Φ. In order to overcome this difficulty of "ambiguity of types" we construct a structure $\mathfrak{A} \in \llbracket \Phi \rrbracket$ and a homomorphism $\mathfrak{A} \to \mathfrak{B}'|_{\tau}$ as follows: first construct the σ-structure $\mathfrak{A}'$ by letting

$$A' = \{(b, \mathfrak{t}) \mid b \in B, b \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{B}}, \mathfrak{t} \in \mathcal{E}_{+,1}\}. \tag{74}$$

For the unary type predicates, let $\text{Tp}_{\mathfrak{t}}^{\mathfrak{A}'} = A' \cap (B \times \{\mathfrak{t}\})$ and for binary type predicates, let

$$\text{Tp}_{\mathfrak{t}}^{\mathfrak{A}'} = \left\{ ((b, \mathfrak{t}'), (b', \mathfrak{t}'')) \;\middle|\; \mathfrak{t}|_{v_1} = \mathfrak{t}', \mathfrak{t}|_{v_2} = \mathfrak{t}'', (b, b') \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{B}} \right\}$$

In a next step, we define $\mathfrak{A}$ from $\mathfrak{A}'$ as follows: let k be the maximum number of distinct $\mathfrak{t}'' \in \mathcal{E}_{+,2}$ with $(a, a') \in \text{Tp}_{\mathfrak{t}''}^{\mathfrak{A}'}$ across all pairs (a, a') from A'. Define a function $\mathfrak{t}^? \colon A' \times A' \times \{0, \ldots, k-1\} \to \mathcal{E}_{+,2}$ such that for every $a, a' \in A'$, the set $\{\mathfrak{t}^?(a, a', i) \mid 0 \le i < k\}$ contains all and only the distinct $\mathfrak{t}''$ with $(a, a') \in \text{Tp}_{\mathfrak{t}''}^{\mathfrak{A}'}$ (with possible repetitions) and $\mathfrak{t}^?(a, a', i) = \mathfrak{t}^?(a', a, i)^-$. Note that the last identity is possible by (67). Now, let

$$A_! = \{(b, \mathfrak{t}) \in A' \mid \mathfrak{t} \in \mathcal{E}_{!,1}\} \tag{75}$$

and let

$$A = A_! \times \{0\} \cup (A' \setminus A_!) \times \{0, \ldots, k-1\}. \tag{76}$$

Note that $A_!$ contains for every $\mathfrak{t} \in \mathcal{E}_{!,1}$ exactly one element of the form $(b, \mathfrak{t})$ since $\mathfrak{B}$ satisfies (71). We assign relations to the type predicates as follows (for all $\mathfrak{t} \in \mathcal{E}_{+,1}$ and $\mathfrak{t}'' \in \mathcal{E}_{+,2}$):

$$\text{Tp}_{\mathfrak{t}}^{\mathfrak{A}} = \{(a, i) \mid a \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{A}'}\} \text{ for } \mathfrak{t} \in \mathcal{E}_{+,1} \tag{77}$$

$$\text{Tp}_{\mathfrak{t}''}^{\mathfrak{A}} = \{((a, i), (a', j)) \mid \mathfrak{t}^?(a, a', i+j \mod k) = \mathfrak{t}''\}. \tag{78}$$

It follows from our construction that the structure $\mathfrak{A}$ satisfies formulae (64), (65), and (71). To see that (66) is satisfied consider $((a, i), (a', j)) \in A^2$ such that $((a, i), (a', j)) \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{A}}$ holds. By the definition of $\text{Tp}_{\mathfrak{t}}^{\mathfrak{A}}$ we get $\mathfrak{t}^?(a, a', i+j \mod k) = \mathfrak{t}$. Let $(a, a') = ((b, \mathfrak{t}^a), (b', \mathfrak{t}^{a'}))$. Since $(a, a') \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{A}'}$ holds by definition of $\mathfrak{t}^?$ we get by the definition of $\text{Tp}_{\mathfrak{t}}^{\mathfrak{A}'}$ that $\mathfrak{t}|_{v_1} = \mathfrak{t}^a$ holds. By the definition of $\text{Tp}_{\mathfrak{t}}^{\mathfrak{A}'}$ and since $\mathfrak{B}$ satisfies (66) we have that $(b, \mathfrak{t}^a)$ is in $\text{Tp}_{\mathfrak{t}^a}^{\mathfrak{A}'}$. This implies that $((b, \mathfrak{t}^a), i)$ is in $\text{Tp}_{\mathfrak{t}^a}^{\mathfrak{A}}$. Since $\mathfrak{t}|_{v_1} = \mathfrak{t}^a$ holds and the same argument works for v_2 and $\mathfrak{t}^{a'}$, we showed that (66) is satisfied in $\mathfrak{A}$.

Formulae (67) clearly hold in $\mathfrak{A}$. In consequence, $\Psi_{(\mathcal{E}_+, \mathcal{E}_!)}$ as well as (71) holds in $\mathfrak{A}$.

For signature elements from τ, choose $\cdot^{\mathfrak{A}}$ such that all (pairs of) domain elements realize exactly the types as indicated by the type predicates. This is possible since, by construction, there is exactly one 2-type-predicate holding for every pair of domain elements. Furthermore, every domain element has been assigned exactly two 1-type-predicates (whose associated types only differ in the used variable name but are otherwise identical by (67)) by the definitions (74) and (77). The 1-type-predicates also fit with all the two-type predicates by what we have shown before.

We now claim that $\mathfrak{A}|_{\tau}$ is a model of Φ. To see this, consider an arbitrary element $a = ((b, \mathfrak{t}), i) \in A$. Since $\mathfrak{B}$ satisfies (73), there exists a $b' \in B$ such that that the inner implication in (73) is satisfied (for $x = b$ and $y = b'$). By the construction of $\mathfrak{A}$, we know that $b \in \text{Tp}_{\mathfrak{t}}^{\mathfrak{B}}$ holds and therefore the premise of the implication is satisfied. This implies the existence of a type $\mathfrak{t}'$ such that $(b, b') \in \text{Tp}_{\mathfrak{t}'}^{\mathfrak{B}}$ and $\mathfrak{t}' \models \psi_i(v_1, v_2)$ hold. By the definition of $\mathfrak{t}^?$, we can find a j such that $\mathfrak{t}^?((b, \mathfrak{t}), (b', \mathfrak{t}'|_{v_2}), i+j \mod k) = \mathfrak{t}'$. Therefore, we get for $a' = ((b', \mathfrak{t}'|_{v_2}), j)$ that $(a, a') \in \text{Tp}_{\mathfrak{t}'}^{\mathfrak{A}}$ holds. By the definition of the τ-relations from $\mathfrak{A}$, it follows that $\psi_j(a, a')$ holds in $\mathfrak{A}$. The proof for the conjunct $\forall xy. \varphi(x, y)$

is by an analogous (but easier, since "type-internal") use of formula (72). Consider arbitrary elements $a = ((b, \mathsf{t}), i) \in A$ and $a' = ((b', \mathsf{t}'), i') \in A$. By construction of $\mathfrak{A}$ we know that $b \in \mathsf{Tp}_{\mathsf{t}}^{\mathfrak{B}}$ and $b' \in \mathsf{Tp}_{\mathsf{t}'}^{\mathfrak{B}}$. By 72 there is a at least one $\mathsf{t}'' \in \mathcal{E}_{+,2}$ with $\mathsf{t}''|_{v_1} = \mathsf{t}$ and $\mathsf{t}''|_{v_2} = \mathsf{t}'$ such that $(b, b') \in \mathsf{Tp}_{\mathsf{t}''}^{\mathfrak{B}}$. Hence by definition of $\mathsf{t}^?$ one of those types t'' sastisfies $\mathsf{t}^?((b, \mathsf{t}), (b', \mathsf{t}'), i + i' \bmod k) = \mathsf{t}''$. Consequently $(a, a') \in \mathsf{Tp}_{\mathsf{t}''}^{\mathfrak{A}}$ by construction. By the definition of the τ-relations from $\mathfrak{A}$ we immediately obtain that (a, a') realizes t''. Now note that, as $\mathsf{t}'' \in \mathcal{E}_{+,2}$, $\mathsf{t}'' \vDash \varphi(v_1, v_2)$ since $\mathcal{E}_+$ collects all eligible types realized by some $\mathfrak{A}^* \in [\![\Phi]\!]$. Hence (a, a') satisfies φ in $\mathfrak{A}$. Thus the conjunct $\forall xy.\varphi(x, y)$ holds for $\mathfrak{A}$.

By our construction, $h \colon \mathfrak{A}|_\tau \to \mathfrak{B}|_\tau$ with $((b, \mathsf{t}), i) \mapsto b$ is a homomorphism, which proves the statement. $\qquad\square$

Before characterizing the homclosures of a $\mathbb{GFO}_=$ or a $\mathbb{TGF}$ sentence, we first discuss normal forms for each fragment, which we will facilitate in the upcoming proofs.

Lemma 70. *Let Φ be a $\mathbb{GFO}_=$ τ-sentence. Then there exists a* normal form Φ' *over an extended signature $\sigma \uplus \tau \uplus \{\mathsf{D}\}$ (with D being a unary relational symbol) such that Φ' has the form $\Phi' = \Phi_\mathsf{D} \wedge \Phi'_\forall \wedge \Phi'_{\forall\exists}$ where*

$$\Phi_\mathsf{D} = \forall x.\mathsf{D}(x), \tag{79}$$

$$\Phi'_\forall = \bigwedge_i \forall \boldsymbol{x}.\Big(\alpha_i(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]\Big), \tag{80}$$

$$\Phi'_{\forall\exists} = \bigwedge_j \forall \boldsymbol{z}.\Big(\beta_j(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma_j(\boldsymbol{y}\boldsymbol{z})\Big), \tag{81}$$

with ϑ_i is a disjunction of literals, α_i, β_j are guard atoms, and γ_j is an atom, such that every model of Φ can be extended to a model of Φ' and any model of Φ' restricted to the signature τ is a model of Φ.

Proof. Let Φ be a $\mathbb{GFO}_=$ τ-sentence. The construction will be done iteratively. As a note on notation we will collect for every iteration step i in the first big transformation some sentences in sets $\boldsymbol{A}_\Phi^i$ and $\boldsymbol{E}_\Phi^i$. The collected sentences will be of specific forms: Sentences from $\boldsymbol{A}_\Phi^i$ have the form

$$\forall \boldsymbol{x}.\alpha(\boldsymbol{x}) \Rightarrow \vartheta[\boldsymbol{x}] \tag{82}$$

(where α is a guard atom and ϑ is quantifier free) and those from $\boldsymbol{E}_\Phi^i$ have the form

$$\forall \boldsymbol{z}.\Big(\beta(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma(\boldsymbol{y}\boldsymbol{z}) \wedge \varphi[\boldsymbol{y}\boldsymbol{z}]\Big) \tag{83}$$

(where β and γ are guard atoms and φ is quantifier free). Finally, $\boldsymbol{A}_\Phi^i$ and $\boldsymbol{E}_\Phi^i$ will constitute two finite monotonically increasing sequences of sets, i.e. $\boldsymbol{A}_\Phi^i \subseteq \boldsymbol{A}_\Phi^{i+1}$ and $\boldsymbol{E}_\Phi^i \subseteq \boldsymbol{E}_\Phi^{i+1}$ for every $i \geq 0$ such that $\boldsymbol{A}_\Phi^{i+1}$ and $\boldsymbol{E}_\Phi^{i+1}$ are defined.

If χ is a (proper) subformula in Φ and ψ a formula with the same free variables as χ, we will write $\Phi[\chi/\psi]$ for the formula we obtain by replacing every occurence of χ in Φ by ψ.

In a preliminary step we will transform Φ to a sentence Φ^0 to ensure that

1) every proper subformula of Φ^0 has at least one free variable,

2) Φ^0 has an outermost universal and guarded quantification (and every other quantification is in the scope of this universal quantifier), and

3) every quantification properly inside the scope of the outermost universal one is existential.

Let u be a fresh variable symbol completely unused. To establish Item 1 we introduce for every relational symbol $\mathsf{R} \in \tau$ a fresh symbol R' (which we add to σ) such that $\mathrm{ar}(\mathsf{R}') = \mathrm{ar}(\mathsf{R}) + 1$. We then substitute every atom $\mathsf{R}(\boldsymbol{t})$ in Φ by the corresponding $\mathsf{R}'(\boldsymbol{t}, u)$. The so obtained formula will be denoted by $\psi(u)$. Note that by uniformly using u the formula $\psi(u)$ stays guarded. We furthermore let $\boldsymbol{A}_\Phi^0$ consist of just the sentences $\forall \boldsymbol{x} \forall u.\mathsf{R}'(\boldsymbol{x}, u) \Rightarrow \mathsf{R}(\boldsymbol{x})$ for each relational symbol $\mathsf{R} \in \tau$. Also we set $\boldsymbol{E}_\Phi^0 = \emptyset$.

To ensure Item 2 we introduce the fresh relation symbol D of arity 1. We now let Ψ^0 be the sentence $\forall u.\mathsf{D}(u) \Rightarrow \psi(u)$.

Ensuring Item 3 is a simple exercise of double negation: Let every occurence of a subformula from $\psi(u)$ of the form $\forall \boldsymbol{x}.\alpha(\boldsymbol{xy}) \Rightarrow \varphi[\boldsymbol{xy}]$ be substituted by $\neg \exists \boldsymbol{x}.\alpha(\boldsymbol{xy}) \wedge \neg\varphi[\boldsymbol{xy}]$. The so obtained formula will be denoted by $\psi'(u)$.

Hence we obtain a guarded sentence $\Phi^0 = \forall u.\mathsf{D}(u) \Rightarrow \psi'(u)$ that satisfies Items 1 - 3. Note that every model of $\mathfrak{A} \in [\![\Phi]\!]$ can be expanded to one of $\Phi_\mathsf{D} \wedge \Phi^0 \wedge \bigwedge_{\Theta \in \boldsymbol{A}_\Phi^0} \Theta$. To see this, let $\mathfrak{A}'$ (with domain A) be the expansion of $\mathfrak{A}$, where $\mathsf{D}^{\mathfrak{A}'} = A$ and for every $\mathsf{R} \in \tau$ we set $(\mathsf{R}')^{\mathfrak{A}'} = \mathsf{R}^{\mathfrak{A}} \times A$. $\mathfrak{A}'$ satisfies $\Phi_\mathsf{D} \wedge \Phi^0 \wedge \bigwedge_{\Theta \in \boldsymbol{A}_\Phi^0} \Theta$.

On the other hand, if $\mathfrak{A}$ is a model of $\Phi_\mathsf{D} \wedge \Phi^0 \wedge \bigwedge_{\Theta \in \boldsymbol{A}_\Phi^0} \Theta$, then $\mathfrak{A}|_\tau$ is one of Φ. To ensure this we added to $\boldsymbol{A}_\Phi^0$ the sentences $\forall \boldsymbol{x} \forall u.\mathsf{R}'(\boldsymbol{x}, u) \Rightarrow \mathsf{R}(\boldsymbol{x})$. Note that, since we do not allow empty models, Φ_D also ensures the non-emptiness of interpretations of D.

We now begin to iteratively "unnest" the quantifiers in Φ^0. Let $\chi_0(\boldsymbol{z})$ be a proper subformula of Φ^0 of the form

$$\chi_0(\boldsymbol{z}) = \exists \boldsymbol{y}.\beta(\boldsymbol{yz}) \wedge \varphi[\boldsymbol{yz}]$$

with φ being quantifier free. We introduce a fresh relation symbol R_{χ_0} (which we add to σ) of the same arity as length of $\boldsymbol{z}$. We obtain a sentence Φ^1 by substituting $\mathsf{R}_{\chi_0}(\boldsymbol{z})$ for each occurence of $\chi_0(\boldsymbol{z})$ in Φ^0, i.e.

$$\Phi^1 = \Phi^0[\chi_0(\boldsymbol{z})/\mathsf{R}_{\chi_0}(\boldsymbol{z})]. \tag{84}$$

Let $\boldsymbol{E}_\Phi^1$ be $\boldsymbol{E}_\Phi^0$ with the added sentence

$$\forall \boldsymbol{z}.\mathsf{R}_{\chi_0}(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\beta(\boldsymbol{yz}) \wedge \varphi[\boldsymbol{yz}] \tag{85}$$

and let $\boldsymbol{A}_\Phi^1$ be $\boldsymbol{A}_\Phi^0$ with the added sentence

$$\forall \boldsymbol{yz}.\beta(\boldsymbol{yz}) \Rightarrow (\varphi[\boldsymbol{yz}] \Rightarrow R_{\chi_0}(\boldsymbol{z})). \tag{86}$$

Note that the later sentence is of the form $\forall \boldsymbol{yz}.\beta(\boldsymbol{yz}) \Rightarrow \vartheta[\boldsymbol{yz}]$ with ϑ quantifier free, as $(\varphi[\boldsymbol{yz}] \Rightarrow R_{\chi_0}(\boldsymbol{z}))$ is quantifier free.

By iterating this process, picking from Φ^i a proper subformula $\chi_i(\boldsymbol{z})$ of the form

$$\chi_i(\boldsymbol{z}) = \exists \boldsymbol{y}.\beta(\boldsymbol{yz}) \wedge \varphi[\boldsymbol{yz}]$$

with φ being quantifier free and introducing a fresh relation symbol $\mathtt{R}_{\chi_i}$ obtaining Φ^{i+1} and enlarging $\boldsymbol{A}_\Phi^i$ and $\boldsymbol{E}_\Phi^i$ to $\boldsymbol{A}_\Phi^{i+1}$ and $\boldsymbol{E}_\Phi^{i+1}$ respectively, we will terminate at some n with the sentence

$$\Phi^n = \forall u.\mathtt{D}(u) \Rightarrow \mathtt{R}_{\chi_{n-1}} \tag{87}$$

and sets $\boldsymbol{A}_\Phi^n$, $\boldsymbol{E}_\Phi^n$. We set $\boldsymbol{A}_\Phi$ to be $\boldsymbol{A}_\Phi^n$ with Φ^n added and $\boldsymbol{E}_\Phi = \boldsymbol{E}_\Phi^n$.

Note that in every iteration, every model $\mathfrak{A} \in [\![\Phi_\mathtt{D} \wedge \Phi^{i-1}]\!]$ can be expanded to one of $\Phi_\mathtt{D} \wedge \Phi^i \wedge \Theta_{\boldsymbol{A},i} \wedge \Theta_{\boldsymbol{E},i}$ with $\Theta_{\boldsymbol{A},i}$ being the unique element from the singleton set $\boldsymbol{A}_\Phi^i \setminus \boldsymbol{A}_\Phi^{i-1}$ and $\Theta_{\boldsymbol{E},i}$ being the unique element from the singleton set $\boldsymbol{E}_\Phi^i \setminus \boldsymbol{E}_\Phi^{i-1}$. The trick is to interpret $\mathtt{R}_{\chi_{i-1}}$ to be the set of tuples satisfying χ_{i-1} in $\mathfrak{A}$. On the other hand, by "forgetting" the relation symbols introduced in the i-th iteration we obtain from a model $\mathfrak{A} \in [\![\Phi_\mathtt{D} \wedge \Phi^i \wedge \Theta_{\boldsymbol{A},i} \wedge \Theta_{\boldsymbol{E},i}]\!]$ (where $\Theta_{\boldsymbol{A},i}$ is the unique element from the singleton set $\boldsymbol{A}_\Phi^i \setminus \boldsymbol{A}_\Phi^{i-1}$ and $\Theta_{\boldsymbol{E},i}$ is the unique element from the singleton set $\boldsymbol{E}_\Phi^i \setminus \boldsymbol{E}_\Phi^{i-1}$) a model of $\Phi_\mathtt{D} \wedge \Phi^{i-1}$. The choice of $\Theta_{\boldsymbol{A},i}$ and $\Theta_{\boldsymbol{E},i}$ in the construction ensures that this works.

As now every information of our initial Φ^0 is stored in the sets $\boldsymbol{A}_\Phi$ and $\boldsymbol{E}_\Phi$, the next manipulations will take place inside those sets. First we look at $\boldsymbol{E}_\Phi$. We know that every sentence from $\boldsymbol{E}_\Phi$ is of the form

$$\forall \boldsymbol{xz}.\beta(\boldsymbol{xz}) \Rightarrow \exists \boldsymbol{y}.\gamma(\boldsymbol{yz}) \wedge \varphi[\boldsymbol{yz}], \tag{88}$$

where β and γ are respective guard atoms and φ is quantifier free. We go through every sentence of $\boldsymbol{E}_\Phi$ and do the following steps:

- For every $\Theta \in \boldsymbol{E}_\Phi$ such that

$$\forall \boldsymbol{xz}.\beta(\boldsymbol{xz}) \Rightarrow \exists \boldsymbol{y}.\gamma(\boldsymbol{yz}) \wedge \varphi[\boldsymbol{yz}] \tag{89}$$

we introduce fresh relation symbols $\mathtt{P}_\Theta$ (arity being equal to the length of $\boldsymbol{z}$) and $\mathtt{Q}_\Theta$ (arity being the length of $\boldsymbol{yz}$) to σ.

- Then we decompose each Θ into the three sentences

$$\forall \boldsymbol{xz}.\beta(\boldsymbol{xz}) \Rightarrow \mathtt{P}_\Theta(\boldsymbol{z}) \tag{90}$$
$$\forall \boldsymbol{z}.\mathtt{P}_\Theta(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\mathtt{Q}_\Theta(\boldsymbol{yz}) \tag{91}$$
$$\forall \boldsymbol{yz}.\mathtt{Q}_\Theta(\boldsymbol{yz}) \Rightarrow \gamma(\boldsymbol{yz}) \wedge \varphi[\boldsymbol{yz}]. \tag{92}$$

- Finally, we substitute in $\boldsymbol{E}_\Phi$ each sentence Θ by the corresponding sentence 91 obtained in the previous step and furthermore we add to $\boldsymbol{A}_\Phi$ all the sentences 90 and 92 yielded by decomposing each Θ.

This step leaves us with particularly simple sentences in $\boldsymbol{E}_\Phi$: They are just of the form

$$\forall \boldsymbol{z}.\beta(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma(\boldsymbol{yz}) \tag{93}$$

where β and γ are atoms. Furthermore, γ contains all the variables contained in β.

Finally we simplify the sentences in the set $\boldsymbol{A}_\Phi$. First we note, that every sentence $\Theta \in \boldsymbol{A}_\Phi$ is of the form

$$\forall \boldsymbol{x}.\alpha(\boldsymbol{x}) \Rightarrow \vartheta[\boldsymbol{x}] \tag{94}$$

with α a guard atom and ϑ being a quantifier free formula. We now transform ϑ into conjunctive normal form, i.e. $\vartheta = \bigwedge_i \vartheta_i$, where the ϑ_i are disjunctions over literals. As the conjunction commutes with the universal quantification, φ is equivalent to the sentence

$$\bigwedge_i \forall \boldsymbol{x}.\alpha(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]. \tag{95}$$

Replacing Θ in $\boldsymbol{A}_\Phi$ by all the sentences

$$\forall \boldsymbol{x}.\alpha(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}] \tag{96}$$

of that conjunction and doing this for every $\Theta \in \boldsymbol{A}_\Phi$ yields us a simpler $\boldsymbol{A}_\Phi$. To summarize, every sentence of $\boldsymbol{A}_\Phi$ now has the form

$$\forall \boldsymbol{x}.\alpha(\boldsymbol{x}) \Rightarrow \vartheta[\boldsymbol{x}] \tag{97}$$

with α a guard atom and ϑ a disjunction of literals.

In consequence we obtain the sentence $\Phi' = \Phi_\mathtt{D} \wedge \Phi'_\forall \wedge \Phi'_{\forall\exists}$ where

$$\Phi_\mathtt{D} = \forall x.\mathtt{D}(x), \tag{98}$$
$$\Phi'_\forall = \bigwedge_{\Theta \in \boldsymbol{A}_\Phi} \Theta, \text{ and} \tag{99}$$
$$\Phi'_{\forall\exists} = \bigwedge_{\Theta \in \boldsymbol{E}_\Phi} \Theta. \tag{100}$$

Note that $\Phi_\mathtt{D}$ is guarded, hence Φ' is, too. Finally, as every step ensured the models could be expanded or restricted to fulfill the respective sentences after or before the step, every model of Φ can be expanded to one of Φ' and every model of Φ' can be restricted to get one of Φ. $\qquad\square$

As a corollary we obtain a quite similar normal form for $\mathbb{TGF}$.

Corollary 71. *Let Φ be a $\mathbb{TGF}$ τ-sentence. Then there exists a* normal form Φ' *over an extended signature $\sigma \uplus \tau \uplus \{\mathtt{Univ}\}$ (where $\mathtt{Univ}$ is a binary auxiliary predicate) such that Φ' has the form $\Phi' = \Phi'_{\mathtt{Univ}} \wedge \Phi'_\forall \wedge \Phi'_{\forall\exists}$ where*

$$\Phi_{\mathtt{Univ}} = \forall xy.\mathtt{Univ}(x,y), \tag{101}$$
$$\Phi'_\forall = \bigwedge_i \forall \boldsymbol{x}.(\alpha_i(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]), \text{ and} \tag{102}$$
$$\Phi'_{\forall\exists} = \bigwedge_j \forall \boldsymbol{z}.(\beta_j(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma_j(\boldsymbol{yz})), \tag{103}$$

with ϑ_i a disjunction of literals, α_i, β_j guard atoms, and γ_j an atom, such that every model of Φ can be extended to a model of Φ' and any model of Φ' restricted to the signature τ is a model of Φ.

Proof. Let Φ be a $\mathbb{TGF}$ τ-sentence. In a preparatory step we guard every unguarded quantification (which can only happen for quantification over one or two variables) inside of Φ by $\mathtt{Univ}$. What we mean by this is the following: Substitute every subformula

- of the form $\exists x.\varphi[x]$ by $\exists x.\mathtt{Univ}(x,x) \wedge \varphi[x]$,
- of the form $\exists x.\varphi[x,y]$ by $\exists x.\mathtt{Univ}(x,y) \wedge \varphi[x,y]$,
- of the form $\exists xy.\varphi[x,y]$ by $\exists xy.\mathtt{Univ}(x,y) \wedge \varphi[x,y]$,
- of the form $\forall x.\varphi[x]$ by $\forall x.\mathtt{Univ}(x,x) \Rightarrow \varphi[x]$,
- of the form $\forall x.\varphi[x,y]$ by $\forall x.\mathtt{Univ}(x,y) \Rightarrow \varphi[x,y]$,
- of the form $\forall xy.\varphi[x,y]$ by $\forall xy.\mathtt{Univ}(x,y) \Rightarrow \varphi[x,y]$.

We denote the so obtained sentence by Ψ. First note that every model of $\Phi_{\mathtt{Univ}} \wedge \Psi$ is a model of Φ after "forgetting" $\mathtt{Univ}$. On the other hand, by interpreting $\mathtt{Univ}$ as the full binary relation every model of Φ can be expanded to one of $\Phi_{\mathtt{Univ}} \wedge \Psi$. More importantly, Ψ itself is a $\mathbb{GFO}_=$ $\tau \uplus \{\mathtt{Univ}\}$-sentence. By applying Lemma 70 we hence obtain a normal form over an extended signature $\tau \uplus \{\mathtt{Univ}\} \uplus \sigma \uplus \{\mathtt{D}\}$ as the sentence $\Psi' = \Phi_{\mathtt{D}} \wedge \Psi'_{\forall} \wedge \Psi'_{\forall\exists}$ where

$$\Phi_{\mathtt{D}} = \forall x.\mathtt{D}(x), \tag{104}$$

$$\Psi'_{\forall} = \bigwedge_i \forall \boldsymbol{x}.\Big(\alpha_i(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]\Big), \tag{105}$$

$$\Psi'_{\forall\exists} = \bigwedge_j \forall \boldsymbol{z}.\Big(\beta_j(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma_j(\boldsymbol{y}\boldsymbol{z})\Big), \tag{106}$$

with ϑ_i a disjunction of literals, α_i, β_j guard atoms, and γ_j an atom. In a final step we will eliminate every subformula in $\Psi'_{\forall}$ and $\Psi'_{\forall\exists}$ of the form $\mathtt{D}(x)$ by $\mathtt{Univ}(x,x)$. We will denote the so obtained $\tau \uplus \{\mathtt{Univ}\} \uplus \sigma$-sentences by $\Phi'_{\forall}$ and $\Phi'_{\forall\exists}$ respectively. By dropping $\Phi_{\mathtt{D}}$ this yields us the desired normal form $\Phi' = \Phi_{\mathtt{Univ}} \wedge \Phi'_{\forall} \wedge \Phi'_{\forall\exists}$. $\qquad\square$

Having established these normal forms, we are in a position of establishing further preliminary results. Though we will first need some notations to actually talk about those.

Definition 72. *Let Φ be a $\mathbb{GFO}_=$ or $\mathbb{TGF}$ τ-sentence in normal form, i.e. $\Phi = \Phi_{\mathtt{D}} \wedge \Phi_{\forall} \wedge \Phi_{\forall\exists}$ or $\Phi_{\mathtt{Univ}} \wedge \Phi_{\forall} \wedge \Phi_{\forall\exists}$ respectively, where*

$$\Phi_{\mathtt{D}} = \forall x.\mathtt{D}(x), \tag{107}$$

$$\Phi_{\mathtt{Univ}} = \forall xy.\mathtt{Univ}(x,y), \tag{108}$$

$$\Phi_{\forall} = \bigwedge_i \forall \boldsymbol{x}.(\alpha_i(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]), \text{ and} \tag{109}$$

$$\Phi_{\forall\exists} = \bigwedge_j \forall \boldsymbol{z}.(\beta_j(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma_j(\boldsymbol{y}\boldsymbol{z})), \tag{110}$$

with ϑ_i a disjunction of literals, α_i, β_j guard atoms, and γ_j an atom. Additionally let $\mathrm{width}(\Phi)$ be the maximal arity of a relation symbol appearing in Φ and $\mathcal{E}$ be the set of all guarded

types that are also rigid and of order $\leq \mathrm{width}(\Phi)$. To every $(\mathcal{E}_+, \mathcal{E}_!) \in (\!|\Phi|\!)$, we associate for each conjunct in

- $\Phi_{\forall}$ a set $T_i^{(\mathcal{E}_+, \mathcal{E}_!)}$ containing all pairs $(\mathfrak{t}, \nu)$ where $\mathfrak{t}$ is a type from $\mathcal{E}_+$ guarded by α_i as witnessed by ν,
- $\Phi_{\forall\exists}$ a set $B_j^{(\mathcal{E}_+, \mathcal{E}_!)}$ containing all pairs $(\mathfrak{t}, \nu)$ where $\mathfrak{t}$ is a type from $\mathcal{E}_+$ guarded by β_j as witnessed by ν, and
- $\Phi_{\forall\exists}$ a set $P_j^{(\mathcal{E}_+, \mathcal{E}_!)}$ containing all pairs $(\mathfrak{t}, \nu)$ where $\mathfrak{t}$ is a type from $\mathcal{E}_+$ guarded by γ_j as witnessed by ν.

For every $(\mathcal{E}_+, \mathcal{E}_!) \in (\!|\Phi|\!)$ we call two elements $(\mathfrak{t}, \nu) \in B_j^{(\mathcal{E}_+, \mathcal{E}_!)}$ and $(\mathfrak{t}', \nu') \in P_j^{(\mathcal{E}_+, \mathcal{E}_!)}$ fitting and write $(\mathfrak{t}, \nu) \Subset_j (\mathfrak{t}', \nu')$ if ν' extends ν, and the subtype of $\mathfrak{t}'$ w.r.t. the variables $\boldsymbol{v}^* = \boldsymbol{v} \cap \nu^{(\prime)}(\boldsymbol{z})$ coincides with $\mathfrak{t}$, that is, $\mathfrak{t} = \mathfrak{t}'|_{\boldsymbol{v}^*}$.

Note that since our types are maximally consistent, each $\mathfrak{t} \in \mathcal{E}_+$ (for $(\mathcal{E}_+, \mathcal{E}_!) \in (\!|\Phi|\!)$) with Φ a sentence from $\mathbb{GFO}_=$ or $\mathbb{TGF}$ in normal form contains for every term (i.e. type variable or constant symbol) t, t_1, t_2 from $\mathfrak{t}$ the atoms $\mathtt{D}(t)$ or $\mathtt{Univ}(t_1, t_2)$ depending on whether Φ is $\mathbb{GFO}_=$ or $\mathbb{TGF}$. This presupposes that Φ is in the respective normal form (see Lemma 70 or Corollary 71 respectively).

Note also, that by our types being maximally consistent each $(\mathfrak{t}, \nu) \in T_i^{(\mathcal{E}_+, \mathcal{E}_!)}$ contains automatically some disjunct from $\nu(\vartheta_i)$.

Lastly, if Φ is from $\mathbb{TGF}$ it does not really matter, whether we state that $\mathcal{E}$ consists of rigid types or not, since $\mathbb{TGF}$ expressly disallows equality atoms.

Definition 73. *Let Φ be given as a $\mathbb{GFO}_=$ or $\mathbb{TGF}$ τ-sentence. Additionally let $\mathrm{width}(\Phi)$ be the maximal arity of a relation symbol appearing in Φ and $\mathcal{E}$ be the set of all guarded types that are also rigid and of order $\leq \mathrm{width}(\Phi)$. We will denote by σ the set of fresh relation symbols $\mathtt{Tp}_{\mathfrak{t}}$ of arity $\mathrm{order}(\mathfrak{t})$ for every $\mathfrak{t} \in \mathcal{E}$. Furthermore let $(\mathcal{E}_+, \mathcal{E}_!) \in (\!|\Phi|\!)$.*

Let $C_{/\sim}$ denote the factorization of the set C of constant symbols from τ by the equivalence relation $\sim$ that is defined by letting $\mathtt{c} \sim \mathtt{d}$ hold whenever the literal $\mathtt{c} = \mathtt{d}$ is contained in some $\mathfrak{t} \in \mathcal{E}_+$. We obtain the C-structure $\mathfrak{C}$ by taking $C_{/\sim}$ as domain and letting $\mathtt{c}^{\mathfrak{C}} = [\mathtt{c}]_{\sim}$ for any $\mathtt{c} \in C$.
Furthermore, the $\sigma \uplus C$-structure $\mathfrak{C} \cdot \breve{\sigma}$ (the extension of $\mathfrak{C}$ to the signature $\sigma \uplus C$ by interpreting the relation symbols from σ as empty) will sometimes also be denoted with $\mathfrak{C}$. In this case we will make clear, or it is clear from the context, that we talk about the $\sigma \uplus C$-structure and not just the C-structure.

For $\mathfrak{t} \in \mathcal{E}_+$, we denote by $\mathfrak{D}_{\mathfrak{t}}$ the relation-minimal σ-structure with domain $D_{\mathfrak{t}} = \boldsymbol{v}_{\mathfrak{t}}$ satisfying $\boldsymbol{v}_{\mathfrak{t}} \in \mathtt{Tp}_{\mathfrak{t}}^{\mathfrak{D}_{\mathfrak{t}}}$ as well as all further σ-relations implied by the conjuncts from 66 and 67.

We define for every $\mathfrak{t} \in \mathcal{E}_+$ a $\tau \uplus \sigma$-structure $(\mathfrak{D}_{\mathfrak{t}})^{\tau}$ with domain $(D_{\mathfrak{t}})^{\tau} = D_{\mathfrak{t}} \uplus C_{/\sim}$ as follows:

- $\mathtt{c}^{(\mathfrak{D}_{\mathfrak{t}})^{\tau}} = [\mathtt{c}]_{\sim}$ for $\mathtt{c} \in C$,
- $\mathtt{T}^{(\mathfrak{D}_{\mathfrak{t}})^{\tau}} = \mathtt{T}^{\mathfrak{D}_{\mathfrak{t}}}$ for every $\mathtt{T} \in \sigma$, and
- $\mathtt{R}^{(\mathfrak{D}_{\mathfrak{t}})^{\tau}} = \{\boldsymbol{b} \in ((D_{\mathfrak{t}})^{\tau})^{\mathrm{ar}(\mathtt{R})} \mid \mathtt{R}(\boldsymbol{b}) \in \mathfrak{t}\}$ for every $\mathtt{R} \in$

$\tau \setminus C$.

Let $\mathfrak{K}_{\mathcal{E}_+}$ be the $\sigma \uplus C$-structure that is the disjoint union of $\mathfrak{C} \cdot \breve{\sigma}$ and all $\mathfrak{D}_t$ for $t \in \mathcal{E}_+$. We assume that taking the disjoint union includes the renaming of all domain elements of every $\mathfrak{D}_t$ with fresh, nowhere-else occurring element names, prior to taking the union. We denote by $f_t^{\mathfrak{K}_{\mathcal{E}_+}} : \mathfrak{D}_t \to \mathfrak{K}_{\mathcal{E}_+}$ the injection that maps $\mathfrak{D}_t$ to its isomorphic copy. Additionally, we extend $f_t^{\mathfrak{K}_{\mathcal{E}_+}}$ to a map $f_{t,\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_+}} : \mathfrak{D}_t \uplus (\mathfrak{C} \cdot \breve{\sigma}) \to \mathfrak{K}_{\mathcal{E}_+}$ with $f_{t,\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_+}}(c^{\mathfrak{C} \cdot \breve{\sigma}}) = c^{\mathfrak{K}_{\mathcal{E}_+}}$ for all $c \in C$.

We will make some notes on Definition 73.

1) For each $\mathcal{E}_+$ we note that if some $t \in \mathcal{E}_+$ contains a literal whose terms are just from C (so no variables are mentioned), i.e. a closed literal, then this literal is contained in every $t \in \mathcal{E}_+$. This follows from the fact that first, our types are maximal consistent and second that $\mathcal{E}_+$ arises as a component of the type summary of a given model $\mathfrak{A} \in [\![\Phi]\!]$. More specifically, let Θ be a literal having as terms just constants (hence being a sentence) and let $t \in \mathcal{E}_+$ with $\Theta \in t$. Let $t' \in \mathcal{E}_+$ some arbitrary, different type, and take $\mathfrak{A} \in [\![\Phi]\!]$ such that $\mathcal{E}(\mathfrak{A}) = (\mathcal{E}_+, \mathcal{E}_!)$. By $t \in \mathcal{E}_+$ we conclude that $\mathfrak{A} \models \Theta$. Suppose, by maximal consistency of types, $\neg\Theta \in t'$. As also $t' \in \mathcal{E}_+$ we obtain $\mathfrak{A} \models \neg\Theta$, a contradiction. Hence every closed literal that is contained in some $t \in \mathcal{E}_+$ automatically is contained in every $t \in \mathcal{E}_+$. An example for this is the closed literal $c = d$ (which assumes Φ to be a $\mathbb{GFO}_=$ sentence).

2) If Φ is from $\mathbb{TGF}$, rigidity always holds for types. And especially, in the definition of $\sim$ we obtain that $C_{/\sim}$ is, up to renaming, equal to C (note that the former is a set of equivalence classes, whereas the latter is not).

3) Note that when we defined $\mathfrak{D}_t$ for some $t \in \mathcal{E}_+$ we made use of (type) variables (which are *syntactic* objects) as domain elements (which are *semantic* objects). This shall not cause any confusion, especially as it is a convenient and intuitive choice for our domain elements. Even more important, the usage of such $\mathfrak{D}_t$ will be only temporary and always followed by a renaming of its domain elements.

4) In the definition of $\mathfrak{D}_t$ we closed $\mathfrak{D}_t$ under taking variable permutations and (valid) subtypes. Note that this plays out on the level of the type predicates and not the types themselves.

5) For $t \in \mathcal{E}_+$ the structure $(\mathfrak{D}_t)^\tau$ satisfies for every $t' \in \mathcal{E}_+$ an equivalence between the set of tuples $\boldsymbol{d}$ from $(\mathfrak{D}_t)^\tau$ satisfying Tp_t and the set of tuples $\boldsymbol{d}$ realizing t' in $(\mathfrak{D}_t)^\tau$. As this is an important fact, we will prove it separately in Lemma 74.

6) For $t \in \mathcal{E}_+$ the structure $(\mathfrak{D}_t)^\tau$ satisfies the following properties: every tuple $\boldsymbol{d}$ satisfying some $\mathrm{Tp}_{t'}$ in $(\mathfrak{D}_t)^\tau$ consists of pairwise distinct elements from $\boldsymbol{v}_t$.

7) When we defined $\mathfrak{K}_{\mathcal{E}_+}$, note that by including $\mathfrak{C}$ in the disjoint union, we add all domain elements denoted by

constants. This part of the model does not participate in any Tp_t relations, due to definition. For our construction later on this will be no issue: The rigidity of the types from $\mathcal{E}$ enforces that the elements from $\mathfrak{C}$ do not participate in any Tp_t relations.

8) As $f_t^{\mathfrak{K}_{\mathcal{E}_+}}$ is injective, so is $f_{t,\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_+}}$ by definition (as we only additionally map bijectively $\mathfrak{C}$ to its copy in $\mathfrak{K}_{\mathcal{E}_+}$). Additionally, $f_{t,\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_+}}$ is an embedding. That it is a homomorphism is clear by definition. So let $T \in \sigma$ and $\boldsymbol{d} \in D_t \uplus C_{/\sim}$ such that $f_{t,\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_+}}(\boldsymbol{d}) \in T^{\mathfrak{K}_{\mathcal{E}_+}}$. As the $\mathfrak{C}$ part of $\mathfrak{K}_{\mathcal{E}_+}$ does not participate in any σ-relations we obtain by definition of $f_{t,\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_+}}$ that $\boldsymbol{d} \in D_t$. And by definition of $\mathfrak{K}_{\mathcal{E}_+}$ we conclude $\boldsymbol{d} \in T^{\mathfrak{D}_t}$ and therefore $\boldsymbol{d} \in T^{\mathfrak{D}_t \uplus (\mathfrak{C} \cdot \breve{\sigma})}$.

In the following lemmata (Lemma 74 and Lemma 75) we will introduce results that essentially allow us to switch between the notion of realizing a type t from some set of eligible types $\mathcal{E}$ and satisfying the associated type predicates Tp_t. As we will start from structures whose relations are just interpretations of type predicates, the general idea is the following: For each tuple $\boldsymbol{a}$ satisfying a type predicate Tp_t, we will look into the associated type t and, according to the information therein, we will introduce τ-relations such that $\boldsymbol{a}$ also realizes t. At this stage this will be restricted to the structures $(\mathfrak{D}_t)^\tau$ and an expanded version of $\mathfrak{K}_{\mathcal{E}_+}$. In the first case, we actually only have to check that the general idea already holds. In the second case we expand the structure according to the general idea. Those might be simple cases but they lie at the heart of our construction we will employ to prove the characterizations of the homclosure for $\mathbb{GFO}_=$ or $\mathbb{TGF}$ sentences.

Lemma 74. *Let Φ be given as a $\mathbb{GFO}_=$ or $\mathbb{TGF}$ τ-sentence. Additionally let $\mathrm{width}(\Phi)$ be the maximal arity of a relation symbol appearing in Φ and $\mathcal{E}$ be the set of all guarded types that are also rigid and of order $\leq \mathrm{width}(\Phi)$. We will denote by σ the set of fresh relation symbols Tp_t of arity $\mathrm{order}(t)$ for every $t \in \mathcal{E}$. Furthermore let $(\mathcal{E}_+, \mathcal{E}_!) \in (\![\Phi]\!)$.*

Then for all $t' \in \mathcal{E}$ it holds that a tuple $\boldsymbol{d}$ from D_t satisfies $\mathrm{Tp}_{t'} \in \sigma$ in $(\mathfrak{D}_t)^\tau$ if and only if $\boldsymbol{d}$ realizes t' in $(\mathfrak{D}_t)^\tau$.

Proof. Let $\boldsymbol{d}$ be a tuple in $(\mathfrak{D}_t)^\tau$ such that $\boldsymbol{d} \in \mathrm{Tp}_{t'}^{(\mathfrak{D}_t)^\tau}$ for some $t' \in \mathcal{E}$. There is a permutation $\kappa: \boldsymbol{v} \hookrightarrow \boldsymbol{v}$ with $\kappa t'$ being a subtype of t. By definition of $(\mathfrak{D}_t)^\tau$ every (positive) literal, i.e. atom $\delta[\boldsymbol{v}_{t'}] \in t'$ is satisfied by $\boldsymbol{d}$ in $(\mathfrak{D}_t)^\tau$. Note that the atom δ might contain constant symbols that we do not mention explicitly. On the other hand for negative literals, suppose $\neg\delta[\boldsymbol{v}_{t'}] \in t'$. Hence $\neg\delta[\boldsymbol{v}_{t'}] \in t$ and by t being a type, thus maximally consistent, it cannot contain the non-negated atom $\delta[\boldsymbol{v}_{t'}]$. Thus by definition, $\boldsymbol{d}$ does not satisfy $\delta[\boldsymbol{v}_{t'}]$ and consequently satisfying the negation of it, i.e. $\neg\delta[\boldsymbol{v}_{t'}]$. Hence we obtain that $\boldsymbol{d}$ realizes t' in $(\mathfrak{D}_t)^\tau$.

Now let $\boldsymbol{d}$ from $(\mathfrak{D}_t)^\tau$ satisfy $t' \in \mathcal{E}_+$. By definition, every literal $\lambda[\boldsymbol{v}_{t'}] \in t'$ is satisfied by $\boldsymbol{d}$. Trivially by assumption $\lambda[\boldsymbol{v}_{t'}] \in t$. Hence there is a permutation $\kappa: \boldsymbol{v} \hookrightarrow \boldsymbol{v}$ such that $\kappa t|_{\boldsymbol{v}_{t'}} = t'$. Without loss of generality we may assume κ being

the identity and $\boldsymbol{v}_{\mathfrak{t}'} \subseteq \boldsymbol{v}_{\mathfrak{t}}$. As the tuple $\boldsymbol{v}_{\mathfrak{t}}$ satisfies $\mathtt{Tp}_{\mathfrak{t}}$ in $\mathfrak{D}_{\mathfrak{t}}$ and the structure satisfies both 66 and 67 we thus obtain via $\mathfrak{t}' = \mathfrak{t}|_{\boldsymbol{v}_{\mathfrak{t}'}}$ that $\boldsymbol{d} \in \mathtt{Tp}_{\mathfrak{t}'}^{(\mathfrak{D}_{\mathfrak{t}})^{\tau}}$. $\qquad\square$

Lemma 75. *Let* Φ *be given as a* $\mathbb{GFO}_{=}$ *or* $\mathbb{TGF}$ τ*-sentence. Additionally let* $\mathrm{width}(\Phi)$ *be the maximal arity of a relation symbol appearing in* Φ *and* $\mathcal{E}$ *be the set of all guarded types that are also rigid and of order* $\leq \mathrm{width}(\Phi)$. *We will denote by* σ *the set of fresh relation symbols* $\mathtt{Tp}_{\mathfrak{t}}$ *of arity* $\mathrm{order}(\mathfrak{t})$ *for every* $\mathfrak{t} \in \mathcal{E}$. *Furthermore let* $(\mathcal{E}_{+}, \mathcal{E}_{!}) \in (\!|\Phi|\!)$.

Then there is a canonical $\sigma \uplus \tau$*-expansion* $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$ *of* $\mathfrak{K}_{\mathcal{E}_{+}}$, *such that for all* $\mathfrak{t} \in \mathcal{E}$ *it holds that a tuple* $\boldsymbol{a}$ *from* $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$ *satisfies* $\mathtt{Tp}_{\mathfrak{t}} \in \sigma$ *in* $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$ *if and only if* $\boldsymbol{a}$ *realizes* $\mathfrak{t}$ *in* $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$.

Additionally, for every $\mathfrak{t} \in \mathcal{E}$ *the map* $f_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}$ *gives rise to an embedding* $(f')_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}} : (\mathfrak{D}_{\mathfrak{t}})^{\tau} \to (\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$, *being the same map as* $f_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}$.

Proof. For every $\mathfrak{t} \in \mathcal{E}$ we use the maps $f_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}$ to expand $\mathfrak{K}_{\mathcal{E}_{+}}$ in the following way: Note first, that $(\mathfrak{D}_{\mathfrak{t}})^{\tau}$ has the same domain as $\mathfrak{D}_{\mathfrak{t}} \uplus (\mathfrak{C} \cdot \breve{\sigma})$. For $\mathtt{R} \in \tau \setminus C$ we let

$$\mathtt{R}^{(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}} = \{ f_{\mathfrak{t}}^{\mathfrak{K}_{\mathcal{E}_{+}}}(\boldsymbol{d}) \mid \mathfrak{t} \in \mathcal{E}_{+} \text{ and } \boldsymbol{d} \in \mathtt{R}^{(\mathfrak{D}_{\mathfrak{t}})^{\tau}} \}. \quad (111)$$

By definition for different $\mathfrak{t}, \mathfrak{t}' \in \mathcal{E}_{+}$ the sets $f_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}(D_{\mathfrak{t}})$ and $f_{\mathfrak{t}',\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}(D_{\mathfrak{t}'})$ are disjoint. Hence for every tuple $\boldsymbol{a} \in \mathtt{R}^{(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}}$ containing at least one element not named by a constant there is a unique $\mathfrak{t} \in \mathcal{E}_{+}$ such that $\boldsymbol{a} = f_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}(\boldsymbol{d})$ for a tuple $\boldsymbol{d}$ consisting exclusively of elements from $\mathfrak{C}$ and $\mathfrak{D}_{\mathfrak{t}}$. If $\boldsymbol{a}$ consists only of elements named by constants, we remember Remark 1 after Definition 73 from which we immediately conclude, that it does not matter which $\mathfrak{t} \in \mathcal{E}_{+}$ we choose in the discussion. Hence in this case we fix one. By $f_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}$ being injective, this $\boldsymbol{d}$ is also unique. Additionally, $f_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}$ gives rise to an embedding $(f')_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}} : (\mathfrak{D}_{\mathfrak{t}})^{\tau} \to (\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$, being the same map as $f_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}$ considering both as maps between the underlying domains of the structures. For simplicity of exposition we will denote it by just $f_{\mathfrak{t}}'$. By the interpretation of the $\mathtt{R} \in \tau \setminus C$ in $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$ as defined in 111 the map $f_{\mathfrak{t}}'$ is a homomorphism. $f_{\mathfrak{t}}'$ is injective since $f_{\mathfrak{t},\mathfrak{c}}^{\mathfrak{K}_{\mathcal{E}_{+}}}$ is. So let $\boldsymbol{d}$ be a tuple from $(\mathfrak{D}_{\mathfrak{t}})^{\tau}$ such that $f_{\mathfrak{t}}'(\boldsymbol{d}) \in \mathtt{R}^{(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}}$. Then by definition $\boldsymbol{d} \in \mathtt{R}^{(\mathfrak{D}_{\mathfrak{t}})^{\tau}}$, establishing that $f_{\mathfrak{t}}'$ is an embedding.

We now show that for every type $\mathfrak{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ from $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$ satisfies $\mathtt{Tp}_{\mathfrak{t}'}^{(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}}$ in $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$ if and only if $\boldsymbol{a}$ realizes $\mathfrak{t}'$ in $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$. Note that by rigidity of types on the one hand and Remark 7 after Definition 73 on the other, only tuples that solely consists of elements not named by constants can satisfy type predicates or realize types.

So let $\boldsymbol{a} \in \mathtt{Tp}_{\mathfrak{t}'}^{(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}}$ for some $\mathfrak{t}' \in \mathcal{E}$. As by definition the domain of $\mathfrak{K}_{\mathcal{E}_{+}}$ is the disjoint union of $\mathfrak{C} \cdot \breve{\sigma}$ and isomorphic copies of $\mathfrak{D}_{\mathfrak{t}}$ there exists a unique $\mathfrak{t} \in \mathcal{E}_{+}$ such that $\boldsymbol{a}$ is from $f_{\mathfrak{t}}'(D_{\mathfrak{t}})$. Since $f_{\mathfrak{t}}'$ is an embedding, we find $\boldsymbol{d}$ in $D_{\mathfrak{t}}$ with

$f_{\mathfrak{t}}'(\boldsymbol{d}) = \boldsymbol{a}$ and $\boldsymbol{d} \in \mathtt{Tp}_{\mathfrak{t}'}^{(\mathfrak{D}_{\mathfrak{t}})^{\tau}}$. Lemma 74 yields that $\boldsymbol{d}$ realizes $\mathfrak{t}'$ in $(\mathfrak{D}_{\mathfrak{t}})^{\tau}$. By invoking $f_{\mathfrak{t}}'$ being an embedding once again, we obtain that $\boldsymbol{a} = f_{\mathfrak{t}}'(\boldsymbol{d})$ realizes $\mathfrak{t}'$ in $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$.

Suppose now that $\boldsymbol{a}$ realizes $\mathfrak{t}'$ for some $\mathfrak{t}' \in \mathcal{E}$. As $\mathfrak{t}'$ is guarded, there exists a guard atom δ (witnessed by ν) such that $\boldsymbol{a}$ satisfies $\nu(\delta)$ in $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$. By 111 this necessitates, that there is a $\mathfrak{t}'' \in \mathcal{E}$ such that $\boldsymbol{a}$ consists of elements exclusively from $f_{\mathfrak{t}''}'(D_{\mathfrak{t}''})$. More importantly, $\boldsymbol{a}$ cannot consist of images from different $D_{\mathfrak{t}''}$ via $f_{\mathfrak{t}''}'$. But since $\boldsymbol{a}$ realizes $\mathfrak{t}'$ we find some $\mathfrak{t} \in \mathcal{E}_{+}$ and a unique $\boldsymbol{d}$ from $D_{\mathfrak{t}}$ with $f_{\mathfrak{t}}'(\boldsymbol{d}) = \boldsymbol{a}$ such that $\boldsymbol{d}$ also realizes $\mathfrak{t}'$ (using that $f_{\mathfrak{t}}'$ is an embedding). By Lemma 74 we obtain that $\boldsymbol{d} \in \mathtt{Tp}_{\mathfrak{t}'}^{(\mathfrak{D}_{\mathfrak{t}})^{\tau}}$ and by $f_{\mathfrak{t}}'$ being an embedding we conclude $\boldsymbol{a} \in \mathtt{Tp}_{\mathfrak{t}'}^{(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}}$. $\qquad\square$

As we are concerned with the homclosure of $\mathbb{GFO}_{=}$ or $\mathbb{TGF}$ sentences, we need to start shedding light on the behaviour of our ("enriched") structures under homomorphisms into some other structures. It is natural to start with the simple case of $\mathfrak{K}_{\mathcal{E}_{+}}$ (or, for the "enriched" structure $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$). The structures to which we will homomorphically map $\mathfrak{K}_{\mathcal{E}_{+}}$ (or $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$) cannot be arbitrary ones but have to already satisfy some sentence, which will be part of the sentence characterizing the homclosure. This is what the following lemma (Lemma 76) does.

Lemma 76. *Let* Φ *be a* $\mathbb{GFO}_{=}$ *or* $\mathbb{TGF}$ τ*-sentence with* C *being the set of constants in* τ, $\mathcal{E}$ *a set of eligible types (with the usual provision of being finite and containing rigid types) closed under taking variable permutations and* $(\mathcal{E}_{+}, \mathcal{E}_{!}) \in (\!|\Phi|\!)$. *Let* σ *be the signature containing for every* $\mathfrak{t} \in \mathcal{E}$ *the predicate* $\mathtt{Tp}_{\mathfrak{t}}$ *having as its arity the order of* $\mathfrak{t}$. *Then* $\mathfrak{K}_{\mathcal{E}_{+}}$ *satisfies* $\Psi_{\mathrm{gen},(\mathcal{E}_{+},\mathcal{E}_{!})}$ *with the additional property that for every model* $\mathfrak{B}$ *(over the signature* $\tau \uplus \sigma$) *of* $\Psi_{\mathrm{hom}} \wedge \Psi_{\mathrm{gen},(\mathcal{E}_{+},\mathcal{E}_{!})}$ *there exists a homomorphism* $h \colon \mathfrak{K}_{\mathcal{E}_{+}} \to \mathfrak{B}|_{\sigma \uplus C}$.
Additionally, h *gives rise to a homomorphism* $h' \colon (\mathfrak{K}_{\mathcal{E}_{+}})^{\tau} \to \mathfrak{B}$ *and* $(\mathfrak{K}_{\mathcal{E}_{+}})^{\tau}$ *satisfies* $\Psi_{\mathrm{hom}} \wedge \Psi_{\mathrm{gen},(\mathcal{E}_{+},\mathcal{E}_{!})}$.

Proof. We show that $\mathfrak{K}_{\mathcal{E}_{+}}$ satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_{+},\mathcal{E}_{!})}$. Since by definition $\mathfrak{K}_{\mathcal{E}_{+}}$ contains for every $\mathfrak{t} \in \mathcal{E}_{+}$ a substructure isomorphic to $\mathfrak{D}_{\mathfrak{t}}$, it contains some tuple $\boldsymbol{a}$ such that $\boldsymbol{a} \in \mathtt{Tp}_{\mathfrak{t}}^{\mathfrak{K}_{\mathcal{E}_{+}}}$. As this holds for every $\mathfrak{t} \in \mathcal{E}_{+}$ Formulae 64 is satisfied.

To show the satisfaction of Formulae 65 observe that in the construction we either added for $\mathfrak{t} \in \mathcal{E}_{+}$ a $\mathfrak{D}_{\mathfrak{t}}$ (hence ensuring the satisfaction of some $\mathtt{Tp}_{\mathfrak{t}}$) or we closed the disjunction ultimately yielding $\mathfrak{K}_{\mathcal{E}_{+}}$ under 66 and 67. Consequently, only by these closure operations could we unwittingly made a $\mathtt{Tp}_{\mathfrak{t}}$ with $\mathfrak{t} \in \mathcal{E} \setminus \mathcal{E}_{+}$ true for some tuple in $\mathfrak{K}_{\mathcal{E}_{+}}$. But this cannot have happened, as by the fact that $\mathcal{E}$ is closed under taking variable permutations and applying Lemma 67 (both Item 1 and 2) we only could have introduced $\mathtt{Tp}_{\mathfrak{t}}$ for $\mathfrak{t} \in \mathcal{E}_{+}$. Consequently, for every $\mathfrak{t} \in \mathcal{E} \setminus \mathcal{E}_{+}$ we obtain $\mathtt{Tp}_{\mathfrak{t}}^{\mathfrak{K}_{\mathcal{E}_{+}}} = \emptyset$, implying the validity of Formulae 65.

Now take $\mathfrak{t}, \mathfrak{t}' \in \mathcal{E}_{+}$ such that $\mathfrak{t}' = \mathfrak{t}|_{\boldsymbol{v}_{\mathfrak{t}'}}$ and let $\boldsymbol{a}$ be some tuple from $\mathfrak{K}_{\mathcal{E}_{+}}$ such that $\boldsymbol{a} \in \mathtt{Tp}_{\mathfrak{t}}^{\mathfrak{K}_{\mathcal{E}_{+}}}$, let $\mu \colon \boldsymbol{v}_{\mathfrak{t}} \to \boldsymbol{a}$ be

the componentwise evaluation witnessing this. By definition of $\mathfrak{K}_{\mathcal{E}_+}$ there exists some type $\mathfrak{t}^* \in \mathcal{E}_+$ with $\mathfrak{t}^*|_{v_\mathfrak{t}} = \mathfrak{t}$, such that the induced substructure by $\boldsymbol{a}$ in $\mathfrak{K}_{\mathcal{E}_+}$ is an induced one of an isomorphic copy of $\mathfrak{D}_{\mathfrak{t}^*}$. Hence we know by definition (closure with respect to eligible subtypes in $\mathfrak{D}_\mathfrak{t}$) that the subtuple $\mu(\boldsymbol{v}_{\mathfrak{t}'})$ of $\boldsymbol{a}$ satisfies $\mu(\boldsymbol{v}_{\mathfrak{t}'}) \in \mathrm{Tp}_{\mathfrak{t}'}^{\mathfrak{K}_{\mathcal{E}_+}}$. Finally, observing that μ assigns the tuple $\boldsymbol{a}$ componentwise to $\boldsymbol{v}_\mathfrak{t}$, this yields the satisfaction of Formulae 66 by $\mathfrak{K}_{\mathcal{E}_+}$.

To show the satisfaction of the last kind of Formulae (i.e. Formulae 67), let $\mathfrak{t} \in \mathcal{E}$, $\kappa\colon \boldsymbol{v} \hookrightarrow \boldsymbol{v}$ a permutation of the variables and $\boldsymbol{a}$ a tuple from $\mathfrak{K}_{\mathcal{E}_+}$ such that $\boldsymbol{a} \in \mathrm{Tp}_\mathfrak{t}^{\mathfrak{K}_{\mathcal{E}_+}}$. Note that, since $\mathcal{E}$ is closed under taking permutations of (type) variables, $\kappa\mathfrak{t} \in \mathcal{E}$. And as for every $\mathfrak{t} \in \mathcal{E} \backslash \mathcal{E}_+$ the interpretation $\mathrm{Tp}_\mathfrak{t}^{\mathfrak{K}_{\mathcal{E}_+}}$ is empty, it suffices to discuss the case when $\mathfrak{t} \in \mathcal{E}_+$. By definition of $\mathfrak{K}_{\mathcal{E}_+}$ there exists some type $\mathfrak{t}^* \in \mathcal{E}_+$ and some permutation $\kappa^*\colon \boldsymbol{v} \hookrightarrow \boldsymbol{v}$ with $\kappa^*\mathfrak{t}^* = \mathfrak{t}$ such that the induced substructure by $\boldsymbol{a}$ in $\mathfrak{K}_{\mathcal{E}_+}$ is an induced one of an isomorphic copy of $\mathfrak{D}_{\mathfrak{t}^*}$. Since every $\mathfrak{D}_\mathfrak{t}$ used in the definition of $\mathfrak{K}_{\mathcal{E}_+}$ is itself closed under 67 (i.e. closed under permutations, which is a fact holding by definition) we obtain that $\eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_\mathfrak{t}}(\boldsymbol{a}) \in \mathrm{Tp}_{\kappa\mathfrak{t}}^{\mathfrak{K}_{\mathcal{E}_+}}$ holds. As this shows the validity of the Formulae 67 we finally have shown that $\mathfrak{K}_{\mathcal{E}_+}$ satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$.

Furthermore, for every $\mathfrak{B} \in [\![\Psi_{\mathrm{hom}} \wedge \Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}]\!]$ the structure $\mathfrak{K}_{\mathcal{E}_+}$ has a homomorphism h into the structure $\mathfrak{B}|_{\sigma \uplus C}$, which we will denote by $\mathfrak{B}^*$. To see this, note that $\mathfrak{K}_{\mathcal{E}_+}$ is the disjoint union of $\mathfrak{C}\cdot\breve{\sigma}$ and structures isomorphic to $\mathfrak{D}_\mathfrak{t}$ for $\mathfrak{t} \in \mathcal{E}_+$, each of which in separation homomorphically maps to $\mathfrak{B}^*$. As for $\mathfrak{C}$, the mapping $[\mathsf{c}]_\sim \mapsto \mathsf{c}^{\mathfrak{B}^*}$ does the trick since for any $\mathsf{c},\mathsf{d} \in C$ with $\mathsf{c} \sim \mathsf{d}$, the fact that $\mathfrak{B}$ is a model of Ψ_{hom} and realizes some $\mathrm{Tp}_\mathfrak{t}$ containing the literal $\mathsf{c}{=}\mathsf{d}$ ensures $\mathsf{c}^\mathfrak{B} = \mathsf{d}^\mathfrak{B}$. For Φ from $\mathbb{TGF}$ this comment is not interesting since $\mathsf{c} \sim \mathsf{d}$ if and only if c and d are the same constant symbols. We can now extend this map stepwise for every component isomorphic to some $\mathfrak{D}_\mathfrak{t}$ of the disjoint union. Because of isomorphy it is enough to discuss $\mathfrak{D}_\mathfrak{t}$ instead of the actual component. By definition each $\mathfrak{D}_\mathfrak{t}$ satisfies a conjunct from 64 saturated under 66 and 67. As $\mathfrak{B} \in [\![\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}]\!]$ we can map the elements $\boldsymbol{v}_\mathfrak{t}$ of $\mathfrak{D}_\mathfrak{t}$ componentwise to some arbitrary $\boldsymbol{b} \in \mathrm{Tp}_\mathfrak{t}^\mathfrak{B}$. Then the extension of the homomorphism is obtained by mapping the tuple $f_\mathfrak{t}^{\mathfrak{K}_{\mathcal{E}_+}}(\boldsymbol{v}_\mathfrak{t})$ componentwise to $\boldsymbol{b}$. As we took the disjoint union, we obtain a homomorphism $h\colon \mathfrak{K}_{\mathcal{E}_+} \to \mathfrak{B}^*$.

Finally we show that h (as a map) constitutes a homomorphism $h'\colon (\mathfrak{K}_{\mathcal{E}_+})^\tau \to \mathfrak{B}$. As h is a homomorphism when restricting to the signature $\sigma \uplus C$, it suffices to discuss $\mathrm{R} \in \tau \setminus C$. Furthermore, we will make use of the embedding $(f')_{\mathfrak{t},\mathfrak{C}}^{\mathfrak{K}_{\mathcal{E}_+}}$ obtained from Lemma 75 (for $\mathfrak{t} \in \mathcal{E}_+$). For brevity, we will denote those maps by just $f_\mathfrak{t}'$. Let $\boldsymbol{a} \in \mathrm{R}^{(\mathfrak{K}_{\mathcal{E}_+})^\tau}$ for some $\mathrm{R} \in \tau \setminus C$. If $\boldsymbol{a}$ contains at least one element not named by a constant there is a unique $\mathfrak{t} \in \mathcal{E}_+$ and a unique $\boldsymbol{d}$ from $\boldsymbol{v}_\mathfrak{t}$ such that $\boldsymbol{d} \in \mathrm{R}^{(\mathfrak{D}_\mathfrak{t})^\tau}$ and $f_\mathfrak{t}'(\boldsymbol{d}) = \boldsymbol{a}$. If $\boldsymbol{a}$ consists of only elements named by constants, we remember Remark 1 after Definition 73 from which we immediately conclude, that it does not matter which $\mathfrak{t} \in \mathcal{E}_+$ we choose in the

discussion. Hence in this case we fix one. Let $\boldsymbol{x}$ be a tuple of pairwise different variables of length $\mathrm{length}(\boldsymbol{d})$ and $\mu\colon \boldsymbol{x} \to \boldsymbol{d}$ componentwise. Then $\mathrm{R}(\boldsymbol{x})$ is satisfied by $\boldsymbol{d}$ as witnessed by μ. Furthermore let $\boldsymbol{v}_{\boldsymbol{d}}$ be the subtuple of $\boldsymbol{d}$ consisting of elements from $\boldsymbol{v}_\mathfrak{t}$, i.e. $\boldsymbol{v}_{\boldsymbol{d}} = \boldsymbol{d} \cap \boldsymbol{v}_\mathfrak{t}$. Take $\delta(\boldsymbol{v}_{\boldsymbol{d}})$ to be the atom obtained from $\mathrm{R}(\boldsymbol{d})$ where each element of the form $[\mathsf{c}]_\sim$ for $\mathsf{c} \in C$ appearing in $\boldsymbol{d}$ is substituted by c. The choice of the representative does only matter in so far as we will not obtain a unique δ. Nevertheless this non-uniqueness will have no impact on the proof. Note for this that $\sim$ is defined on C in terms of equality atoms $\mathsf{c} = \mathsf{d}$ in the types $\mathfrak{t} \in \mathcal{E}_+$. And by the note after Definition 73, realization of the types from $\mathcal{E}_+$ automatically enforces this equivalence relation. As $\boldsymbol{v}_\mathfrak{t}$ realizes $\mathfrak{t}$ in $(\mathfrak{D}_\mathfrak{t})^\tau$, $\boldsymbol{v}_\mathfrak{t} \in \mathrm{Tp}_\mathfrak{t}^{(\mathfrak{D}_\mathfrak{t})^\tau}$ by Lemma 74. As $\delta \in \mathfrak{t}$ the subtype $\mathfrak{t}' = \mathfrak{t}|_{\boldsymbol{v}_{\boldsymbol{d}}}$ is guarded by δ (as witnessed by $\nu\colon \boldsymbol{v}_{\boldsymbol{d}} \to \boldsymbol{v}_{\mathfrak{t}'} \uplus C, v \mapsto v$). Hence $\mathfrak{t}' \in \mathcal{E}_+$ by Lemma 67 and by $\mathfrak{D}_\mathfrak{t}$ (hence $(\mathfrak{D}_\mathfrak{t})^\tau$) satisfying 66 (by definition) we obtain that $\boldsymbol{v}_{\boldsymbol{d}} \in \mathrm{Tp}_{\mathfrak{t}'}^{(\mathfrak{D}_\mathfrak{t})^\tau}$. Applying $f_\mathfrak{t}'$ yields $\boldsymbol{a}^* \in \mathrm{Tp}_{\mathfrak{t}'}^{(\mathfrak{K}_{\mathcal{E}_+})^\tau}$ with $f_\mathfrak{t}'(\boldsymbol{v}_{\boldsymbol{d}}) = \boldsymbol{a}^*$, where $\boldsymbol{a}^*$ is the subtuple of $\boldsymbol{a}$ consisting of all elements that are no interpretations of constant symbols. Hence $h'(\boldsymbol{a}^*) \in \mathrm{Tp}_{\mathfrak{t}'}^\mathfrak{B}$. By $\mathfrak{B}$ satisfying Ψ_{hom}, $h'(\boldsymbol{a}^*)$ satisfies the atom δ in $\mathfrak{B}$. But this means that $h'(\boldsymbol{a}) \in \mathrm{R}^\mathfrak{B}$ showing that h' is also a homomorphism for the $\tau \uplus \sigma$-structures. $\square$

We will now start to make more precise the sentence Ψ that will characterize the homclosure of some $\mathbb{GFO}_=$ or $\mathbb{TGF}$ sentence Φ. For this we will utilize our normal forms (see Lemma 70 and Corollary 71). Thanks to those normal forms, the $\mathbb{GFO}_=$ and the $\mathbb{TGF}$ case will differ only slightly. This comes as little surprise after observing that our normal form for $\mathbb{TGF}$ sentences is in and of itself "essentially guarded" with the exception of the $\Phi_{\mathtt{Univ}}$ part. Hence the difference in Ψ depending on whether Φ is from $\mathbb{GFO}_=$ or $\mathbb{TGF}$ is connected to $\Phi_{\mathtt{Univ}}$.

We will first define the, more or less, "last puzzle pieces" for Ψ. Additionally we introduce the notion of a violation tuple which will be put to good use in a stepwise construction procedure later on.

Definition 77. *Let Φ be a $\mathbb{GFO}_=$ or $\mathbb{TGF}$ τ-sentence in normal form, i.e. $\Phi = \Phi_{\mathtt{D}} \wedge \Phi_\forall \wedge \Phi_{\forall\exists}$ or $\Phi = \Phi_{\mathtt{Univ}} \wedge \Phi_\forall \wedge \Phi_{\forall\exists}$ respectively, where*

$$\Phi_{\mathtt{D}} = \forall x.\mathrm{D}(x), \tag{112}$$

$$\Phi_{\mathtt{Univ}} = \forall xy.\mathtt{Univ}(x,y), \tag{113}$$

$$\Phi_\forall = \bigwedge_i \forall \boldsymbol{x}.(\alpha_i(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]), \text{ and} \tag{114}$$

$$\Phi_{\forall\exists} = \bigwedge_j \forall \boldsymbol{z}.(\beta_j(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma_j(\boldsymbol{y}\boldsymbol{z})), \tag{115}$$

with ϑ_i a disjunction of literals, α_i, β_j guard atoms, and γ_j an atom. Additionally let $\mathcal{E}$ be the set of all guarded rigid types of order $\leq \mathrm{width}(\Phi)$, and σ be the set of all type predicates

associated to the types from $\mathcal{E}$. Then for every $(\mathcal{E}_+, \mathcal{E}_!)$ we define $\Psi_{\text{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ to be the conjunction of the σ-sentences

$$\forall \boldsymbol{v}_{\mathsf{t}}.\mathsf{Tp}_{\mathsf{t}}(\boldsymbol{v}_{\mathsf{t}}) \Rightarrow \bigvee_{(\mathsf{t},\nu) \Subset_j (\mathsf{t}',\nu')} \exists \boldsymbol{v}_{\mathsf{t}'} \backslash \nu(\boldsymbol{z}).\mathsf{Tp}_{\mathsf{t}'}(\boldsymbol{v}_{\mathsf{t}'}) \tag{116}$$

where (t,ν) ranges over $B_j^{(\mathcal{E}_+,\mathcal{E}_!)}$ and j ranges over the number of conjuncts in $\Phi_{\forall\exists}$

We call $(\boldsymbol{a}, \mu, \mathsf{t}, \nu)$ a violation tuple *of some structure* $\mathfrak{A} \in \llbracket \Phi \rrbracket$ *with* $\mathcal{E}(\mathfrak{A}) = (\mathcal{E}_+, \mathcal{E}_!)$ *if* $\boldsymbol{a}$ *is a tuple from* $\mathfrak{A}$ *and* $(\mathsf{t},\nu) \in B_j^{(\mathcal{E}_+,\mathcal{E}_!)}$ *such that the componentwise variable assignment* $\mu \colon \boldsymbol{v}_{\mathsf{t}} \to \boldsymbol{a}$ *witnesses the violation of the* (t,ν)-*conjunct from* (116) *in the structure* $\mathfrak{A}$, *i.e.,* $(\mathfrak{A},\mu) \vDash \mathsf{Tp}_{\mathsf{t}}$ *and for every* $(\mathsf{t}',\nu') \in P_j^{(\mathcal{E}_+,\mathcal{E}_!)}$ *with* $(\mathsf{t},\nu) \Subset_j (\mathsf{t}',\nu')$ *and every* $\mu' \colon \boldsymbol{v}_{\mathsf{t}'} \to A$ *with* $\mu'|_{\boldsymbol{v}_{\mathsf{t}}} = \mu$ *the condition* $(\mathfrak{A},\mu') \nvDash \mathsf{Tp}_{\mathsf{t}'}$ *holds.*

Additionally, if Φ *is from* $\mathbb{TGF}$, *we define for every* $(\mathcal{E}_+, \mathcal{E}_!)$ *the* σ *sentence* $\Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$ *to be the conjunction of the following sentences for every* $\mathsf{t}, \mathsf{t}' \in \mathcal{E}_+$ *with* $\boldsymbol{v}_{\mathsf{t}} = v$ *and* $\boldsymbol{v}_{\mathsf{t}'} = v'$:

$$\forall vv'.\mathsf{Tp}_{\mathsf{t}}(v) \wedge \mathsf{Tp}_{\mathsf{t}'}(v') \Rightarrow \bigvee_{\substack{\mathsf{t}'' \in \mathcal{E}_+, \ \boldsymbol{v}_{\mathsf{t}''} = vv' \\ \mathsf{t}''|_v = \mathsf{t}, \ \mathsf{t}''|_{v'} = \mathsf{t}'}} \mathsf{Tp}_{\mathsf{t}''}(v,v') \tag{117}$$

For the upcoming lemmata (Lemma 78 to Lemma 83), which record some technical machinery needed, we will fix the following:

Let Φ be a $\mathbb{GFO}_=$ or $\mathbb{TGF}$ τ-sentence in normal form, i.e. $\Phi = \Phi_{\mathsf{D}} \wedge \Phi_{\forall} \wedge \Phi_{\forall\exists}$ or $\Phi = \Phi_{\mathsf{Univ}} \wedge \Phi_{\forall} \wedge \Phi_{\forall\exists}$ respectively, where

$$\Phi_{\mathsf{D}} = \forall x.\mathsf{D}(x), \tag{118}$$

$$\Phi_{\mathsf{Univ}} = \forall xy.\mathsf{Univ}(x,y), \tag{119}$$

$$\Phi_{\forall} = \bigwedge_i \forall \boldsymbol{x}.(\alpha_i(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]), \text{ and} \tag{120}$$

$$\Phi_{\forall\exists} = \bigwedge_j \forall \boldsymbol{z}.(\beta_j(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma_j(\boldsymbol{y}\boldsymbol{z})), \tag{121}$$

with ϑ_i a disjunction of literals, α_i, β_j guard atoms, and γ_j an atom. Additionally let $\mathcal{E}$ be the set of all guarded rigid types of order $\leq \mathrm{width}(\Phi)$, and σ be the set of all type predicates associated to the types from $\mathcal{E}$. Let C be the set of all constants from τ and σ the signature containing for every t a fresh predicate symbol Tp_{t} of arity $\mathrm{order}(\mathsf{t})$. We furthermore fix some $(\mathcal{E}_+, \mathcal{E}_!) \in (\![\Phi]\!)$.

We will prove an additional "preservation" lemma akin to Lemma 68.

Lemma 78. *Then for* $\mathfrak{A} \in \llbracket \Phi \rrbracket$ *the* $\mathcal{E}$-*adornment* $\mathfrak{A} \cdot \mathcal{E}$ *satisfies* Ψ_{hom} *and* $\Psi_{\text{guard},\mathcal{E}(\mathfrak{A})}$. *If* Φ *is from* $\mathbb{TGF}$, $\mathfrak{A} \cdot \mathcal{E}$ *additionally satisfies* $\Psi_{\mathbb{TGF},\mathcal{E}(\mathfrak{A})}$.

Furthermore, for every $\mathfrak{B}$ *a homomorphic codomain of* $\mathfrak{A}$ *via a homomorphism* $h \colon \mathfrak{A} \to \mathfrak{B}$ *the* $\tau \uplus \sigma$-*structure* $\mathfrak{B} \cdot h(\mathfrak{A} \cdot \mathcal{E}|_\sigma)$ *equally satisfies* Ψ_{hom} *and* $\Psi_{\text{guard},\mathcal{E}(\mathfrak{A})}$. *If* Φ *is from* $\mathbb{TGF}$, $\mathfrak{B} \cdot h(\mathfrak{A} \cdot \mathcal{E}|_\sigma)$ *additionally satisfies* $\Psi_{\mathbb{TGF},\mathcal{E}(\mathfrak{A})}$.

Proof. Let $\mathfrak{A} \in \llbracket \Phi \rrbracket$. We set $\mathcal{E}(\mathfrak{A}) = (\mathcal{E}_+, \mathcal{E}_!)$. We first prove everything for Φ from $\mathbb{GFO}_=$ or $\mathbb{TGF}$. We then conclude with the special case that Φ is just from $\mathbb{TGF}$.

Note that $\mathfrak{A} \cdot \mathcal{E}$ trivially satisfies Ψ_{hom} by definition, as $\boldsymbol{a} \in \mathsf{Tp}_{\mathsf{t}}^{\mathfrak{A} \cdot \mathcal{E}}$ if and only if $\boldsymbol{a}$ realizes the type t. Hence it realizes every positive literal in t.

We show that $\mathfrak{A} \cdot \mathcal{E}$ satisfies $\Psi_{\text{guard},\mathcal{E}(\mathfrak{A})}$. For this we will make use of the fact that we assumed Φ to be in normal form (see also Lemma 70). Let $\boldsymbol{a}$ be a tuple from $\mathfrak{A} \cdot \mathcal{E}$ (and hence from $\mathfrak{A}$) and $(\mathsf{t},\nu) \in B_j^{\mathcal{E}(\mathfrak{A})}$ for some j such that $\boldsymbol{a} \in \mathsf{Tp}_{\mathsf{t}}^{\mathfrak{A} \cdot \mathcal{E}}$. Let furthermore $\mu_{\boldsymbol{a}} \colon \boldsymbol{v}_{\mathsf{t}} \to \boldsymbol{a}$ be the componentwise variable assignment witnessing this fact, i.e. $(\mathfrak{A}, \mu_{\boldsymbol{a}}) \vDash \mathsf{Tp}_{\mathsf{t}}$. We extend $\mu_{\boldsymbol{a}}$ to $\mu'_{\boldsymbol{a}} \colon \boldsymbol{v}_{\mathsf{t}} \uplus C \to A$ where for every $\mathsf{c} \in C$ we set $\mu'_{\boldsymbol{a}}(\mathsf{c}) = \mathsf{c}^{\mathfrak{A}}$. By definition $\boldsymbol{a}$ realizes t and thus (by virtue of (t,ν) being in $B_j^{\mathcal{E}(\mathfrak{A})}$) we obtain by componentwise application that $(\mathfrak{A}, \mu'_{\boldsymbol{a}} \circ \nu) \vDash \beta_j(\boldsymbol{z})$. By $\Phi_{\forall\exists}$ there is a supertuple $\boldsymbol{a}^*$ of $\boldsymbol{a}$ with $\nu' \colon \boldsymbol{y}\boldsymbol{z} \to \boldsymbol{v}^* \uplus C$ extending ν (i.e. $\nu'|_{\boldsymbol{z}} = \nu$) and $\mu_{\boldsymbol{a}^*} \colon \boldsymbol{v}^* \uplus C \to A$ (where $\mu_{\boldsymbol{a}^*}(\boldsymbol{v}^*) = \boldsymbol{a}^*$ holds componentwise) extending $\mu'_{\boldsymbol{a}}$ such that $(\mathfrak{A}, \mu_{\boldsymbol{a}^*} \circ \nu') \vDash \gamma_j(\boldsymbol{y}\boldsymbol{z})$. Then $\boldsymbol{a}^*$ has a unique type t' with $\boldsymbol{v}_{\mathsf{t}'} = \boldsymbol{v}^*$ componentwise and t' being guarded by γ_j as witnessed by ν'. So $\mathsf{t}' \in \mathcal{E}$ and hence $(\mathsf{t}',\nu') \in P_j^{\mathcal{E}(\mathfrak{A})}$. Since $\boldsymbol{v}_{\mathsf{t}} \subseteq \boldsymbol{v}_{\mathsf{t}'}$, ν' extends ν, $\mu_{\boldsymbol{a}^*}$ extends $\mu_{\boldsymbol{a}}$ we obtain that $\mathsf{t}'|_{\boldsymbol{v}_{\mathsf{t}}} = \mathsf{t}$, hence $(\mathsf{t},\nu) \Subset_j (\mathsf{t}',\nu')$. By definition of $\mathfrak{A} \cdot \mathcal{E}$ the tuple $\boldsymbol{a}^*$ is in $\mathsf{Tp}_{\mathsf{t}'}^{\mathfrak{A} \cdot \mathcal{E}}$. In conclusion, this argument shows that $\mathfrak{A} \cdot \mathcal{E}$ satisfies $\Psi_{\text{guard},\mathcal{E}(\mathfrak{A})}$.

Now let $\mathfrak{B}$ be a τ-structure with $h \colon \mathfrak{A} \to \mathfrak{B}$ homomorphism. We let $\mathfrak{B}^*$ denote the $\tau \uplus \sigma$-structure $\mathfrak{B} \cdot h(\mathfrak{A} \cdot \mathcal{E}|_\sigma)$. Note that by definition h is also a homomorphism $\mathfrak{A} \cdot \mathcal{E} \to \mathfrak{B} \cdot h(\mathfrak{A} \cdot \mathcal{E}|_\sigma)$. Let $\boldsymbol{b}$ be a tuple from $\mathfrak{B}^*$ satisfying Tp_{t} for some $\mathsf{t} \in \mathcal{E}_+$. By definition there is $\boldsymbol{a}$ in $\mathfrak{A}^*$ (and hence in $\mathfrak{A}$) such that $\boldsymbol{a} \in \mathsf{Tp}_{\mathsf{t}}^{\mathfrak{A}^*}$ and $h(\boldsymbol{a}) = \boldsymbol{b}$. As $\mathfrak{A}^*$ satisfies Ψ_{hom} by definition we obtain that $\boldsymbol{a}$ realizes t. Especially it satisfies the positive literals from t. We extend the componentwise variable assignment $\mu \colon \boldsymbol{v}_{\mathsf{t}} \to \boldsymbol{a}$ to a map $\mu' \colon \boldsymbol{v}_{\mathsf{t}} \uplus C \to A$ such that $\mu'|_{\boldsymbol{v}_{\mathsf{t}}} = \mu$ and for every $\mathsf{c} \in C$ we have $\mu'(\mathsf{c}) = \mathsf{c}^{\mathfrak{A}}$. Hence for every atom $\mathsf{R}(\boldsymbol{t}) \in \mathsf{t}$ with $\boldsymbol{t} \subseteq \boldsymbol{v}_{\mathsf{t}} \uplus C$ we have (applying μ' componentwise) $\mu'(\boldsymbol{t}) \in \mathsf{R}^{\mathfrak{A}}$. Consequently, the (componentwise) image $h(\mu'(\boldsymbol{t}))$ is a subtuple from $\boldsymbol{b}$ such that $h(\mu'(\boldsymbol{t})) \in \mathsf{R}^{\mathfrak{B}^*}$. We thus obtain that $\mathfrak{B}^* \vDash \Psi_{\text{hom}}$.

Next we will show that $\mathfrak{B}^*$ also satisfies $\Psi_{\text{guard},(\mathcal{E}_+,\mathcal{E}_!)}$. So let $\boldsymbol{b}$ be some tuple from $\mathfrak{B}^*$ and $(\mathsf{t},\nu) \in B_j^{(\mathcal{E}_+,\mathcal{E}_!)}$ for some j such that $\boldsymbol{b} \in \mathsf{Tp}_{\mathsf{t}}^{\mathfrak{B}^*}$. Additionally let $\mu_{\boldsymbol{b}} \colon \boldsymbol{v}_{\mathsf{t}} \to \boldsymbol{b}$ be the componentwise variable assignment. By definition of $\mathfrak{B}^*$ there is $\boldsymbol{a}$ in $\mathfrak{A}^*$ such that $\boldsymbol{a} \in \mathsf{Tp}_{\mathsf{t}}^{\mathfrak{A}^*}$, $\mu_{\boldsymbol{a}} \colon \boldsymbol{v}_{\mathsf{t}} \to \boldsymbol{a}$ componentwise variable assignment and $h(\boldsymbol{a}) = \boldsymbol{b}$ (again, componentwise). Since $\mathfrak{A}^*$ satisfies $\Psi_{\text{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ we find $(\mathsf{t}',\nu') \in P_j^{(\mathcal{E}_+,\mathcal{E}_!)}$ with $(\mathsf{t},\nu) \Subset_j (\mathsf{t}',\nu')$ and a tuple $\boldsymbol{a}^*$ such that the componentwise variable assignment $\mu_{\boldsymbol{a}^*} \colon \boldsymbol{v}_{\mathsf{t}'} \to \boldsymbol{a}^*$ extends $\mu_{\boldsymbol{a}}$, i.e. $\mu_{\boldsymbol{a}^*}|_{\boldsymbol{v}_{\mathsf{t}}} = \mu_{\boldsymbol{a}}$. Applying h to $\boldsymbol{a}^*$ we obtain $\boldsymbol{b}^* = h(\boldsymbol{a}^*)$ a supertuple of $\boldsymbol{b}$. The assignment $\mu_{\boldsymbol{b}^*} = h \circ \mu_{\boldsymbol{a}^*}$ extends $\mu_{\boldsymbol{b}}$:

$$\mu_{\boldsymbol{b}^*}|_{\boldsymbol{v}_{\mathsf{t}}} = h \circ (\mu_{\boldsymbol{a}^*}|_{\boldsymbol{v}_{\mathsf{t}}}) = h \circ \mu_{\boldsymbol{a}} = \mu_{\boldsymbol{b}}. \tag{122}$$

As h is a homomorphism we obtain $b^* \in \mathrm{Tp}_{t'}^{\mathfrak{B}^*}$. This argument shows that $\mathfrak{B}^* \vDash \Psi_{\mathrm{guard},\mathcal{E}(\mathfrak{A})}$.

Now suppose in addition, that Φ is from $\mathbb{TGF}$. We first show that then $\mathfrak{A} \cdot \mathcal{E}$ satisfies $\Psi_{\mathbb{TGF},\mathcal{E}(\mathfrak{A})}$. Take a_1, a_2 from $\mathfrak{A} \cdot \mathcal{E}$ (hence from $\mathfrak{A}$) satisfying for some types $t_1, t_2 \in \mathcal{E}_+$ the following: $a_1 \in \mathrm{Tp}_{t_1}^{\mathfrak{A} \cdot \mathcal{E}}$ and $a_2 \in \mathrm{Tp}_{t_2}^{\mathfrak{A} \cdot \mathcal{E}}$. Without loss of generality we may assume that $v_{t_1} = \{v_1\}$ and $v_{t_2} = \{v_2\}$ for $v_1 \neq v_2$. By definition a_1 realizes t_1 and a_2 realizes t_2 in $\mathfrak{A}$ (and $\mathfrak{A} \cdot \mathcal{E}$). For (a_1, a_2) there is a unique 2-type t' over τ with $v_{t'} = v_1 v_2$ (as a subsequence from v) such that (a_1, a_2) realizes t' in $\mathfrak{A}$ and $t'|_{v_1} = t_1$ and $t'|_{v_2} = t_2$. Because of $\mathfrak{A} \vDash \Phi_{\mathrm{Univ}}$ the type t' contains automatically the atoms $\mathrm{Univ}(v_1, v_2)$ and $\mathrm{Univ}(v_2, v_1)$. Hence t' is guarded and so we see that $t' \in \mathcal{E}$ (and of course also in $\mathcal{E}_+$). By definition of $\mathfrak{A} \cdot \mathcal{E}$ we obtain $(a_1, a_2) \in \mathrm{Tp}_{t'}^{\mathfrak{A} \cdot \mathcal{E}}$. Hence also $\mathfrak{A} \cdot \mathcal{E} \vDash \Psi_{\mathbb{TGF},\mathcal{E}(\mathfrak{A})}$.

Finally, also $\mathfrak{B}^*$ satisfies $\Psi_{\mathbb{TGF},\mathcal{E}(\mathfrak{A})}$. To see this, let b_1, b_2 be from $\mathfrak{B}^*$ such that there are $t_1, t_2 \in \mathcal{E}_+$ with $b_1 \in \mathrm{Tp}_{t_1}^{\mathfrak{B}^*}$ and $b_2 \in \mathrm{Tp}_{t_2}^{\mathfrak{B}^*}$. Without loss of generality we may assume that $v_{t_1} = \{v_1\}$ and $v_{t_2} = \{v_2\}$ for $v_1 \neq v_2$. By definition of $\mathfrak{B}^*$ there are a_1, a_2 in $\mathfrak{A} \cdot \mathcal{E}$ such that $a_1 \in \mathrm{Tp}_{t_1}^{\mathfrak{A} \cdot \mathcal{E}}$ and $a_2 \in \mathrm{Tp}_{t_2}^{\mathfrak{A} \cdot \mathcal{E}}$ with $h(a_1) = b_1$ and $h(a_2) = b_2$. Since we already established that $\mathfrak{A} \cdot \mathcal{E}$ satisfies $\Psi_{\mathbb{TGF},\mathcal{E}(\mathfrak{A})}$, there is a unique 2-type $t' \in \mathcal{E}_+$ with $v_{t'} = v_1 v_2$ (as a subsequence from v) with $t'|_{v_1} = t_1$ and $t'|_{v_2} = t_2$ such that $(a_1, a_2) \in \mathrm{Tp}_{t'}^{\mathfrak{A} \cdot \mathcal{E}}$. As previously mentioned, h as a map is also a homomorphism $\mathcal{A} \cdot \mathcal{E} \to \mathfrak{B}^*$. Hence, by $h((a_1, a_2)) = (b_1, b_2)$, we obtain $(b_1, b_2) \in \mathrm{Tp}_{t'}^{\mathfrak{B}'}$. This argument shows $\mathfrak{B}^* \vDash \Psi_{\mathbb{TGF},\mathcal{E}(\mathfrak{A})}$ and concludes our proof. $\qquad \square$

We will now prove a lemma that will allow us to extend a structure not satisfying $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ to one that does. Of course, this will not be so simple. Hence we first show that, if there is a "problem" for satisfaction of $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ then we are able to "repair" this problem without introducing additional ones with respect to the original structure. This will not necessarily preclude obtaining new "problems" in the extended structure, though those new ones will be located in the new additions. Dealing with this will be part of Lemma 81 and Lemma 82. The notion of "problem" in this context has been made formal by the definition of violation tuples in Definition 77. The main idea is the following: If we have a "problem" (formalized via a violation tuple), we can take some type "fitting" for the position where the "problem" occurs (consistent with $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$) and glue $\mathfrak{D}_t$ (or $(\mathfrak{D}_t)^{\tau}$ in the enriched setting) on accordingly. And as we are concerned with homclosures we will also discuss the cases when we have homomorphisms from a structure "having problems" to one that has none.

Lemma 79. *Let $\mathfrak{A}$ be a finite $\sigma \uplus C$-structure satisfying $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. Let $(t, \nu) \in B_j^{(\mathcal{E}_+,\mathcal{E}_!)}$ and $(t', \nu') \in P_j^{(\mathcal{E}_+,\mathcal{E}_!)}$ be fitting, i.e. $(t, \nu) \in_j (t', \nu')$. Furthermore let a be a tuple of length $\mathrm{order}(t)$ in $\mathfrak{A}$ and $\mu: v_t \to a$ be the componentwise variable assignment such that $(\mathfrak{A}, \mu) \vDash \mathrm{Tp}_t(v_t)$, i.e. $a \in \mathrm{Tp}_t^{\mathfrak{A}}$. Then we can extend $\mathfrak{A}$ to a $\sigma \uplus C$-structure $\mathfrak{A}^*$ and the compo-*

nentwise assignment μ to a componentwise one $\mu^*: v_{t'} \to \mathfrak{A}^*$ in a canonical way such that $(\mathfrak{A}^*, \mu^*) \vDash \mathrm{Tp}_{t'}(v_{t'})$, and $\mu^*(v_{t'} \setminus v_t)$ has no element in common with the domain A of $\mathfrak{A}$. Furthermore $\mathfrak{A}$ is an induced substructure of $\mathfrak{A}^*$ and $\mathfrak{A}^*$ satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$.

If there is a $\sigma \uplus C$-structure $\mathfrak{B}$ satisfying

$$\forall v_t. \mathrm{Tp}_t(v_t) \Rightarrow \exists v_{t'} \setminus \nu(z). \mathrm{Tp}_{t'}(v_{t'}) \tag{123}$$

and a homomorphism $h: \mathfrak{A} \to \mathfrak{B}$ we can extend it to a homomorphism $h^*: \mathfrak{A}^* \to \mathfrak{B}$.

Additionally, if there are $\tau \uplus \sigma$-structures $\mathfrak{A}'$ and $\mathfrak{B}'$ such that $\mathfrak{A} = \mathfrak{A}'|_{\sigma \uplus C}$ and $\mathfrak{B} = \mathfrak{B}'|_{\sigma \uplus C}$, $\mathfrak{B}' \vDash \Psi_{\mathrm{hom}}$, for all $t \in \mathcal{E}$ it holds that a tuple a without constants from $\mathfrak{A}'$ satsifies Tp_t in $\mathfrak{A}'$ if and only if a realizes t in $\mathfrak{A}'$, and there is a homomorphism $h': \mathfrak{A}' \to \mathfrak{B}'$, then we can extend $\mathfrak{A}'$ to a structure $(\mathfrak{A}')^*$ with $(\mathfrak{A}')^*|_{\sigma \uplus C} = \mathfrak{A}^*$ and h' to a homomorphism $(h')^*: (\mathfrak{A}')^* \to \mathfrak{B}'$. Furthermore, $\mathfrak{A}'$ is an induced substructure of $(\mathfrak{A}')^*$, and $(\mathfrak{A}')^*$ satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. Additionally, for all $t \in \mathcal{E}$ it holds that a tuple a without constants from $(\mathfrak{A}')^*$ satsifies Tp_t in $(\mathfrak{A}')^*$ if and only if a realizes t in $(\mathfrak{A}')^*$.

Let Q_a be the set of all violation tuples of $\mathfrak{A}$ (or $\mathfrak{A}'$) of the form $q_i = (c_i, \mu_i, t_i, \nu_i)$ where the c_i consist only of elements from a. Not only is Q_a finite but it contains already all violation tuples of the form (d, μ_*, t_*, ν_*) of $\mathfrak{A}^*$ (or $(\mathfrak{A}')^*$) where d consists of elements from a.

Proof. Let $\mathfrak{A}$ be a finite $\sigma \uplus C$ structure satisfying $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. Let $(t, \nu) \in B_j^{(\mathcal{E}_+,\mathcal{E}_!)}$ and $(t', \nu') \in P_j^{(\mathcal{E}_+,\mathcal{E}_!)}$ be fitting, i.e. $(t, \nu) \in_j (t', \nu')$. Furthermore let a be a tuple of length $\mathrm{order}(t)$ in $\mathfrak{A}$ and $\mu: v_t \to a$ be the componentwise variable assignment such that $(\mathfrak{A}, \mu) \vDash \mathrm{Tp}_t(v_t)$, i.e. $a \in \mathrm{Tp}_t^{\mathfrak{A}}$.

Now, consider the σ-structure $\mathfrak{D}_{t'}$ and be reminded that its domain are the elements from $v_{t'}$. We obtain $\mathfrak{D}_{t'}'$ from $\mathfrak{D}_{t'}$ by renaming all elements $\nu'(z) \in v_{t'} \cap \nu'(z)$ into $\mu(z)$ (i.e., into elements already present in $\mathfrak{A}$) and assigning a fresh name to any other element $v \in v_{t'} \setminus \nu'(z)$. Now we obtain $\mathfrak{A}^*$ by taking the (non-disjoint) union of $\mathfrak{A}$ with $\mathfrak{D}_{t'}'$.

Note that by (t, ν) and (t', ν') being fitting we obtain that $\mathfrak{A}$ is an induced substructure of $\mathfrak{A}^*$ by construction. As $(t', \nu') \in P_j^{(\mathcal{E}_+,\mathcal{E}_!)}$, clearly $t' \in \mathcal{E}_+$. By definition and applying Lemma 67 this means that $\mathfrak{D}_{t'}$ and hence $\mathfrak{D}_{t'}'$ satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. More specifically, the validity of Formulae 66 and 67 is clear from the closure under those formulae in the definition, whereas 64 and 65 follow from $t \in \mathcal{E}_+$ and Lemma 67 and the definition. Note that the saturation with respect to Formulae 66 and 67 is closely tied to the relation of the types to which the type predicates are associated. And, by again applying that (t, ν) and (t', ν') are fitting, we conclude that the union of $\mathfrak{A}$ and $\mathfrak{D}_{t'}'$, which is $\mathfrak{A}^*$ inherits the satisfaction of $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. This is clear while observing that the substructure induced by a in $\mathfrak{A}$ and the one in $\mathfrak{D}_{t'}'$ are the same. Hence we do not create any new relations with respect to relation symbols from σ. By this token, $\mathfrak{A}$ is an

induced substructure of $\mathfrak{A}^*$, as no new relations are introduced on tuples purely from $\mathfrak{A}$.

If there is a $\sigma \uplus C$-structure $\mathfrak{B}$ satisfying

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\mathrm{Tp}_{\mathfrak{t}} \Rightarrow \exists \boldsymbol{v}_{\mathfrak{t}'} \setminus \nu(\boldsymbol{z}).\mathrm{Tp}_{\mathfrak{t}'}(\boldsymbol{v}_{\mathfrak{t}'}) \qquad (124)$$

and a homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$ we adjust the construction in the following way: Since $h \colon \mathfrak{A} \to \mathfrak{B}$ is a homomorphism the tuple $\boldsymbol{b} = h(\boldsymbol{a})$ satisfies $\boldsymbol{b} \in \mathrm{Tp}_{\mathfrak{t}}^{\mathfrak{B}}$. This is witnessed by the componentwise assignment $\mu' \colon \boldsymbol{v}_{\mathfrak{t}} \to \boldsymbol{b}$ defined as $h \circ \mu$. As $\mathfrak{B}$ satisfies Formula 124 we obtain a tuple $\boldsymbol{b}^*$ containing $\boldsymbol{b}$ as a subtuple and can extend μ' to the componentwise assignment $\mu^* \colon \boldsymbol{v}_{\mathfrak{t}'} \to \boldsymbol{b}^*$ (i.e. $\mu^*|_{\boldsymbol{v}_{\mathfrak{t}}} = \mu'$). Then, by what we have seen earlier, the componentwise map $g \colon \boldsymbol{v}_{\mathfrak{t}'} \to \boldsymbol{b}^*, v \mapsto \mu'(v)$ is a homomorphism from $\mathfrak{D}_{\mathfrak{t}'}$ to $\mathfrak{B}$. By moving from $\mathfrak{D}_{\mathfrak{t}'}$ to $\mathfrak{D}_{\mathfrak{t}'}'$ via renaming (which establishes an isomorphism $f \colon \mathfrak{D}_{\mathfrak{t}'} \to \mathfrak{D}_{\mathfrak{t}'}'$) we obtain via $g \circ f^{-1}$ a homomorphism $\dot{g} \colon \mathfrak{D}_{\mathfrak{t}'}' \to \mathfrak{B}$

After obtaining $\mathfrak{A}^*$ via the union of $\mathfrak{A}$ with $\mathfrak{D}_{\mathfrak{t}'}'$, we let h^* be the union of the maps h and $\dot{g}$, i.e. $h \cup \dot{g}$. Note that this is well defined, since $h(a) = \dot{g}(a)$ for every $a \in \boldsymbol{a}$. To be more precise, by the variable assignments μ, μ' being componentwise and $(\mathfrak{t}, \nu)$ and $(\mathfrak{t}', \nu')$ being fitting we obtain for $v \in \boldsymbol{v}_{\mathfrak{t}}$ with $\mu(v) = a$:

$$h(a) = h(\mu(v)) = \mu'(v) = g(v) = g \circ f^{-1}(a) = \dot{g}(a). \quad (125)$$

Since the substructures induced by $\boldsymbol{a}$ in $\mathfrak{D}_{\mathfrak{t}'}'$ and in $\mathfrak{A}$ are equal, this implies together with h and $\dot{g}$ being local homomorphisms that $h^* \colon \mathfrak{A}^* \to \mathfrak{B}$ is a homomorphism.

We turn to the part about $\tau \uplus \sigma$-structures $\mathfrak{A}'$ and $\mathfrak{B}'$. As $\mathfrak{A}'|_{\sigma \uplus C} = \mathfrak{A}$ and $\mathfrak{B}'|_{\sigma \uplus C} = \mathfrak{B}$, $\mathfrak{A}'$ satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ whereas $\mathfrak{B}'$ satisfies 124. Additionally, $h' \colon \mathfrak{A}' \to \mathfrak{B}'$ is a homomorphism $\mathfrak{A} \to \mathfrak{B}$. Hence we apply the previous results from this lemma/proof and extend $\mathfrak{A}'$ to a structure $\mathfrak{A}_{\mathrm{temp}}'$ by adding $\mathfrak{D}_{\mathfrak{t}'}$ as previously described. Additionally we extend h' to $(h')^*$ again as previously described. We hence obtain $\mathfrak{A}_{\mathrm{temp}}'|_{\sigma \uplus C} = \mathfrak{A}^*$ and $(h')^* \colon \mathfrak{A}_{\mathrm{temp}}'|_{\sigma \uplus C} \to \mathfrak{B}'|_{\sigma \uplus C}$ being a homomorphism. Let $f' \colon \mathfrak{C} \uplus \mathfrak{D}_{\mathfrak{t}'} \to \mathfrak{A}_{\mathrm{temp}}'$ be the map that bijectively maps $\mathfrak{D}_{\mathfrak{t}'}$ to $\mathfrak{D}_{\mathfrak{t}'}'$ in $\mathfrak{A}_{\mathrm{temp}}'$ and $f'(\mathtt{c}^{\mathfrak{C}}) = \mathtt{c}^{(\mathfrak{A}')^*}$. For $\mathtt{R} \in \tau \setminus C$ we define

$$\mathtt{R}^{(\mathfrak{A}')^*} = \mathtt{R}^{\mathfrak{A}_{\mathrm{temp}}'} \cup \{ f'(\boldsymbol{d}) \mid \boldsymbol{d} \in \mathtt{R}^{(\mathfrak{D}_{\mathfrak{t}'})^\tau} \}. \qquad (126)$$

By definition, f' gives rise to an embedding $f'_{\mathfrak{t}'} \colon (\mathfrak{D}_{\mathfrak{t}'})^\tau \to (\mathfrak{A}')^*$: That $f'_{\mathfrak{t}'}$ is an injective homomorphism is plain. So assume we have $\boldsymbol{d}$ such that $f'_{\mathfrak{t}'}(\boldsymbol{d}) \in \mathtt{R}^{(\mathfrak{A}')^*}$ for $\mathtt{R} \in \tau \setminus C$. By definition $\boldsymbol{d} \in \mathtt{R}^{(\mathfrak{D}_{\mathfrak{t}'})^\tau}$. Hence we have to discuss $f'_{\mathfrak{t}'}(\boldsymbol{d}) \in \mathtt{T}^{(\mathfrak{A}')^*}$ for $\mathtt{T} \in \sigma$. As the $\mathtt{T} \in \sigma$ relations holding for elements from $f'_{\mathfrak{t}'}((\mathfrak{D}_{\mathfrak{t}'})^\tau)$ have been taken over exactly from $\mathfrak{D}_{\mathfrak{t}'}$ we easily obtain $\boldsymbol{d} \in \mathtt{T}^{(\mathfrak{D}_{\mathfrak{t}'})^\tau}$ showing that $f'_{\mathfrak{t}'}$ is an embedding. Additionally, every tuple $\boldsymbol{a}$ from $(\mathfrak{A}')^*$ such that $\boldsymbol{a} \in \mathrm{Tp}_{\mathfrak{t}''}^{(\mathfrak{A}')^*}$ realizes $\mathfrak{t}''$ in $(\mathfrak{A}')^*$ for arbitrary $\mathfrak{t}'' \in \mathcal{E}$. If $\boldsymbol{a}$ is purely from $\mathfrak{A}'$ this holds by definition. Hence we only need to discuss the case where $\boldsymbol{a}$ is from $f'_{\mathfrak{t}'}((\mathfrak{D}_{\mathfrak{t}'})^\tau)$. Since $(\mathfrak{t}, \nu)$ and $(\mathfrak{t}', \nu')$ where fitting, we immediately obtain that as soon as $\boldsymbol{a}$ contains at least one element from $f'_{\mathfrak{t}'}((\mathfrak{D}_{\mathfrak{t}'})^\tau)$ all of its elements are from this set. But by $f'_{\mathfrak{t}'}$ being an embedding there is a unique

$\boldsymbol{d}$ in $(\mathfrak{D}_{\mathfrak{t}'})^\tau$ with $\boldsymbol{d} \in \mathrm{Tp}_{\mathfrak{t}''}^{(\mathfrak{D}_{\mathfrak{t}'})^\tau}$ such that $f'_{\mathfrak{t}'}(\boldsymbol{d}) = \boldsymbol{a}$. By Lemma 74 $\boldsymbol{d}$ realizes $\mathfrak{t}''$ in $(\mathfrak{D}_{\mathfrak{t}'})^\tau$ and hence by $f'_{\mathfrak{t}'}$ being an embedding, $\boldsymbol{a} = f'_{\mathfrak{t}'}(\boldsymbol{d})$ realizes $\mathfrak{t}''$ in $(\mathfrak{A}')^*$.

For the other direction suppose $\boldsymbol{a}$ realizes some type $\mathfrak{t}''$ in $(\mathfrak{A}')^*$. Then, by construction, it is a tuple either from just $\mathfrak{A}'$, where by assumption $\boldsymbol{a} \in \mathrm{Tp}_{\mathfrak{t}''}^{(\mathfrak{A}')^*}$ holds, or it is from $f'_{\mathfrak{t}'}((\mathfrak{D}_{\mathfrak{t}'})^\tau)$. As $f'_{\mathfrak{t}'}$ is an embedding there exists a unique $\boldsymbol{d}$ from $(\mathfrak{D}_{\mathfrak{t}'})^\tau$ such that $f'_{\mathfrak{t}'}(\boldsymbol{d}) = \boldsymbol{a}$ and $\boldsymbol{d}$ realizes $\mathfrak{t}''$ in $(\mathfrak{D}_{\mathfrak{t}'})^\tau$. By Lemma 74 $\boldsymbol{d} \in \mathrm{Tp}_{\mathfrak{t}''}^{(\mathfrak{D}_{\mathfrak{t}'})^\tau}$ and hence via $f'_{\mathfrak{t}'}$ being an embedding we obtain $\boldsymbol{a} \in \mathrm{Tp}_{\mathfrak{t}''}^{(\mathfrak{A}')^*}$.

Lastly, $(h')^*$ is a homomorphism $(\mathfrak{A}')^* \to \mathfrak{B}'$. For this let $\mathtt{R} \in \tau \setminus C$ and $\boldsymbol{a}$ be a tuple from $(\mathfrak{A}')^*$ with $\boldsymbol{a} \in \mathtt{R}^{(\mathfrak{A}')^*}$. We continue the discussion by using the atom $\mathtt{R}(\boldsymbol{x})$ containing only the (pairwise distinct) variables from $\boldsymbol{x}$. Set $\mu \colon \boldsymbol{x} \to \boldsymbol{a}$ as the componentwise valuation of $\boldsymbol{x}$ by $\boldsymbol{a}$. Let $\boldsymbol{y} = (y_1, \ldots, y_m)$ be a tuple of (fresh) variables, where m is the number of all components (allowing repetitions of elements to be counted) of $\boldsymbol{a}$ that are not the interpretation of constant symbols. We define $\nu_\mu \colon \boldsymbol{x} \to \boldsymbol{y} \uplus C$ as follows: $\nu_\mu(x) = \mathtt{c} \in C$ if and only if $\mu(x) = \mathtt{c}^{(\mathfrak{A}')^*}$, $\nu_\mu(x) \in \boldsymbol{y}$ otherwise. Additionally, ν_μ shall map the subtuple $\boldsymbol{x}^*$ of $\boldsymbol{x}$ induced by $\{x \in \boldsymbol{x} \mid \nu_\mu(x) \in \boldsymbol{y}\}$ componentwise to the tuple $\boldsymbol{y}$. By definition μ_ν is a bijection between $\boldsymbol{x}^*$ and $\boldsymbol{y}$. Finally note that $\mathtt{c}$ in the definition of ν_μ might not be unique. In this case we choose one of the possible options. Still this map will be well defined for our purposes. We now define the atom $\delta(\boldsymbol{y})$ as $\nu_\mu(\mathtt{R}(\boldsymbol{x}))$. Additionally, we acknowledge that $\boldsymbol{a}^*$, as a tuple being the componentwise image of $\boldsymbol{x}^*$ of μ, satisfies $\delta(\boldsymbol{y})$ in $(\mathfrak{A}')^*$. Let $\mathfrak{t}^*$ be the τ-type realized by $\boldsymbol{a}^*$ in $(\mathfrak{A}')^*$. As $\boldsymbol{x}^*$ satisfies $\delta(\boldsymbol{y})$, $\mathfrak{t}^*$ is guarded by δ (witnessed by the componentwise map $\nu \colon \boldsymbol{y} \to \boldsymbol{v}_{\mathfrak{t}^*}$). Hence $\mathfrak{t}^* \in \mathcal{E}$ and $\boldsymbol{a}^*$ satisfies $\mathrm{Tp}_{\mathfrak{t}^*}$ in $(\mathfrak{A}')^*$ by construction. As $\mathfrak{A}'$ satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ we obtain that $\mathfrak{t}^* \in \mathcal{E}_+$ (by applying Formula 65). As $(h')^*$ has also been shown to be a homomorphism $(\mathfrak{A}')^*|_{\sigma \uplus C} \to \mathfrak{B}'|_{\sigma \uplus C}$, we obtain $(h')^*(\boldsymbol{a}^*) \in \mathrm{Tp}_{\mathfrak{t}^*}^{\mathfrak{B}'}$. Since $\mathfrak{B}' \models \Psi_{\mathrm{hom}}$, $(h')^*(\boldsymbol{a}^*)$ satisfies $\delta(\boldsymbol{y})$ in $\mathfrak{B}'$. Again, as $(h')^*$ has also been shown to be a homomorphism $(\mathfrak{A}')^*|_{\sigma \uplus C} \to \mathfrak{B}'|_{\sigma \uplus C}$, we conclude further that since $(h')^*(\mathtt{c}^{(\mathfrak{A}')^*}) = \mathtt{c}^{\mathfrak{B}'}$, application of $(h')^*$ to $\boldsymbol{a}$ yields $(h')^*(\boldsymbol{a}) \in \mathtt{R}^{\mathfrak{B}'}$. Consequently, $(h')^* \colon (\mathfrak{A}')^* \to \mathfrak{B}'$ is a homomorphism.

We show that $\mathfrak{A}'$ is an induced substructure of $(\mathfrak{A}')^*$. It is clear by construction, that the injective map $\iota \colon \mathfrak{A}' \to (\mathfrak{A}')^*, a \mapsto a$ is a homomorphism. As by construction no new type predicate has been introduced on tuples from purely $\mathfrak{A}'$, there also did not arise the need to introduce new τ-relations on tuples from purely $\mathfrak{A}'$ in order to saturate for a correspondence between type predicates and τ-types (as this was already given by assumption). Hence ι is already an embedding, as every τ-relation $\mathtt{R}$ holding for some $\iota(\boldsymbol{a})$ in $(\mathfrak{A}')^*$ (where $\boldsymbol{a}$ is a tuple from $\mathfrak{A}'$) already held in $\mathfrak{A}'$. Hence, $\mathfrak{A}'$ is an induced substructure of $(\mathfrak{A}')^*$, and $(\mathfrak{A}')^*$ satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$.

For the remainder it is enough to restrict our argument to the $\sigma \uplus C$-structures. Let $Q_{\boldsymbol{a}}$ be the set of all violation tuples of $\mathfrak{A}$ of the form $q_i = (\boldsymbol{c}_i, \mu_i, \mathfrak{t}_i, \nu_i)$ where the $\boldsymbol{c}_i$ consist only

of elements from $\boldsymbol{a}$. As $\boldsymbol{a}$ is finite and so are the number if types and type variables, $Q_{\boldsymbol{a}}$ is finite. Hence we can assume $Q_{\boldsymbol{a}} = \{q_1, \ldots, q_m\}$ with the q_i as mentioned.

It remains to show, that all violation tuples of the form $(\boldsymbol{d}, \mu_*, \mathsf{t}_*, \nu_*)$ of $\mathfrak{A}^*$ where $\boldsymbol{d}$ consists of elements from $\boldsymbol{a}$ are already in $Q_{\boldsymbol{a}}$. Suppose we have such a violation tuple $(\boldsymbol{d}, \mu_*, \mathsf{t}_*, \nu_*)$ of $\mathfrak{A}^*$. Consequently $(\mathfrak{A}^*, \mu_*) \vDash \mathsf{Tp}_{\mathsf{t}_*}(\boldsymbol{v}_{\mathsf{t}_*})$, i.e. $(\boldsymbol{d}) \in \mathsf{Tp}_{\mathsf{t}_*}^{\mathfrak{A}^*}$. But since in the construction of $\mathfrak{A}^*$ we have that $(\mathsf{t}, \nu) \in_j (\mathsf{t}', \nu')$, i.e. they are fitting, the substructures of $\mathfrak{A}$ and $\mathfrak{D}'_{\mathsf{t}'}$ are the same (hence isomorphic), every predicate $\mathsf{Tp}_{\mathsf{t}''} \in \sigma$ that holds for tuples built from elements of $\boldsymbol{a}$ in $\mathfrak{A}^*$ already held in $\mathfrak{A}$ and vice versa. Consequently $\boldsymbol{d} \in \mathsf{Tp}_{\mathsf{t}_*}^{\mathfrak{A}}$. But then we conclude by construction that $(\boldsymbol{d}, \mu_*, \mathsf{t}_*, \nu_*)$ was already a violation tuple of $\mathfrak{A}$. By definition this means it is in $Q_{\boldsymbol{a}}$. $\qquad \square$

As we (essentially) just did for $\Psi_{\mathrm{guard}, (\mathcal{E}_+, \mathcal{E}_!)}$ in Lemma 79 we will show how we can enrich structures in order to make $\Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)}$ true.

Lemma 80. *We restrict Φ to be a $\mathbb{TGF}$ sentence in normal form. Let $\mathfrak{A}, \mathfrak{B}$ be $\tau \uplus \sigma$-structures with $\mathfrak{A}$ being finite and a homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$ such that both $\mathfrak{A}$ and $\mathfrak{B}$ satisfy $\Psi_{\mathrm{gen}, (\mathcal{E}_+, \mathcal{E}_!)}$. Assume that $\mathfrak{A}$ satisfies the following: For all $\mathsf{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ from $\mathfrak{A}$ without constants satisfies $\mathsf{Tp}_\mathsf{t} \in \sigma$ in $\mathfrak{A}$ if and only if $\boldsymbol{a}$ realizes t in $\mathfrak{A}$. Further let $\mathfrak{B}$ additionally satisfy $\Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)}$ and Ψ_{hom}. Then there is a $\tau \uplus \sigma$-structure $\mathfrak{A}'$ with domain A satisfying $\Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)}$ in addition to $\Psi_{\mathrm{gen}, (\mathcal{E}_+, \mathcal{E}_!)}$, and for every type $\mathsf{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ without constants from $\mathfrak{A}'$ satisfies $\mathsf{Tp}_\mathsf{t} \in \sigma$ in $\mathfrak{A}'$ if and only if $\boldsymbol{a}$ realizes t in $\mathfrak{A}'$. Lastly the identity map $\iota \colon \mathfrak{A} \to \mathfrak{A}', a \mapsto a$ is a homomorphism.*

Furthermore, every induced substructure of $\mathfrak{A}$ satisfying $\Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)}$ also is an induced substructure of $\mathfrak{A}'$.

Proof. Let $\mathfrak{A}, \mathfrak{B}$ be $\tau \uplus \sigma$-structures with $\mathfrak{A}$ being finite, $h \colon \mathfrak{A} \to \mathfrak{B}$ a homomorphism, $\mathfrak{A}, \mathfrak{B} \vDash \Psi_{\mathrm{gen}, (\mathcal{E}_+, \mathcal{E}_!)}$ and $\mathfrak{B} \vDash \Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)} \wedge \Psi_{\mathrm{hom}}$. Furthermore, for every tuple $\boldsymbol{a}$ from $\mathfrak{A}$, containing no $\mathsf{c}^{\mathfrak{A}}$ for $\mathsf{c} \in C$ and every $\mathsf{t} \in \mathcal{E}$ the following equivalence holds: $\boldsymbol{a}$ satisfies $\mathsf{Tp}_\mathsf{t} \in \sigma$ in $\mathfrak{A}$ if and only if $\boldsymbol{a}$ realizes t in $\mathfrak{A}$. If $\mathfrak{A}$ already satisfies $\Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)}$ we are already done, so suppose this were not the case. We will start by successively saturating the σ-relations of $\mathfrak{A}$ in a minimal manner such that the so obtained structure (denoted by $\mathfrak{A}^*$) satisfies $\Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)}$. Beside that, nothing more will be done in this first step. This also means, that the interpretations of $\mathsf{R} \in \tau$ relations will not change, equally the domain remains to be A. Hence we do the following for every two elements a_1, a_2 from $\mathfrak{A}$ such that there are 1-types $\mathsf{t}_1(v_1), \mathsf{t}_2(v_2) \in \mathcal{E}_+$ with $a_1 \in \mathsf{Tp}_{\mathsf{t}_1}^{\mathfrak{A}}, a_2 \in \mathsf{Tp}_{\mathsf{t}_2}^{\mathfrak{A}}$: If there is already a 2-type $\mathsf{t}^* \in \mathcal{E}_+$ with $\mathsf{Tp}_{\mathsf{t}^*}^{\mathfrak{A}^*}$ containing (a_1, a_2) then, by Formula 66, $\Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)}$ cannot be violated by this pair. This is also the case, when $a_1 = a_2$. We see this by the following argument: Since Φ is a normal form sentence from $\mathbb{TGF}$ every type $\mathsf{t} \in \mathcal{E}$ is for each $v \in \boldsymbol{v}_\mathsf{t}$ the atom

$\mathsf{Univ}(v, v)$. Hence, by $a_1 \in \mathsf{Tp}_{\mathsf{t}_1}^{\mathfrak{A}}$ and the assumption that type predicate satisfaction and type realization of the associated types coincides we obtain $(a_1, a_1) \in \mathsf{Univ}^{\mathfrak{A}}$. By $a_1 = a_2$ this is the same as $(a_1, a_2) \in \mathsf{Univ}^{\mathfrak{A}}$. Hence (a_1, a_2) satisfies $\mathsf{Univ}(v_1, v_2)$. So the type t' of (a_1, a_2) with variables (v_1, v_2) is guarded (by $\mathsf{Univ}(v_1, v_2)$) and consequently $\mathsf{t}' \in \mathcal{E}$. By going back to type predicates, applying that $\mathfrak{A} \vDash \Psi_{\mathrm{gen}, (\mathcal{E}_+, \mathcal{E}_!)}$ we obtain that $\mathsf{t}' \in \mathcal{E}_+$. Hence $(a_1, a_2) \in \mathsf{Tp}_{\mathsf{t}'}^{\mathfrak{A}}$.

So suppose the pair (a_1, a_2) is not yet in $\mathsf{Tp}_{\mathsf{t}^*}^{\mathfrak{A}^*}$ for any 2-type t^*. We know by h being a homomorphism that $h(a_1) \in \mathsf{Tp}_{\mathsf{t}_1}^{\mathfrak{B}}, h(a_2) \in \mathsf{Tp}_{\mathsf{t}_2}^{\mathfrak{B}}$. Since $\mathfrak{B}$ satisfies $\Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)}$ we find some 2-type $\mathsf{t}^* \in \mathcal{E}_+$ satisfying $\boldsymbol{v}_{\mathsf{t}^*} = v_1 v_2$, $\mathsf{t}^*|_{v_1} = \mathsf{t}_1$, $\mathsf{t}^*|_{v_2} = \mathsf{t}_2$, and $(h(a_1), h(a_2)) \in \mathsf{Tp}_{\mathsf{t}^*}^{\mathfrak{B}}$. We now have to pick exactly one such t^* and then add (a_1, a_2) to the extension of $\mathsf{Tp}_{\mathsf{t}^*}^{\mathfrak{A}^*}$. Moreover, for every permutation $\kappa \colon \boldsymbol{v} \hookrightarrow \boldsymbol{v}$ we also add $\eta_{\boldsymbol{v}}^{\kappa \boldsymbol{v}_{\mathsf{t}^*}}((a_1, a_2))$ to $\mathsf{Tp}_{\kappa \mathsf{t}^*}^{\mathfrak{A}^*}$. This ensures that even after augmenting, $\Psi_{\mathrm{gen}, (\mathcal{E}_+, \mathcal{E}_!)}$ remains satisfied. Furthermore the construction enforces that now $\Psi_{\mathbb{TGF}, (\mathcal{E}_+, \mathcal{E}_!)}$ also holds for $\mathfrak{A}^*$. Lastly, as we only ever added tuples to the relations in order to obtain $\mathfrak{A}^*$ the identity map is clearly a homomorphism $\iota \colon \mathfrak{A} \to \mathfrak{A}^*$.

We will collect the tuples $((a_1, a_2), \mathsf{t}^*)$ consisting of pairs (a_1, a_2) not contained in any type predicates interpretation of $\mathfrak{A}$ and the type t^* chosen in the construction, in the set $\mathcal{N}$. Note that this means the following: For all $((a_1, a_2), \mathsf{t}^*)$ we have $(a_1, a_2) \in \mathsf{Tp}_{\mathsf{t}^*}^{\mathfrak{A}^*}$ but $(a_1, a_2) \notin \mathsf{Tp}_{\mathsf{t}^*}^{\mathfrak{A}}$.

Now let $((a_1, a_2), \mathsf{t}^*) \in \mathcal{N}$. To prepare for the next step, note the following: We ensured, that there is a homomorphism $f_{\mathsf{t}^*}^{(a_1, a_2)} \colon \mathfrak{D}_{\mathsf{t}^*} \to \mathfrak{A}^*|_\sigma$ via $f_{\mathsf{t}^*}^{(a_1, a_2)}(v_1) = a_1$ and $f_{\mathsf{t}^*}^{(a_1, a_2)}(v_2) = a_2$. Furthermore, $f_{\mathsf{t}^*}^{(a_1, a_2)}$ is even injective. We will extend $f_{\mathsf{t}^*}^{(a_1, a_2)}$ to the map $f_{\mathsf{t}^*, \mathfrak{C}}^{(a_1, a_2)} \colon (\mathfrak{D}_{\mathsf{t}^*} \uplus (\mathfrak{C} \cdot \breve{\sigma})) \to \mathfrak{A}^*|_{\sigma \uplus C}$, where $f_{\mathsf{t}^*, \mathfrak{C}}^{(a_1, a_2)}(\mathsf{c}^{\mathfrak{C}}) = \mathsf{c}^{\mathfrak{A}^*}$ which is a homomorphism, too.

In the next step, we will make use of $f_{\mathsf{t}^*, \mathfrak{C}}^{(a_1, a_2)}$ by applying it to $(\mathfrak{D}_{\mathsf{t}^*})^\tau$ and homomorphically transferring the τ relations. The so obtained structure will be denoted by $\mathfrak{A}'$. Hence for $\mathsf{R} \in \tau \setminus C$ we obtain

$$\mathsf{R}^{\mathfrak{A}'} = \mathsf{R}^{\mathfrak{A}^*} \cup \bigcup_{((a_1, a_2), \mathsf{t}^*)} \{f_{\mathsf{t}^*, \mathfrak{C}}^{(a_1, a_2)}(\boldsymbol{d}) \mid \boldsymbol{d} \in \mathsf{R}^{(\mathfrak{D}_{\mathsf{t}^*})^\tau}\}, \quad (127)$$

where the union extends over all $((a_1, a_2), \mathsf{t}^*)$ such that a $f_{\mathsf{t}^*, \mathfrak{C}}^{(a_1, a_2)}$ exists, i.e. we introduced in the previous step a binary type predicate for (a_1, a_2). We thus lift $f_{\mathsf{t}^*, \mathfrak{C}}^{(a_1, a_2)}$ to a homomorphism $(f')_{\mathsf{t}^*, \mathfrak{C}}^{(a_1, a_2)} \colon (\mathfrak{D}_{\mathsf{t}^*})^\tau \to \mathfrak{A}'$. This is, additionally, an embedding. By assumption, (a_1, a_2) did not realize any 2-type $\mathsf{t}' \in \mathcal{E}_+$. If that were the case, by assumption on $\mathfrak{A}$, (a_1, a_2) would have satisfied $\mathsf{Tp}_{\mathsf{t}'}(v_1, v_2)$ in $\mathfrak{A}$ and as $\mathfrak{A}$ satisfies $\Psi_{\mathrm{gen}, (\mathcal{E}_+, \mathcal{E}_!)}$ it holds that $a_1 \in \mathsf{Tp}_{\mathsf{t}'|_{v_1}}^{\mathfrak{A}}$ and $a_2 \in \mathsf{Tp}_{\mathsf{t}'|_{v_2}}^{\mathfrak{A}}$. Hence there would have been no addition of (a_1, a_2) to some $\mathsf{Tp}_{\mathsf{t}'}^{\mathfrak{A}^*}$. And again by assumption on $\mathfrak{A}$ there cannot be any binary τ relation holding for (a_1, a_2) as this would give rise to a guarded 2-type realized by (a_1, a_2), leading to the same argument. By assumption we exclude the case where one of

a_1 or a_2 is an interpretation of a constant symbol.

We now show that realization of types and satisfaction of the associated type predicate coincides for tuples without constants. Let $\boldsymbol{a}$ be such a tuple from $\mathfrak{A}'$ satisfying $\mathrm{Tp}_{\mathfrak{t}'}$ for some $\mathfrak{t}' \in \mathcal{E}$, i.e. $\boldsymbol{a} \in \mathrm{Tp}_{\mathfrak{t}'}^{\mathfrak{A}'}$. By construction, $\boldsymbol{a} \in \mathrm{Tp}_{\mathfrak{t}'}^{\mathfrak{A}}$ holds for all $\mathfrak{t}'$ being not of order 2. Then by assumption, $\boldsymbol{a}$ realizes $\mathfrak{t}'$ in $\mathfrak{A}$ and hence in $\mathfrak{A}'$. So suppose $\mathfrak{t}'$ being of order 2 and such that $\boldsymbol{a} \notin \mathrm{Tp}_{\mathfrak{t}'}^{\mathfrak{A}}$ (for order 2 and $\boldsymbol{a} \in \mathrm{Tp}_{\mathfrak{t}'}^{\mathfrak{A}}$ a repeat of the previous argument yields the realization of the type). Hence $\boldsymbol{a} = (a_1, a_2)$ and we have the embedding $(f')_{\mathfrak{t}',\mathfrak{C}}^{(a_1,a_2)} \colon (\mathfrak{D}_{\mathfrak{t}'})^\tau \to \mathfrak{A}'$ by construction. More specifically, we know that $(f')_{\mathfrak{t}',\mathfrak{C}}^{(a_1,a_2)}(\boldsymbol{v}_{\mathfrak{t}'}) = (a_1, a_2)$ and $\boldsymbol{v}_{\mathfrak{t}'}$ satisfies $\mathrm{Tp}_{\mathfrak{t}'}$ in $(\mathfrak{D}_{\mathfrak{t}'})^\tau$. By Lemma 74 $\boldsymbol{v}_{\mathfrak{t}'}$ realizes $\mathfrak{t}'$ in $(\mathfrak{D}_{\mathfrak{t}'})^\tau$ and by $(f')_{\mathfrak{t}',\mathfrak{C}}^{(a_1,a_2)}$ being an embedding we obtain that (a_1, a_2) realizes $\mathfrak{t}'$ in $\mathfrak{A}'$. On the other hand, let $\boldsymbol{a}$ realize $\mathfrak{t}' \in \mathcal{E}$ in $\mathfrak{A}'$. If $\boldsymbol{a}$ already realizes $\mathfrak{t}'$ in $\mathfrak{A}$ we immediately obtain $\boldsymbol{a} \in \mathrm{Tp}_{\mathfrak{t}'}^{\mathfrak{A}}$ and hence $\boldsymbol{a} \in \mathrm{Tp}_{\mathfrak{t}'}^{\mathfrak{A}'}$. By construction, the only non trivial cases are those for $\mathrm{order}(\mathfrak{t}') = 2$ and $\boldsymbol{a}$ not already realizing $\mathfrak{t}'$ in $\mathfrak{A}$. The latter is equivalent to saying, that $\boldsymbol{a}$ does not satisfy a binary τ relation in $\mathfrak{A}$. Hence there is the embedding $(f')_{\mathfrak{t}',\mathfrak{C}}^{(a_1,a_2)} \colon (\mathfrak{D}_{\mathfrak{t}'})^\tau \to \mathfrak{A}'$ by construction. Using this embedding, we obtain $(f')_{\mathfrak{t}',\mathfrak{C}}^{(a_1,a_2)}(\boldsymbol{v}_{\mathfrak{t}}) = (a_1, a_2)$ and $\boldsymbol{v}_{\mathfrak{t}'}$ realizing $\mathfrak{t}'$ in $(\mathfrak{D}_{\mathfrak{t}'})^\tau$. Hence by Lemma 74 $\boldsymbol{v}_{\mathfrak{t}'} \in \mathrm{Tp}_{\mathfrak{t}'}^{(\mathfrak{D}_{\mathfrak{t}'})^\tau}$. Applying $(f')_{\mathfrak{t}',\mathfrak{C}}^{(a_1,a_2)}$ we obtain $(f')_{\mathfrak{t}',\mathfrak{C}}^{(a_1,a_2)}(\boldsymbol{v}_{\mathfrak{t}'}) = (a_1, a_2) \in \mathrm{Tp}_{\mathfrak{t}'}^{\mathfrak{A}'}$.

If $\mathfrak{D}$ is an induced substructure of $\mathfrak{A}$ already satisfying $\Psi_{\mathrm{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$, then in the construction we have for every pair of elements from D already some type predicate in whose interpretation they are. Consequently the construction does not touch any elements from D, hence $\mathfrak{D}$ is also an induced substructure of $\mathfrak{A}'$. $\square$

Now we will employ the local "problem fixing" of Lemma 79 to ensure $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$. This is split into two lemmata (Lemma 81 and Lemma 82).

Lemma 81. *Given a $\tau \uplus \sigma$-structure $\mathfrak{B}$ satisfying Ψ_{hom} as well as $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ and $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$. Suppose we have a $\tau \uplus \sigma$-structure $\mathfrak{A}$ with a homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$ that already satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. Additionally, for all $\mathfrak{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ without constants from $\mathfrak{A}$ satisfies $\mathrm{Tp}_{\mathfrak{t}} \in \sigma$ in $\mathfrak{A}$ if and only if $\boldsymbol{a}$ realizes $\mathfrak{t}$ in $\mathfrak{A}$. Let $q = (\boldsymbol{a}, \mu, \mathfrak{t}, \nu)$ be a violation tuple of $\mathfrak{A}$. Then there is a finite $\tau \uplus \sigma$-structure $\mathfrak{A}'$ and a homomorphism $h' \colon \mathfrak{A}' \to \mathfrak{B}$ with the following properties:*

1) *$\mathfrak{A}'$ satisfies conjuncts $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$,*

2) *$\mathfrak{A}$ is an induced substructure of $\mathfrak{A}'$,*

3) *h' extends h, i.e., $h'|_A = h$,*

4) *q is no violation tuple of $\mathfrak{A}'$,*

5) *no new violation tuple is being added,*

6) *for all $\mathfrak{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ without constants from $\mathfrak{A}'$ satisfies $\mathrm{Tp}_{\mathfrak{t}} \in \sigma$ in $\mathfrak{A}'$ if and only if $\boldsymbol{a}$ realizes $\mathfrak{t}$ in $\mathfrak{A}'$.*

If $\mathfrak{B}$ and $\mathfrak{A}$ additionally satisfy $\Psi_{\mathrm{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$ we can ensure that $\mathfrak{A}'$ also satisfies $\Psi_{\mathrm{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$.

Proof. Let $q = (\boldsymbol{a}, \mu, \mathfrak{t}, \nu)$ be a violation tuple of $\mathfrak{A}$ which implies $\boldsymbol{a} \in \mathrm{Tp}_{\mathfrak{t}}^{\mathfrak{A}}$. Since h is a homomorphism, the tuple $h(\boldsymbol{a})$ in $\mathfrak{B}$ also satisfies the premise of the conjunct from $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ (see 116) indexed by $(\mathfrak{t}, \nu)$. By assumption, $\mathfrak{B}$ satisfies $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ which implies the existence of $(\mathfrak{t}', \nu')$ and a tuple $\boldsymbol{b}$ from $\mathfrak{B}$ (and hence $\mathfrak{B}^*$) such that $(\mathfrak{t}, \nu) \Subset_j (\mathfrak{t}', \nu')$ and $\boldsymbol{b} \in \mathrm{Tp}_{\mathfrak{t}'}^{\mathfrak{B}}$ hold.

Applying Lemma 79 we obtain a structure $\mathfrak{A}^*$ and a homomorphism $h^* \colon \mathfrak{A}^* \to \mathfrak{B}$ such that q is no violation tuple anymore. We also note two things: For one, we do not add any violation tuples via Lemma 79. This is clear when observing the fact that analyzing only locally the violation tuples of $\mathfrak{A}$ at $\boldsymbol{a}$, i.e. via $Q_{\boldsymbol{a}}$, is enough since the rest of the structure (everything not inside the induced substructure by $\boldsymbol{a}$ of $\mathfrak{A}$) remains completely unmoved by the construction. Secondly, Lemma 79 ensures that $\mathfrak{A}$ is an induced substructure of $\mathfrak{A}^*$. Finally, we obtain that $\mathfrak{A}^*$ satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. This is equally achieved by Lemma 79. Hence Items 1 to 5 are established. Lastly Item 6 is also guaranteed Lemma 79.

If either $\mathfrak{A}$ or $\mathfrak{B}$ does not satisfy $\Psi_{\mathrm{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$, we rename $\mathfrak{A}^*$ to $\mathfrak{A}'$ and h^* to h' and are done. If on the other hand, additionally $\Psi_{\mathrm{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$ holds for $\mathfrak{B}$ and $\mathfrak{A}$, we apply Lemma 80 to $\mathfrak{A}^*$ and $\mathfrak{B}$ and the homomorphism h^* and get an structure $\mathfrak{A}'$. Finally, as $\mathfrak{A}$ does satisfy $\Psi_{\mathrm{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$, again Lemma 80 yields that $\mathfrak{A}^*$ is an induced substructure of $\mathfrak{A}'$. And by transitivity of being induced substructures we preserve Item 2 even after this step. Preservation of the remaining items is ensured by Lemma 80. $\square$

Lemma 82. *Given a $\tau \uplus \sigma$-structure $\mathfrak{B}$ satisfying Ψ_{hom} as well as $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ and $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$. Suppose we have a finite $\tau \uplus \sigma$-structure $\mathfrak{A}$ with a homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$ that already satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. Additionally, for every tuple $\boldsymbol{a}$ without constants from $\mathfrak{A}$ it holds for all $\mathfrak{t} \in \mathcal{E}$ that $\boldsymbol{a}$ satisfies $\mathrm{Tp}_{\mathfrak{t}} \in \sigma$ in $\mathfrak{A}$ if and only if $\boldsymbol{a}$ realizes $\mathfrak{t}$ in $\mathfrak{A}$. Then there is a finite $\tau \uplus \sigma$-structure $\mathfrak{A}'$ and a homomorphism $h' \colon \mathfrak{A}' \to \mathfrak{B}$ with the following properties:*

1) *$\mathfrak{A}'$ satisfies all $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$,*

2) *$\mathfrak{A}$ is an induced substructure of $\mathfrak{A}'$,*

3) *h' extends h, i.e., $h'|_A = h$,*

4) *every conjunct in (116) is satisfied in $\mathfrak{A}'$ if the universal quantification is restricted to elements from $\mathfrak{A}$,*

5) *for all $\mathfrak{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ without constants from $\mathfrak{A}'$ satisfies $\mathrm{Tp}_{\mathfrak{t}} \in \sigma$ in $\mathfrak{A}'$ if and only if $\boldsymbol{a}$ realizes $\mathfrak{t}$ in $\mathfrak{A}'$.*

If $\mathfrak{A}$ additionally satisfies $\Psi_{\mathrm{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$ we can ensure that $\mathfrak{A}'$ also satisfies $\Psi_{\mathrm{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$.

Proof. Let $\mathfrak{B}^*$ denote $\mathfrak{B}|_{\sigma \uplus C}$ and Q be the set of violation tuples of $\mathfrak{A}$. Since Q is clearly finite (as we have finitely many

types to consider and $\mathfrak{A}$ is a finite structure) we can assume an enumeration $\{q_1, \ldots, q_m\}$ of its elements. We define inductively homomorphisms $h^j \colon \mathfrak{A}^j \to \mathfrak{B}^*$ and structures $\mathfrak{A}^j$ for $j \in \{0, \ldots, m\}$ such that for every $j \in \{1, \ldots, m\}$ the following holds

i) $\mathfrak{A}$ is a substructure of $\mathfrak{A}^j$,

ii) A^j is finite,

iii) $\mathfrak{A}^j$ satisfies conjuncts $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$,

iv) $\mathfrak{A}^{j-1}$ is an induced substructure of $\mathfrak{A}^j$,

v) h^j extends h^{j-1}, i.e. $h^j|_{A^{j-1}} = h^{j-1}$,

vi) no element from $\{q_1, \ldots, q_{j-1}\}$ is a violation tuple of $\mathfrak{A}^j$,

vii) at most the elements from $\{q_j, \ldots, q_m\}$ are violation tuples of $\mathfrak{A}^j$, i.e. we do not add new violation tuples in each step),

viii) for all $\mathsf{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ without constants from $\mathfrak{A}^j$ satisfies $\mathrm{Tp}_{\mathsf{t}} \in \sigma$ in $\mathfrak{A}^j$ if and only if $\boldsymbol{a}$ realizes t in $\mathfrak{A}^j$.

We set $\mathfrak{A}^0 := \mathfrak{A}$ and $h^0 := h$. Suppose $\mathfrak{A}^k$ and h^k are defined for all $k \in \{1, \ldots, j\}$ and satisfy the requirements.

Now consider $q_j = (\boldsymbol{a}, \mu, \mathsf{t}, \nu)$ which implies $\boldsymbol{a} \in \mathrm{Tp}_{\mathsf{t}}^{\mathfrak{A}}$. Since $\mathfrak{A}$ is a substructure of $\mathfrak{A}^{j-1}$ we find the tuple $\boldsymbol{a}$ in $\mathfrak{A}^{j-1}$. Hence q_j is also a violation tuple of $\mathfrak{A}^{j-1}$. Applying Lemma 81 we obtain $\mathfrak{A}^j$ and h^j with Items ii)-v) holding. Obviously Item i) holds by transitivity of the substructure relation. Furthermore q_j is not a violation tuple of $\mathfrak{A}^j$, hence Item vi) is true, too. Item 5 of Lemma 81 yields vii) and Item 6 of Lemma 81 yields viii).

This process terminates after m steps and we obtain a structure $\mathfrak{A}^m$ from it, which we will denote now by $\mathfrak{A}'$ and a homomorphism $h^m \colon \mathfrak{A}^m \to \mathfrak{B}^*$ renamed to h'. As every step ensured that Items i)-vii) hold, $\mathfrak{A}'$ is a finite structure,

- that satisfies $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$,

- has $\mathfrak{A}$ as induced substructure (by transitivity of this relation), and

- h' extends h, i.e. $h'|_A = h$.

Hence $\mathfrak{A}'$ with h' satisfy Items 1-3. Since no violation tuple of $\mathfrak{A}$ is one of $\mathfrak{A}'$ and there never have been created new ones along the way, Item 4 is true as well. Item 5 holds thanks to Item viii).

For the second part of the lemma note that if $\Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$ holds for $\mathfrak{B}$ and $\mathfrak{A}$ we can additionally apply in each step the second part of Lemma 81. This then ensures that $\Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$ holds for every A^j with $j \in \{1, \ldots, m\}$ and hence also for $\mathfrak{A}'$. $\qquad\square$

We will now put those lemmata together in order to do the following: For a $\tau \uplus \sigma$-model $\mathfrak{B}$ satisfying Ψ_{hom}, all of the $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ and $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ (and in the case of Φ being a

TGF sentence all the $\Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$) we will construct a $\tau \uplus \sigma$-model $\mathfrak{A}$ of Φ and a homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$. Note that we do not fix a $(\mathcal{E}_+, \mathcal{E}_!) \in (\![\Phi]\!)$ this time.

Lemma 83. *Let Φ be a $\mathbb{GFO}_=$ sentence $\Phi = \Phi_{\mathrm{D}} \wedge \Phi_\forall \wedge \Phi_{\forall\exists}$ with $\mathrm{D} \in \tau$ unary or a $\mathbb{TGF}$ sentence $\Phi = \Phi_{\mathrm{Univ}} \wedge \Phi_\forall \wedge \Phi_{\forall\exists}$ with $\mathrm{Univ} \in \tau$ binary, where*

$$\Phi_{\mathrm{D}} = \forall x.\mathrm{D}(x), \tag{128}$$

$$\Phi_{\mathrm{Univ}} = \forall xy.\mathrm{Univ}(x, y), \tag{129}$$

$$\Phi_\forall = \bigwedge_i \forall \boldsymbol{x}.\Big(\alpha_i(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]\Big) \tag{130}$$

$$\Phi_{\forall\exists} = \bigwedge_j \forall \boldsymbol{z}.\Big(\beta_j(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma_j(\boldsymbol{y}\boldsymbol{z})\Big) \tag{131}$$

Define the $\tau \uplus \sigma$-sentences

$$\Psi = \Psi_{\mathrm{hom}} \wedge \bigvee_{(\mathcal{E}_+,\mathcal{E}_!) \in (\![\Phi]\!)} \Psi_{(\mathcal{E}_+,\mathcal{E}_!)} \tag{132}$$

and

$$\Psi^{\mathbb{TGF}} = \Psi_{\mathrm{hom}} \wedge \bigvee_{(\mathcal{E}_+,\mathcal{E}_!) \in (\![\Phi]\!)} \Psi^{\mathbb{TGF}}_{(\mathcal{E}_+,\mathcal{E}_!)} \tag{133}$$

where $\Psi_{(\mathcal{E}_+,\mathcal{E}_!)}$ is the following sentences:

$$\Psi_{(\mathcal{E}_+,\mathcal{E}_!)} = \Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)} \wedge \Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)} \tag{134}$$

and $\Psi^{\mathbb{TGF}}_{(\mathcal{E}_+,\mathcal{E}_!)}$ is the following sentences:

$$\Psi^{\mathbb{TGF}}_{(\mathcal{E}_+,\mathcal{E}_!)} = \Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)} \wedge \Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)} \wedge \Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)} \tag{135}$$

Depending on Φ being a sentence from $\mathbb{GFO}_=$ or $\mathbb{TGF}$ let $\mathfrak{B}$ be a model of Ψ or $\Psi^{\mathbb{TGF}}$. Then there exists a $\tau \uplus \sigma$-structure $\mathfrak{A}$ satisfying Φ and a homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$.

Proof. Let $\mathfrak{B}$ be a model of Ψ or a model of $\Psi^{\mathbb{TGF}}$ (depending on whether Φ is from $\mathbb{GFO}_=$ or $\mathbb{TGF}$ respectively) and let $(\mathcal{E}_+, \mathcal{E}_!) \in (\![\Phi]\!)$ such that the associated disjunct $\Psi_{(\mathcal{E}_+,\mathcal{E}_!)}$ (or $\Psi^{\mathbb{TGF}}_{(\mathcal{E}_+,\mathcal{E}_!)}$) is satisfied in $\mathfrak{B}$.

Applying Lemma 76, let $\mathfrak{A}_0$ be $(\mathfrak{K}_{\mathcal{E}_+})^\tau$, hence for every type $\mathsf{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ (without elements named by constants) from $\mathfrak{A}_0$ satisfies Tp_{t} in $\mathfrak{A}_0$ if and only if $\boldsymbol{a}$ realizes t in $\mathfrak{A}_0$. Let h_0 be the in Lemma 76 constructed homomorphism $h' \colon (\mathfrak{K}_{\mathcal{E}_+})^\tau \to \mathfrak{B}$. If Φ is a $\mathbb{TGF}$ sentence, Lemma 80 allows us to assume $\mathfrak{A}_0$ to satisfy $\Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$.

This will constitute the induction beginning for a proof of the following statement:

<u>Claim:</u> There exist a sequence $(\mathfrak{A}_n)_{n \in \mathbb{N}}$ of $\tau \uplus \sigma$-structures and a sequence $(h_n)_{n \in \mathbb{N}}$ of homomorphisms $h_n \colon \mathfrak{A}_n \to \mathfrak{B}$ such that for every $i \in \mathbb{N}$, the following holds:

1) A_i is finite,

2) $\mathfrak{A}_i$ satisfies all $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ (and if Φ is from $\mathbb{TGF}$ also $\Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$),

3) $\mathfrak{A}_i$ is an induced substructure of $\mathfrak{A}_{i+1}$,

4) h_{i+1} extends h_i, i.e., $h_{i+1}|_{A_i} = h_i$,

5) every conjunct $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ is satisfied in $\mathfrak{A}_{i+1}$ if the universal quantification is restricted to elements from $\mathfrak{A}_i$,

6) for all $\mathsf{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ without constants from $\mathfrak{A}_i$ satisfies $\mathrm{Tp}_{\mathsf{t}} \in \sigma$ in $\mathfrak{A}_i$ if and only if $\boldsymbol{a}$ realizes t in $\mathfrak{A}_i$.

As mentioned before, we prove the statement by induction on n and assume therefore that the claim holds for structures $(\mathfrak{A}_n)_{n \leq i}$. In each induction step we apply Lemma 82 to obtain $\mathfrak{A}^{i+1}$ from $\mathfrak{A}^i$. This immediately yields the claim.

We continue by using the claim for the proof of $[\![\Phi]\!]|_{\tau}^{\rightarrow} \supseteq [\![\Psi]\!]|_{\tau}$ by defining a limit structure for the sequence $(\mathfrak{A}_n)_{n \in \mathbb{N}}$. By Item 3, we know that $A_i \subseteq A_{i+1}$ as well as $\mathrm{R}^{\mathfrak{A}_i} \subseteq \mathrm{R}^{\mathfrak{A}_{i+1}}$ for every $\mathrm{R} \in \sigma$ holds. We define the $\tau \uplus \sigma$-structure $\mathfrak{A}$ as follows: The domain is given by $A := \bigcup_{n \in \mathbb{N}} A_n$ and for $\mathrm{R} \in \tau \setminus C$ we have $\mathrm{R}^{\mathfrak{A}} := \bigcup_{n \in \mathbb{N}} \mathrm{R}^{\mathfrak{A}_n}$. By the choice of $\mathfrak{A}_0$, Item 2 and Item 5 imply that the structure $\mathfrak{A}$ satisfies $\Psi_{(\mathcal{E}_+,\mathcal{E}_!)}$. If Φ is a $\mathbb{TGF}$ sentence, we also obtain $\mathfrak{A} \vDash \Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$. Additionally, Item 5 guarantees that $\mathfrak{A}$ satisfies $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$. If this were not the case we would find some violation tuple $q = (\boldsymbol{a}, \mu, \mathsf{t}, \nu)$ of $\mathfrak{A}$. As $\mathfrak{A}$ is the union of an increasing sequence of structures, there is an index k such that q is a violation tuple of $\mathfrak{A}_k$. This also means that $\boldsymbol{a}$ is a tuple from $\mathfrak{A}_k$. By construction (applying Lemma 79) q is no violation tuple in $\mathfrak{A}_{k+1}$. Since in every application of Lemma 79 we never add a fresh violation tuple this implies, that q could not have been one of $\mathfrak{A}$. As violation tuples are witnessing violations of $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ and we just argued that there cannot be any violation tuples in $\mathfrak{A}$, this yields $\mathfrak{A} \vDash \Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$. Finally, Item 6 guarantees that type predicates correspond to the associated types. To be more precise: For every type $\mathsf{t} \in \mathcal{E}$ it holds that a tuple $\boldsymbol{a}$ (without constants) from $\mathfrak{A}$ satisfies Tp_{t} in $\mathfrak{A}$ if and only if $\boldsymbol{a}$ realizes t in $\mathfrak{A}$. Note that, since $\mathfrak{A} \vDash \Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ the only interesting case is $\mathsf{t} \in \mathcal{E}_+$. This allows us to switch freely between $\mathsf{t} \in \mathcal{E}$ and $\mathrm{Tp}_{\mathsf{t}} \in \sigma$ for any further discussions.

Hence we can establish the validity of the τ-sentence Φ based on the type predicates in $\mathfrak{A}$ alone. To be precise, consider at first (130) and suppose $\boldsymbol{a}$ from $\mathfrak{A}$ satisfies the premise $\alpha_i(\boldsymbol{x})$ of some conjunct from $\Phi_\forall$. Then the τ type $\mathsf{t}_{\boldsymbol{a}}$ realized by $\boldsymbol{a}$ in $\mathfrak{A}$ is guarded by $\alpha_i(\boldsymbol{x})$ (as witnessed by some ν). Consequently $\mathsf{t}_{\boldsymbol{a}} \in \mathcal{E}$ and hence $\boldsymbol{a} \in \mathrm{Tp}_{\mathsf{t}_{\boldsymbol{a}}}^{\mathfrak{A}}$ by construction. With $\mathfrak{A} \vDash \Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ we obtain $\mathsf{t}_{\boldsymbol{a}} \in \mathcal{E}_+$. As $(\mathcal{E}_+, \mathcal{E}_!) \in (\!|\Phi|\!)$, $\mathsf{t}_{\boldsymbol{a}}$ contains some disjunct from $\vartheta_i[\boldsymbol{v}_{\mathsf{t}_{\boldsymbol{a}}}]$. Hence $\boldsymbol{a}$ satisfies some disjunct from $\vartheta_i[\boldsymbol{x}]$ in $\mathfrak{A}$.

Now consider (131) and assume that a premise $\beta_j(\boldsymbol{z})$ of a conjunct from $\Phi_{\forall\exists}$ is satisfied by a tuple $\boldsymbol{c}$ from $\mathfrak{A}$. Hence the τ type $\mathsf{t}_{\boldsymbol{c}}$ of $\boldsymbol{c}$ in $\mathfrak{A}$ is guarded by $\beta_j(\boldsymbol{z})$ (as witnessed by some ν), so $\mathsf{t}_{\boldsymbol{c}} \in \mathcal{E}$. By construction this means that $\boldsymbol{c}$ satisfies $\mathrm{Tp}_{\mathsf{t}_{\boldsymbol{c}}}$ in $\mathfrak{A}$. By $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ we obtain $\mathsf{t}_{\boldsymbol{c}} \in \mathcal{E}_+$ which in turn implies that $(\mathsf{t}_{\boldsymbol{c}}, \nu) \in B_j$. Since $\mathfrak{A}$ satisfies Ψ and especially $\Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}$, it follows that there exist $\boldsymbol{b}$ in $\mathfrak{A}$ and a type $\mathsf{t}' \in \mathcal{E}_+$ guarded by $\gamma_j(\boldsymbol{yz})$ as witnessed by ν',

i.e. $(\mathsf{t}', \nu') \in P_j$ such that $(\mathsf{t}_{\boldsymbol{c}}, \nu)$ and (t', ν') are fitting and $\mathrm{Tp}_{\mathsf{t}'}(\boldsymbol{bc})$ holds. And by construction we obtain $\boldsymbol{bc}$ realizing t' hence satisfying $\gamma_j(\boldsymbol{yz})$ in $\mathfrak{A}$.

For the case Φ being a $\mathbb{GFO}_=$ sentence, Φ_{D} holds since, for every a in $\mathfrak{A}$ its type $\mathsf{t}_a(v_1)$ is, as a 1-type, trivially guarded (hence in $\mathcal{E}$). By construction $a \in \mathrm{Tp}_{\mathsf{t}_a}^{\mathfrak{A}}$, hence by $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ we obtain $\mathsf{t}_a \in \mathcal{E}_+$. As $\mathrm{D}(v_1) \in \mathsf{t}_a$, $a \in \mathrm{D}^{\mathfrak{A}}$.

Lastly, if Φ is a $\mathbb{TGF}$ sentence, Φ_{Univ} holds since, for every (a_1, a_2) in $\mathfrak{A}$ we obtain via $\mathfrak{A} \vDash \Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$ its type $\mathsf{t}_a(v_1)$ is, as a 1-type, trivially guarded (hence in $\mathcal{E}$). By construction $a \in \mathrm{Tp}_{\mathsf{t}_a}^{\mathfrak{A}}$, hence by $\Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)}$ we obtain $\mathsf{t}_a \in \mathcal{E}_+$. As $\mathrm{D}(v_1) \in \mathsf{t}_a$, $a \in \mathrm{D}^{\mathfrak{A}}$.

This implies that $\mathfrak{A}$ indeed satisfies Φ, as desired.

By Item 4, the union $h := \bigcup_{n \in \mathbb{N}} h_n$ is a well-defined map. Moreover, it is a homomorphism from $\mathfrak{A}$ to $\mathfrak{B}$. This proves the statement. $\qquad\square$

We are now ready to prove several results characterizing the homclosures.

Proposition 84. *For every* $\mathbb{GFO}_=$ *τ-sentence Φ, there exists a* $\mathbb{GFO}_=$ *sentence Ψ such that* $[\![\Phi]\!]^{\rightarrow} = [\![\Psi]\!]|_{\tau}$.

Proof. Let Φ be a $\mathbb{GFO}_=$ τ-sentence, which, by Lemma 65, can without loss of generality be assumed to be in the following normal form (as discussed above in Lemma 70): $\Phi = \Phi_{\mathrm{D}} \wedge \Phi_\forall \wedge \Phi_{\forall\exists}$ with $\mathrm{D} \in \tau$ unary, where

$$\Phi_{\mathrm{D}} = \forall x.\mathrm{D}(x), \tag{136}$$

$$\Phi_\forall = \bigwedge_i \forall \boldsymbol{x}.\Big(\alpha_i(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]\Big) \tag{137}$$

$$\Phi_{\forall\exists} = \bigwedge_j \forall \boldsymbol{z}.\Big(\beta_j(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma_j(\boldsymbol{yz})\Big) \tag{138}$$

with ϑ_i a disjunction of literals, α_i, β_j guard atoms and γ_j an atom. We define the *width* of Φ (denoted by $\mathrm{width}(\Phi)$) to be the maximal arity of a relation symbol appearing in Φ. Furthermore, we denote by C the set of all constant symbols from τ.

Now, let the set $\mathcal{E}$ of eligible types consist of all rigid guarded types of order $\leq \mathrm{width}(\Phi)$.

Let σ contain for every $\mathsf{t} \in \mathcal{E}$ a fresh relational symbol Tp_{t} of appropriate arity, and define the $\tau \uplus \sigma$-sentence

$$\Psi = \Psi_{\mathrm{hom}} \wedge \bigvee_{(\mathcal{E}_+,\mathcal{E}_!) \in (\!|\Phi|\!)} \Psi_{(\mathcal{E}_+,\mathcal{E}_!)}, \tag{139}$$

where $\Psi_{(\mathcal{E}_+,\mathcal{E}_!)}$ is the following sentences:

$$\Psi_{(\mathcal{E}_+,\mathcal{E}_!)} = \Psi_{\mathrm{gen},(\mathcal{E}_+,\mathcal{E}_!)} \wedge \Psi_{\mathrm{guard},(\mathcal{E}_+,\mathcal{E}_!)}. \tag{140}$$

We obtain $[\![\Phi]\!]^{\rightarrow} = [\![\Psi]\!]|_{\tau}$ as follows:

"$\subseteq$": Let $\mathfrak{A} \in [\![\Phi]\!]$ and let $h : \mathfrak{A} \to \mathfrak{B}$. Take $\mathcal{E}$ to be the set of all eligible types for Φ and σ the set of fresh relation symbols

Tp_t associated to each $t \in \mathcal{E}$ with the appropriate arity. Then the $\mathcal{E}$-adornment of $\mathfrak{A}$, $\mathfrak{A} \cdot \mathcal{E}$, is a $\tau \uplus \sigma$-structure which we will denote by $\mathfrak{A}^*$. Furthermore, let $\mathfrak{B} \cdot h(\mathfrak{A}^*|_\sigma)$ be the $\tau \uplus \sigma$-structure described before, which we denote by $\mathfrak{B}^*$. Obviously $\mathfrak{B}^*|_\tau = \mathfrak{B}$ holds.

By Lemma 68 $\mathfrak{B}^*$ satisfies $\Psi_{\text{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. Additionally, after applying Lemma 78 $\mathfrak{B}^*$ satisfies $\Psi_{\text{guard},(\mathcal{E}_+,\mathcal{E}_!)}$ and Ψ_{hom}. Consequently $\mathfrak{B}^* \vDash \Psi$.

"$\supseteq$": Let $\mathfrak{B}$ be a model of Ψ. Application of Lemma 83 yields a $\tau \uplus \sigma$-structure $\mathfrak{A}$ satisfying Φ and a homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$. Restricting both $\mathfrak{A}$ and $\mathfrak{B}$ to τ yields the claim. $\qquad \square$

Lemma 85. *For every $\mathbb{GNFO}_=$ τ-sentence Φ, there exists some $\mathbb{GFO}_=$ sentence Φ^* with $[\![\Phi]\!]^\rightarrow = [\![\Phi^*]\!]^\rightarrow_\tau$.*

Proof. By [1, Proposition 3.2], there exist a $\mathbb{GFO}_=$ sentence Φ' and an $\exists^*\mathbb{FO}^+_=$ sentence (more precisely: a union of conjunctive queries) Φ'' such that $[\![\Phi]\!] = [\![\Phi' \wedge \neg\Phi'']\!]|_\tau$ (†).

Then, by [2, Lemma 2.4] and the corresponding proof, there exists a $\mathbb{GFO}_=$ sentence Φ''' such that (i) $[\![\Phi' \wedge \neg\Phi'']\!] \subseteq [\![\Phi' \wedge \neg\Phi''']\!]$ and (ii) for every model $\mathfrak{A}$ of $\Phi' \wedge \neg\Phi'''$ there is a model $\mathfrak{B}$ of $\Phi' \wedge \neg\Phi''$ such that a homomorphism $\mathfrak{B} \to \mathfrak{A}$ exists. Therefore, $[\![\Phi' \wedge \neg\Phi'']\!]^\rightarrow = [\![\Phi' \wedge \neg\Phi''']\!]^\rightarrow$ (‡).

So we obtain:

$$
\begin{aligned}
[\![\Phi]\!]^\rightarrow \;\; &\overset{\dagger}{=} \;\; [\![\Phi' \wedge \neg\Phi'']\!]^\rightarrow_\tau && (141)\\
&\overset{*}{=} \;\; [\![\Phi' \wedge \neg\Phi'']\!]^\rightarrow|_\tau && (142)\\
&\overset{\ddagger}{=} \;\; [\![\Phi' \wedge \neg\Phi''']\!]^\rightarrow|_\tau && (143)\\
&\overset{*}{=} \;\; [\![\Phi' \wedge \neg\Phi''']\!]^\rightarrow_\tau. && (144)
\end{aligned}
$$

The equivalent transformation used in the steps denoted by $\overset{*}{=}$ has been justified in the proof of Lemma 65. Note that $\Phi' \wedge \neg\Phi''' \in \mathbb{GFO}_=$ and let $\Phi^* = \Phi' \wedge \neg\Phi'''$. $\qquad \square$

Proposition 86. *For every $\mathbb{GNFO}_=$ τ-sentence Φ, there exists a $\mathbb{GFO}_=$ sentence Ψ such that $[\![\Phi]\!]^\rightarrow = [\![\Psi]\!]|_\tau$.*

Proof. Given Φ, we apply Lemma 85 to obtain a $\mathbb{GFO}_=$ τ'-sentence Φ^* for some $\tau' \supseteq \tau$ satisfying $[\![\Phi]\!]^\rightarrow = [\![\Phi^*]\!]^\rightarrow_{\tau'}$. Applying Proposition 84 to Φ^*, we obtain a $\mathbb{GFO}_=$ τ''-sentence Ψ for some $\tau'' \supseteq \tau'$ satisfying $[\![\Phi^*]\!]^\rightarrow = [\![\Psi]\!]|_{\tau'}$. Yet, since $\tau \subseteq \tau'$, we also obtain

$$
[\![\Psi]\!]|_\tau = [\![\Psi]\!]|_{\tau'}|_\tau = [\![\Phi^*]\!]^\rightarrow|_\tau \overset{*}{=} [\![\Phi^*]\!]^\rightarrow_\tau = [\![\Phi]\!]^\rightarrow
$$

(the equality denoted by $\overset{*}{=}$ has been shown in the proof of Lemma 65), therefore Ψ serves the desired purpose. $\qquad \square$

Proposition 87. *For every $\mathbb{TGF}$ τ-sentence Φ, there exists a $\mathbb{TGF}$ sentence Ψ such that $[\![\Phi]\!]^\rightarrow = [\![\Psi]\!]|_\tau$.*

Proof. Let Φ be a $\mathbb{TGF}$ τ-sentence which we assume to be in normal form (see Corollary 71), i.e. with $\text{Univ} \in \tau$

$$
\Phi = \Phi_{\text{Univ}} \wedge \Phi_\forall \wedge \Phi_{\forall\exists} \qquad (145)
$$

where

$$
\Phi_{\text{Univ}} \;\; = \;\; \forall xy.\text{Univ}(x,y), \qquad (146)
$$

$$
\Phi_\forall \;\; = \;\; \bigwedge_i \forall \boldsymbol{x}.(\alpha_i(\boldsymbol{x}) \Rightarrow \vartheta_i[\boldsymbol{x}]), \text{ and} \qquad (147)
$$

$$
\Phi_{\forall\exists} \;\; = \;\; \bigwedge_j \forall \boldsymbol{z}.(\beta_j(\boldsymbol{z}) \Rightarrow \exists \boldsymbol{y}.\gamma_j(\boldsymbol{y}\boldsymbol{z})), \qquad (148)
$$

with α_i, β_j guard atoms, γ_j just atoms, and ϑ_i being a disjunction of literals. We define the *width* of Φ (denoted by $\text{width}(\Phi)$) to be the maximal arity of a relation symbol appearing in Φ. We denote by C the set of all constant symbols from τ.

Let the set $\mathcal{E}$ of eligible types consist of all rigid guarded types of order $\leq \text{width}(\Phi)$.

Let σ contain for each $t \in \mathcal{E}$ a fresh relational symbols Tp_t. Then we define the $\tau \uplus \sigma$-sentence

$$
\Psi = \Psi_{\text{hom}} \wedge \bigvee_{(\mathcal{E}_+,\mathcal{E}_!) \in (\![\Phi]\!)} \Psi_{(\mathcal{E}_+,\mathcal{E}_!)}, \qquad (149)
$$

where $\Psi_{(\mathcal{E}_+,\mathcal{E}_!)}$ is the sentence:

$$
\Psi_{(\mathcal{E}_+,\mathcal{E}_!)} = \Psi_{\text{gen},(\mathcal{E}_+,\mathcal{E}_!)} \wedge \Psi_{\text{guard},(\mathcal{E}_+,\mathcal{E}_!)} \wedge \Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}. \qquad (150)
$$

We now show $[\![\Phi]\!]^\rightarrow = [\![\Psi]\!]|_\tau$ by proving the two inclusions separately:

"$\subseteq$": Let $\mathfrak{A} \in [\![\Phi]\!]$ and let $h \colon \mathfrak{A} \to \mathfrak{B}$. Take $\mathcal{E}$ to be the set of all eligible types for Φ and σ the set of fresh relation symbols Tp_t associated to each $t \in \mathcal{E}$ with the appropriate arity. Then the $\mathcal{E}$-adornment of $\mathfrak{A}$, $\mathfrak{A} \cdot \mathcal{E}$, is a $\tau \uplus \sigma$-structure which we will denote by $\mathfrak{A}^*$. Furthermore, let $\mathfrak{B} \cdot h(\mathfrak{A}^*|_\sigma)$ be the $\tau \uplus \sigma$-structure described before, which we denote by $\mathfrak{B}^*$. Obviously $\mathfrak{B}^*|_\tau = \mathfrak{B}$ holds.

By Lemma 68 $\mathfrak{B}^*$ satisfies $\Psi_{\text{gen},(\mathcal{E}_+,\mathcal{E}_!)}$. Additionally, after applying Lemma 78 $\mathfrak{B}^*$ satisfies $\Psi_{\text{guard},(\mathcal{E}_+,\mathcal{E}_!)}$, Ψ_{hom}, and $\Psi_{\mathbb{TGF},(\mathcal{E}_+,\mathcal{E}_!)}$. Consequently $\mathfrak{B}^* \vDash \Psi$.

"$\supseteq$": Let $\mathfrak{B}$ be a model of Ψ. Application of Lemma 83 yields a $\tau \uplus \sigma$ structure $\mathfrak{A}$ satisfying Φ and a homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$. Restricting both $\mathfrak{A}$ and $\mathfrak{B}$ to τ yields the claim. $\qquad \square$

We can "piggyback" on this result to obtain the comparable statement for $\exists^*\forall\forall\exists^*\mathbb{FO}$.

Proposition 88. *For every $\exists^*\forall\forall\exists^*\mathbb{FO}$ τ-sentence Φ, there exists a $\mathbb{TGF}$ sentence Ψ such that $[\![\Phi]\!]^\rightarrow = [\![\Psi]\!]|_\tau$.*

Proof. Φ can be projectively characterized by a $\mathbb{TGF}$ sentence Φ' obtained as follows: skolemize the outer existential quantifiers into extra constants and add an atom using a fresh predicate of adequate arity as a guard following every inner existential quantifier. Clearly, Φ' is in $\mathbb{TGF}$ and it projectively characterizes Φ, so the former can be seen as the "$\mathbb{TGF}$ normal form" of the latter. Let now Ψ be obtained from Φ' following Proposition 87. We then obtain $[\![\Phi]\!]^\rightarrow = [\![\Psi]\!]|_\tau$ thanks to Lemma 65. $\qquad \square$

Definition 45. *Given an instance of 3SAT $\mathcal{S} = \{C_1, \ldots, C_n\}$ consisting of n clauses each containing 3 literals over propositional variables $p_1, \ldots, p_m$, we define the structure $\mathfrak{A}_{\mathcal{S}}$ by*

$$A_{\mathcal{S}} = \{a_i \mid 1 \leq i \leq n+1\} \cup \{b_i, b_i' \mid 1 \leq i \leq m\} \quad (28)$$

$$\mathtt{First}^{\mathfrak{A}_{\mathcal{S}}} = \{a_1\} \quad (29)$$

$$\mathtt{CLst}^{\mathfrak{A}_{\mathcal{S}}} = \{a_i \mid 1 \leq i \leq n\} \quad (30)$$

$$\mathtt{Nil}^{\mathfrak{A}_{\mathcal{S}}} = \{a_{n+1}\} \quad (31)$$

$$\mathtt{Next}^{\mathfrak{A}_{\mathcal{S}}} = \{(a_i, a_{i+1}) \mid 1 \leq i \leq n\} \quad (32)$$

$$\mathtt{Lit}^{\mathfrak{A}_{\mathcal{S}}} = \{(a_i, b_j) \mid p_j \in C_i\} \cup \{(a_i, b_j') \mid \neg p_j \in C_i\} \quad (33)$$

$$\mathtt{Sel}^{\mathfrak{A}_{\mathcal{S}}} = \{b_i, b_i' \mid 1 \leq i \leq m\} \quad (34)$$

$$\mathtt{Cmp}^{\mathfrak{A}_{\mathcal{S}}} = \{(b_i, b_j), (b_i', b_j'), (b_i, b_k'), (b_i', b_k) \mid i \neq k\} \quad (35)$$

We let $\Phi_{3\text{SAT}}$ be the sentence containing the conjunction of the following sentences (leading universal quantifier omitted):

$$\exists x.\mathtt{First}(x) \quad (36)$$

$$\mathtt{First}(x) \Rightarrow \mathtt{CLst}(x) \quad (37)$$

$$\mathtt{CLst}(x) \Rightarrow \exists y.\mathtt{Next}(x, y) \wedge \big(\mathtt{CLst}(y) \vee \mathtt{Nil}(y)\big) \quad (38)$$

$$\mathtt{CLst}(x) \Rightarrow \exists y.\mathtt{Lit}(x, y) \wedge \mathtt{Sel}(y) \quad (39)$$

$$\mathtt{Sel}(x) \wedge \mathtt{Sel}(y) \Rightarrow \mathtt{Cmp}(x, y). \quad (40)$$

Lemma 89. *1) $\Phi_{3\text{SAT}}$ is contained in constant-free, equality-free $\mathbb{FO}^2$ (and therefore in $\mathbb{FO}^2_=$ and $\mathbb{TGF}$).*

2) There exists a sentence Ψ in constant-free, equality-free $\forall\forall\exists\exists\mathbb{FO}$ (and therefore in $\exists^*\forall\forall\exists^*\mathbb{FO}$) such that $\llbracket\Psi\rrbracket^{\rightarrow}_{(<\omega)}|_\tau \subseteq \llbracket\Phi_{3\text{SAT}}\rrbracket^{\rightarrow}_{(<\omega)}$.

3) The size of $\mathfrak{A}_{\mathcal{S}}$ is polynomial in the size of $\mathcal{S}$.

4) $\bigwedge_{\{\ell,\ell',\ell''\}\in\mathcal{S}} \ell \vee \ell' \vee \ell''$ is satisfiable exactly if $\mathfrak{A}_{\mathcal{S}} \in \llbracket\Phi_{3\text{SAT}}\rrbracket^{\rightarrow}_{(<\omega)}$.

5) Checking membership in $\llbracket\Phi_{3\text{SAT}}\rrbracket^{\rightarrow}_{(<\omega)}$ is NP-hard in the size of the structure.

6) Checking membership in $\llbracket\Psi\rrbracket^{\rightarrow}_{(<\omega)}$ is NP-hard in the size of the structure.

Proof. Item 1 follows immediately from the definition of $\Phi_{3\text{SAT}}$.

For Item 2: We let $\Psi = \forall xy \exists zv.\psi$ where ψ is the conjunction over

$$\mathtt{First}(z) \quad (151)$$

$$\mathtt{Eq}(x, x) \quad (152)$$

$$\mathtt{First}(x) \Rightarrow \mathtt{CLst}(x) \quad (153)$$

$$\mathtt{CLst}(x) \wedge \mathtt{Eq}(x, y) \Rightarrow \mathtt{Lit}(x, v) \wedge \mathtt{Sel}(v) \quad (154)$$

$$\mathtt{CLst}(x) \wedge \mathtt{Lit}(x, y) \Rightarrow \mathtt{Next}(x, v) \\ \wedge \big(\mathtt{CLst}(v) \vee \mathtt{Nil}(v)\big) \quad (155)$$

$$\mathtt{Sel}(x) \wedge \mathtt{Sel}(y) \Rightarrow \mathtt{Cmp}(x, y). \quad (156)$$

The containment $\llbracket\Psi\rrbracket_{(<\omega)}|_\tau \subseteq \llbracket\Phi_{3\text{SAT}}\rrbracket_{(<\omega)}$ follows immediately and therefore also $\llbracket\Psi\rrbracket^{\rightarrow}_{(<\omega)}|_\tau \subseteq \llbracket\Phi_{3\text{SAT}}\rrbracket^{\rightarrow}_{(<\omega)}$ holds.

Item 3 follows directly from the definition of $\mathfrak{A}_{\mathcal{S}}$.

For Item 4: Let $V = \{p_1, \ldots, p_m\}$ be the variable set of the propositional formula $\varphi := \bigwedge_{\{\ell,\ell',\ell''\}\in\mathcal{S}} \ell \vee \ell' \vee \ell''$.

Assume that there exists an assignment $\alpha\colon V \to \{\top, \bot\}$ that witnesses the satisfiability of φ. Let τ be the signature of $\mathfrak{A}_{\mathcal{S}}$ and let $\mathfrak{B}'$ be the $\tau\backslash\{\mathtt{Sel}\}$-reduct of $\mathfrak{A}_{\mathcal{S}}$. The τ-structure $\mathfrak{B}$ is the expansion of $\mathfrak{B}'$ where the relation $\mathtt{Sel}^{\mathfrak{B}}$ is defined as follows:

1) $\mathtt{Sel}^{\mathfrak{B}} \subset \{b_i, b_i' \mid 1 \leq i \leq m\}$,

2) $b_i \in \mathtt{Sel}^{\mathfrak{B}}$ if and only $\alpha(p_i) = \top$,

3) $b_i' \in \mathtt{Sel}^{\mathfrak{B}}$ if and only $\alpha(p_i) = \bot$.

One can easily check that $\mathfrak{B}$ satisfies $\Phi_{3\text{SAT}}$. Furthermore, the identity map from $\mathfrak{B}$ to $\mathfrak{A}_{\mathcal{S}}$ is a homomorphism which proves that $\mathfrak{A}_{\mathcal{S}} \in \llbracket\Phi_{3\text{SAT}}\rrbracket^{\rightarrow}_{(<\omega)}$ holds.

For the opposite direction suppose that there exists a structure $\mathfrak{B}$ from $\llbracket\Phi_{3\text{SAT}}\rrbracket_{(<\omega)}$ that has a homomorphism h to $\mathfrak{A}_{\mathcal{S}}$. Note that Equation 36 of Definition 45 ensures that the structure $\mathfrak{B}$ contains an element x_f such that $x_f \in \mathtt{First}^{\mathfrak{B}}$ holds. Since h is a homomorphism $h(x_f) = a_1$ holds. For every element x in $\mathtt{CLst}^{\mathfrak{B}}$ exists by Equation 38 an element y such that $(x, y) \in \mathtt{Next}^{\mathfrak{B}}$ holds. In the structure $\mathfrak{A}_{\mathcal{S}}$ the same holds but with unique elements y. This implies that there exists a set $W \subseteq \mathtt{CLst}^{\mathfrak{B}}$ with $x_f \in W$ such that h restricted to W is a bijection from W to $\{a_i \mid 1 \leq i \leq n+1\}$. We denote this restricted map by h'. Let $f\colon W \to \mathtt{Sel}^{\mathfrak{B}}$ be a map such that $(x, f(x)) \in \mathtt{Lit}^{\mathfrak{B}}$ holds for all $x \in W$. Such a map exists by Equation 39. Observe that $h(f(W)) \subset \{b_i, b_i' \mid 1 \leq i \leq m\}$ holds. Moreover since we have $f(W) \subset \mathtt{Sel}^{\mathfrak{B}}$ by defintion, all pairs of distinct elements in $f(W)$ are in the relation $\mathtt{Cmp}^{\mathfrak{B}}$ by Equation 40. This implies that for all $i \in \{1, \ldots, m\}$ at most one of b_i and b_i' is in $h(f(W))$. We define a variable assignment $\alpha\colon V \to \{\top, \bot\}$ as follows:

1) $\alpha(p_i) := \top$ if $b_i \in h(f(W))$,

2) $\alpha(p_i) := \bot$ if $b_i' \in h(f(W))$,

3) $\alpha(p_i) := \top$ if b_i and b_i' are not in $h(f(W))$.

This is a well defined map by what we saw before. Furthermore, to see that the clause C_j is satisfied, consider the element $y := h'^{-1}(a_j) \in Q$. By the definition of f we have that $(y, f(y)) \in \mathtt{Lit}^{\mathfrak{B}}$ and therefore $(h(y), h(f(y))) \in \mathtt{Lit}^{\mathfrak{A}_{\mathcal{S}}}$ holds. Assume without loss of generality that $h(f(y)) = b_l$ for some $l \in \{1, \ldots, m\}$. Then we have $(a_j, b_l) \in \mathtt{Lit}^{\mathfrak{A}_{\mathcal{S}}}$ which means by the definition of $\mathfrak{A}_{\mathcal{S}}$ that p_l is a literal of C_j. By the definition of α we have $\alpha(p_l) = \top$ which proves that C_j is satisfied.

For Item 5: The 3SAT problem is a well known NP-hard problem [19]. We have a reduction from the 3SAT problem to the membership problem for $\llbracket\Phi_{3\text{SAT}}\rrbracket^{\rightarrow}_{(<\omega)}$ by mapping a 3SAT instance $\mathcal{S}$ to the structure $\mathfrak{A}_{\mathcal{S}}$. This is indeed a reduction by Item 4 and it is polynomial time computable by Item 3.

For Item 6: We reduce the 3SAT problem to the membership problem of $\llbracket\Psi\rrbracket^{\rightarrow}_{(<\omega)}$. We map an instance $\mathcal{S}$ to the structure $\mathfrak{A}_{\mathcal{S}}^*$ that is the expansion of $\mathfrak{A}_{\mathcal{S}}$ by the relation $\mathtt{Eq}^{\mathfrak{A}_{\mathcal{S}}} = \{(x, x) \mid$

$x \in A_{\mathcal{S}}\}$.

Assume that $\mathcal{S}$ is satisfiable. Let $\mathfrak{B}$ be the structure from the first direction of 5). It is easy to see that its expansion $\mathfrak{B}^*$ satisfies Ψ and therefore $\mathfrak{A}_{\mathcal{S}}^*$ is in $[\![\Psi]\!]_{(<\omega)}^{\rightarrow}$.

For the other direction assume that $\mathfrak{A}_{\mathcal{S}}^* \in [\![\Psi]\!]_{(<\omega)}^{\rightarrow}$ holds. By 2) we get that then also $\mathfrak{A}_{\mathcal{S}} \in [\![\Phi_{3\text{SAT}}]\!]_{(<\omega)}^{\rightarrow}$ holds which implies by 4) that $\mathcal{S}$ is satisfiable. $\qquad\square$

Proposition 46. *Checking homclosure membership is NP-hard for constant-free, equality-free $\mathbb{FO}^2$ and $\forall\forall\exists\exists\mathbb{FO}$ with predicate arity ≤ 2.*

Proof. The proposition follows from Item 1, Item 2, Item 5 and Item 6 of Lemma 89. $\qquad\square$

Theorem 47. *Homclosures of $\mathbb{GNFO}_=$ and $\mathbb{GFO}_=$ sentences can be characterized in $\mathbb{FO}_=^{\mathbf{lfp}}$. Thus, for a fixed $\mathbb{GNFO}_=$ or $\mathbb{GFO}_=$ sentence Φ, checking $\mathfrak{A} \in [\![\Phi]\!]^{\rightarrow}$ is P-complete.*

Proof. Given a $\mathbb{GNFO}_=$ τ-sentence Φ, bring it in $\mathbb{GFO}_=$ "homclosure normal form" according to Lemma 85, obtaining a $\mathbb{GFO}_=$ τ'-sentence Φ'. From Φ' compute Ψ as in Proposition 84, adjusting the definition of Ψ_{hom} as follows:

$$\bigwedge_{\mathfrak{t} \in \mathcal{E}} \forall \boldsymbol{v}_{\mathfrak{t}}.\big(\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Rightarrow \bigwedge \mathfrak{t}|_\tau^+\big), \tag{157}$$

where

$$\mathfrak{t}|_\tau^+ = \{\alpha \in \mathfrak{t} \mid \alpha = \mathrm{P}(\boldsymbol{t}) \text{ with } \mathrm{P} \in \tau\} \tag{158}$$

so, compared to the original Ψ_{hom}, the new one disregards auxiliary symbols from $\tau' \setminus \tau$. Clearly, we still obtain $[\![\Phi]\!]^{\rightarrow} = [\![\Psi]\!]|_\tau$ and hence also $[\![\Phi]\!]^{\rightarrow} = [\![\exists\sigma.\Psi]\!]$ where

$$\exists\sigma.\Psi = \exists\sigma.\Big(\Psi_{\text{hom}} \wedge \bigvee_{(\mathcal{E}_+,\mathcal{E}_!) \in (\!|\Phi|\!)} \Psi_{(\mathcal{E}_+,\mathcal{E}_!)}\Big) \tag{159}$$

$$\equiv \bigvee_{(\mathcal{E}_+,\mathcal{E}_!) \in (\!|\Phi|\!)} \exists\sigma.\big(\Psi_{(\mathcal{E}_+,\mathcal{E}_!)} \wedge \Psi_{\text{hom}}\big) \tag{160}$$

$$=: \bigvee_{(\mathcal{E}_+,\mathcal{E}_!) \in (\!|\Phi|\!)} \exists\sigma.\big(\Psi_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}\big), \tag{161}$$

with $\Psi_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}$ being the conjunction over

$$\exists \boldsymbol{v}_{\mathfrak{t}}.\big(\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}})\big) \quad \text{for } \mathfrak{t} \in \mathcal{E}_+ \tag{162}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Rightarrow \bot\big) \quad \text{for } \mathfrak{t} \in \mathcal{E} \setminus \mathcal{E}_+ \tag{163}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Rightarrow \mathrm{Tp}_{\mathfrak{t}'}(\boldsymbol{v}_{\mathfrak{t}'})\big) \quad \text{for } \mathfrak{t},\mathfrak{t}' \in \mathcal{E}_+, \mathfrak{t}' = \mathfrak{t}|_{\boldsymbol{v}_{\mathfrak{t}'}} \tag{164}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Rightarrow \mathrm{Tp}_{\kappa\mathfrak{t}}(\eta_{\boldsymbol{v}}^{\kappa\boldsymbol{v}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}))\big) \quad \text{for } \mathfrak{t} \in \mathcal{E}, \kappa: \boldsymbol{v} \hookrightarrow \boldsymbol{v} \tag{165}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Rightarrow \bigvee_{(\mathfrak{t},\nu) \in_j^z (\mathfrak{t}',\nu')} \exists \boldsymbol{v}_{\mathfrak{t}'} \setminus \nu(\boldsymbol{z}).\mathrm{Tp}_{\mathfrak{t}'}(\boldsymbol{v}_{\mathfrak{t}'})\big) \quad \text{for } (\mathfrak{t},\nu) \in B_j \tag{166}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Rightarrow \bigwedge \mathfrak{t}|_\tau^+\big) \quad \text{for } \mathfrak{t} \in \mathcal{E} \tag{167}$$

where j ranges over the number of conjuncts in $\Phi_{\forall\exists}$ (cf. the proof of Proposition 84).

In a next step, we equivalently rewrite all constituents of $\Psi_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}$ as follows (all side conditions remain untouched):

$$\exists \boldsymbol{v}_{\mathfrak{t}}.\neg\big(\neg\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}})\big) \tag{168}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\neg\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \top\big) \tag{169}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\neg\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \neg\mathrm{Tp}_{\mathfrak{t}'}(\boldsymbol{v}_{\mathfrak{t}'})\big) \tag{170}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\neg\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \neg\mathrm{Tp}_{\kappa\mathfrak{t}}(\eta_{\boldsymbol{v}}^{\kappa\boldsymbol{v}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}))\big) \tag{171}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\neg\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \bigwedge_{(\mathfrak{t},\nu) \in_j^z (\mathfrak{t}',\nu')} \forall \boldsymbol{v}_{\mathfrak{t}'} \setminus \nu(\boldsymbol{z}).\neg\mathrm{Tp}_{\mathfrak{t}'}(\boldsymbol{v}_{\mathfrak{t}'})\big) \tag{172}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\neg\mathrm{Tp}_{\mathfrak{t}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \neg\bigwedge \mathfrak{t}|_\tau^+\big) \tag{173}$$

We next let for $\overline{\sigma} = \{\overline{\mathrm{Tp}_{\mathfrak{t}}} \mid \mathrm{Tp}_{\mathfrak{t}} \in \sigma\}$ be a set of "complement predicates", with the intention that $\overline{\mathrm{Tp}_{\mathfrak{t}}}$ is meant to denote the complement relation of $\mathrm{Tp}_{\mathfrak{t}}$. With this in mind, we rewrite the above sentences in the straightforward way, obtaining $\overline{\Psi}_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}$ as the conjunction over the following $\tau \uplus \overline{\sigma}$-sentences:

$$\exists \boldsymbol{v}_{\mathfrak{t}}.\neg\big(\overline{\mathrm{Tp}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}})\big) \quad \text{for } \mathfrak{t} \in \mathcal{E}_+ \tag{174}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\overline{\mathrm{Tp}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \top\big) \quad \text{for } \mathfrak{t} \in \mathcal{E} \setminus \mathcal{E}_+ \tag{175}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\overline{\mathrm{Tp}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \overline{\mathrm{Tp}_{\mathfrak{t}'}}(\boldsymbol{v}_{\mathfrak{t}'})\big) \quad \text{for } \mathfrak{t},\mathfrak{t}' \in \mathcal{E}_+, \mathfrak{t}' = \mathfrak{t}|_{\boldsymbol{v}_{\mathfrak{t}'}} \tag{176}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\overline{\mathrm{Tp}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \overline{\mathrm{Tp}_{\kappa\mathfrak{t}}}(\eta_{\boldsymbol{v}}^{\kappa\boldsymbol{v}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}))\big) \quad \text{for } \mathfrak{t} \in \mathcal{E}, \kappa: \boldsymbol{v} \hookrightarrow \boldsymbol{v} \tag{177}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\overline{\mathrm{Tp}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \bigwedge_{(\mathfrak{t},\nu) \in_j^z (\mathfrak{t}',\nu')} \forall \boldsymbol{v}_{\mathfrak{t}'} \setminus \nu(\boldsymbol{z}).\overline{\mathrm{Tp}_{\mathfrak{t}'}}(\boldsymbol{v}_{\mathfrak{t}'})\big) \quad \text{for } (\mathfrak{t},\nu) \in B_j \tag{178}$$

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\overline{\mathrm{Tp}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \neg\bigwedge \mathfrak{t}|_\tau^+\big) \quad \text{for } \mathfrak{t} \in \mathcal{E} \tag{179}$$

Note that, for the mentioned reasons,

$$\exists\overline{\sigma}.\big(\overline{\Psi}_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}\big) \equiv \exists\sigma.\big(\Psi_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}\big). \tag{180}$$

Also note that the set of sentences of the forms (176), (178), and (179) can be grouped such that, for every $\overline{\mathrm{Tp}_{\mathfrak{t}}} \in \overline{\sigma}$, there exists precisely one sentence of the form

$$\forall \boldsymbol{v}_{\mathfrak{t}}.\big(\overline{\mathrm{Tp}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}}) \Leftarrow \xi_{\overline{\mathrm{Tp}_{\mathfrak{t}}}}[\boldsymbol{v}_{\mathfrak{t}}]\big), \tag{181}$$

where $\xi_{\overline{\mathrm{Tp}_{\mathfrak{t}}}}[\boldsymbol{v}_{\mathfrak{t}}]$ is a $\mathbb{FO}_=$ sentence with only positive occurrences of $\overline{\sigma}$-atoms (obtain $\xi_{\overline{\mathrm{Tp}_{\mathfrak{t}}}}[\boldsymbol{v}_{\mathfrak{t}}]$ as the disjunction of all the bodies of those implications whose head is $\overline{\mathrm{Tp}_{\mathfrak{t}}}(\boldsymbol{v}_{\mathfrak{t}})$). Therefore, the obtained set of sentences gives rise to a monotonic operator over $\tau \uplus \overline{\sigma}$-structures simultaneously inflating all $\overline{\mathrm{Tp}_{\mathfrak{t}}}$ relations. Hence, $\exists\overline{\sigma}.\big(\overline{\Psi}_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}\big)$ is equivalent to the simultaneous inductive $\mathbb{FO}_=$ least fixed-point expression $\overline{\Psi}_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}^{\mathbf{sim\text{-}lfp}}$ defined by

$$\bigwedge_{\mathfrak{t} \in \mathcal{E}_+} \exists\boldsymbol{x}.\neg\Big[\mathbf{lfp}_{\overline{\mathrm{Tp}_{\mathfrak{t}}}}\big\{ \overline{\mathrm{Tp}_{\mathfrak{t}'}}(\boldsymbol{v}_{\mathfrak{t}}) \leftarrow \xi_{\overline{\mathrm{Tp}_{\mathfrak{t}'}}}[\boldsymbol{v}_{\mathfrak{t}}] \mid \mathfrak{t}' \in \mathcal{E} \big\}\Big](\boldsymbol{x}). \tag{182}$$

As simultaneous induction can be expressed using plain least-fixed-point logic [14], there exists some plain $\mathbb{FO}_=^{\mathbf{lfp}}$ sentence $\overline{\Psi}_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}^{\mathbf{lfp}}$ equivalent to $\overline{\Psi}_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}^{\mathbf{sim\text{-}lfp}}$. By defining

$$\Psi^{\mathbf{lfp}} = \bigvee_{(\mathcal{E}_+,\mathcal{E}_!) \in (\!|\Phi|\!)} \overline{\Psi}_{(\mathcal{E}_+,\mathcal{E}_!),\text{hom}}^{\mathbf{lfp}}, \tag{183}$$

we obtain $\exists\sigma.\Psi \equiv \Psi^{\mathbf{lfp}}$ and therefore $[\![\Psi^{\mathbf{lfp}}]\!] = [\![\Phi]\!]^{\rightarrow}$ as desired. Using the fact that evaluating $\mathbb{FO}_=^{\mathbf{lfp}}$ sentences over structures can be done in polynomial time [16, 30], we obtain polytime data complexity of homclosure membership. $\qquad\square$

G. Characterizing Finite-Model Homclosures

Proposition 48. $\llbracket \Phi_\infty \rrbracket^{\rightarrow}_{<\omega}$ *cannot be characterized in* $\exists \mathbb{SO}$.

Proof. $\llbracket \Phi_\infty \rrbracket^{\rightarrow}_{<\omega}$ contains precisely all – finite or infinite – $\{P\}$-structures that have a directed P-cycle. Toward a contradiction, let us assume that there were an $\exists \mathbb{SO}$ formula Ψ characterizing $\llbracket \Phi_\infty \rrbracket^{\rightarrow}_{<\omega}$. This means the first-order part of Ψ, denoted Ψ', projectively characterizes $\llbracket \Phi_\infty \rrbracket^{\rightarrow}_{<\omega}$. Now, for a given number k, let $\Gamma^{\varnothing}_k$ be the $\mathbb{FO}$ sentence disallowing directed P-cycles of length k defined by

$$\Gamma^{\varnothing}_k \quad = \quad \forall x_1 \ldots x_k. \neg P(x_k, x_1) \vee \bigvee_{i=1}^{k-1} \neg P(x_i, x_{i+1}). \quad (184)$$

Now define the infinite $\mathbb{FO}$ theory $\Xi := \{\Psi'\} \cup \{\Gamma^{\varnothing}_k \mid k > 0\}$. Obviously any finite subset of Ξ is satisfiable, while Ξ is not. However this contradicts the compactness theorem of $\mathbb{FO}_=$ and therefore the presumed existence of Ψ is refuted. $\square$

Theorem 49. *Let* Φ *be some* $\mathbb{SO}$ *sentence and let* Ψ *be such that* $\llbracket \Psi \rrbracket_{<\omega} = \llbracket \Phi \rrbracket^{\rightarrow <\omega}_{<\omega}$. *Then* $\llbracket \Phi \rrbracket^{\rightarrow}_{<\omega} = \llbracket \exists^{<\omega} U. \Psi^{\mathrm{rel}(U)} \rrbracket$.

Proof. "$\subseteq$": let $\mathfrak{B} \in \llbracket \Phi \rrbracket^{\rightarrow}_{<\omega}$. Then, by definition, there exists some finite $\mathfrak{A} \in \llbracket \Phi \rrbracket_{<\omega}$ and a homomorphism $h : \mathfrak{A} \to \mathfrak{B}$. Let $\mathfrak{C}$ be the finite substructure of $\mathfrak{B}$ induced by $h(A)$. Then h also serves as a homomorphism from $\mathfrak{A}$ to $\mathfrak{C}$ and therefore $\mathfrak{C} \in \llbracket \Phi \rrbracket^{\rightarrow <\omega}_{<\omega}$ and, by assumption, $\mathfrak{C} \in \llbracket \Psi \rrbracket_{<\omega}$. We now obtain $\mathfrak{D}$ from $\mathfrak{B}$ by adding U to its signature and letting $U^{\mathfrak{D}} = h(A)$. Then, obviously, $\mathfrak{C} = [\mathfrak{D}]_U$ and, thanks to $\mathfrak{C} \in \llbracket \Psi \rrbracket_{<\omega}$ we obtain $\mathfrak{D} \in \llbracket \Psi^{\mathrm{rel}(U)} \rrbracket$ by the remarks from Section III. By construction, $U^{\mathfrak{D}}$ is finite, therefore $\mathfrak{B} \in \llbracket \exists^{<\omega} U. \Psi^{\mathrm{rel}(U)} \rrbracket$ as desired.

"$\supseteq$": let $\mathfrak{B} \in \llbracket \exists^{<\omega} U. \Psi^{\mathrm{rel}(U)} \rrbracket$. Then there must be a structure $\mathfrak{D} \in \llbracket \Psi^{\mathrm{rel}(U)} \rrbracket$ expanding $\mathfrak{B}$ by interpreting an additional signature element U such that $U^{\mathfrak{D}}$ is finite. Let $\mathfrak{C} = [\mathfrak{D}]_U$, which is finite by construction and an induced substructure of $\mathfrak{B}$. Then by the properties of relativization follows $\mathfrak{C} \in \llbracket \Psi \rrbracket$ and due to finiteness of $\mathfrak{C}$ even $\mathfrak{C} \in \llbracket \Psi \rrbracket_{<\omega}$, therefore, by assumption $\mathfrak{C} \in \llbracket \Phi \rrbracket^{\rightarrow <\omega}_{<\omega}$, meaning there exists some $\mathfrak{A} \in \llbracket \Phi \rrbracket_{<\omega}$ and a homomorphism $h : \mathfrak{A} \to \mathfrak{C}$. But then $id_C \circ h$ is a homomorphism from $\mathfrak{A}$ to $\mathfrak{B}$, witnessing $\mathfrak{B} \in \llbracket \Phi \rrbracket^{\rightarrow}_{<\omega}$ as desired. $\square$

H. Proofs for Normal Form Section

Proposition 51. *For every homclosed* $\exists^n \forall^* \mathbb{FO}_=$ *sentence there exists an equivalent* $\exists^n \mathbb{FO}^+_=$ *sentence.*

Proof. Assume $\Phi \in \exists^n \forall^* \mathbb{FO}_=$ is over the signature $\tau \uplus \sigma$ with predicates τ and constants σ. Let σ^{sk} be a set of n Skolem constants. Let Φ^{sk} be the $\forall^* \mathbb{FO}_= \ \tau \uplus \sigma \uplus \sigma^{\mathrm{sk}}$-sentence obtained from Φ via skolemization of all the existentially quantified variables into (additional) constants. Clearly then $\llbracket \Phi \rrbracket_{(<\omega)} = \llbracket \Phi^{\mathrm{sk}} \rrbracket_{(<\omega)}|_\tau$. Since Φ^{sk} is a $\forall^* \mathbb{FO}_=$ sentence, any model $\mathfrak{A}$ gives rise to a small submodel by taking the substructure $\mathfrak{A}'$ induced by the set $A' \subseteq A$ containing all elements denoted

by constants. The identity on A' is a homomorphism from $\mathfrak{A}'$ to $\mathfrak{A}$. Let $\mathcal{A}$ be the set of small submodels thus obtained, for which therefore holds $\mathcal{A}^{\rightarrow} = \llbracket \Phi^{\mathrm{sk}} \rrbracket^{\rightarrow}$ and hence also $\mathcal{A}|_\tau^{\rightarrow} = \llbracket \Phi \rrbracket^{\rightarrow} = \llbracket \Phi \rrbracket$. Now let us obtain $\Psi = \exists \sigma^{\mathrm{sk}}. \bigvee_{\mathfrak{A} \in \mathcal{A}} \mathrm{cq}'(\mathfrak{A})$ where

$$\mathrm{cq}'(\mathfrak{A}) = \bigwedge_{\substack{c, c' \in \sigma \cup \sigma^{\mathrm{sk}} \\ \mathfrak{A} \models c = c'}} c = c' \quad \wedge \bigwedge_{\substack{P \in \tau, \ k = \mathrm{ar}(P) \\ \mathfrak{A} \models P(c_1, \ldots, c_k)}} P(c_1, \ldots, c_k). \quad (185)$$

Then we get $\llbracket \Psi \rrbracket = \mathcal{A}|_\tau^{\rightarrow} = \llbracket \Phi \rrbracket$. Moreover $\Psi \in \exists^n \mathbb{FO}^+_=$. $\square$

Lemma 52. *For any class* $\mathcal{C}$ *of (finite)* τ-*structures holds:*

1) $\mathcal{C}^{\hookrightarrow}_{(<\omega)}$ *is closed under (finite) superstructures. Moreover, if* $\mathcal{C}$ *is closed under (finite) superstructures, then* $\mathcal{C}^{\hookrightarrow}_{(<\omega)} = \mathcal{C}$.

2) $\mathcal{C}^{\mathrm{sH}(<\omega)}$ *is closed under (finite) surjective homomorphisms. Moreover, if* $\mathcal{C}$ *is closed under (finite) surjective homomorphisms, then* $\mathcal{C}^{\mathrm{sH}(<\omega)} = \mathcal{C}$.

3) $(\mathcal{C}^{\mathrm{sH}(<\omega)})^{\hookrightarrow}_{(<\omega)}$ *is (finitely) homclosed. Moreover, every (finitely) homclosed* $\mathcal{C}$ *satisfies* $(\mathcal{C}^{\mathrm{sH}(<\omega)})^{\hookrightarrow}_{(<\omega)} = \mathcal{C}$.

Proof. 1) and 2) follow immediately since "being superstructure" and "having surjective homomorphisms" are transitive relations on structures.

For 3) note that a class is closed under (finite) homomorphism if and only if it is closed under (finite) superstructures and (finite) surjective homomorphisms. By 1) we get that $(\mathcal{C}^{\mathrm{sH}(<\omega)})^{\hookrightarrow}_{(<\omega)}$ is closed under (finite) superstructures. Therefore it remains to show the closure under (finite) surjective homomorphisms.

Let $\mathfrak{A} \in (\mathcal{C}^{\mathrm{sH}(<\omega)})^{\hookrightarrow}_{(<\omega)}$ and suppose that there exists a (finite) surjective homomorphism $h \colon \mathfrak{A} \to \mathfrak{B}$ where $\mathfrak{B}$ is a τ-structure. By definition of $(\mathcal{C}^{\mathrm{sH}(<\omega)})^{\hookrightarrow}_{(<\omega)}$ there exists a substructure $\mathfrak{A}_0$ of $\mathfrak{A}$ such that $\mathfrak{A}_0 \in \mathcal{C}^{\mathrm{sH}(<\omega)}$. Let $\mathfrak{B}_0$ be the substructure of $\mathfrak{B}$ that is induced by $h(A_0)$. The structure $\mathfrak{A}_0$ has clearly a (finite) surjective homomorphism to $\mathfrak{B}_0$ and since $\mathcal{C}^{\mathrm{sH}(<\omega)}$ is by 2) closed under (finite) surjective homomorphisms it follows that $\mathfrak{B}_0 \in \mathcal{C}^{\mathrm{sH}(<\omega)}$. Since $\mathfrak{B}_0$ is a substructure of $\mathfrak{B}$ it follows by the definition of $(\mathcal{C}^{\mathrm{sH}(<\omega)})^{\hookrightarrow}_{(<\omega)}$ that $\mathfrak{B} \in (\mathcal{C}^{\mathrm{sH}(<\omega)})^{\hookrightarrow}_{(<\omega)}$ holds. $\square$

Proposition 54. *Let* Ψ *be an* $\mathbb{SO}$ τ-*sentence. Then*

$$\llbracket \Psi \rrbracket^{\hookrightarrow}_{(<\omega)} = \llbracket \Psi^{\mathrm{sup}} \rrbracket_{(<\omega)}. \quad (47)$$

Proof. Let $\mathfrak{A} \in \llbracket \Psi \rrbracket^{\hookrightarrow}_{(<\omega)}$ and let $\mathfrak{B}$ a substructure of $\mathfrak{A}$ such that $\mathfrak{B} \in \llbracket \Psi \rrbracket_{(<\omega)}$. Such a substructure exists by the definition of $\llbracket \Psi \rrbracket^{\hookrightarrow}_{(<\omega)}$. We can therefore choose $B \subseteq A$ as a witness for the set variable $\dot{U}$ in $\exists \dot{U}. \Psi^{\mathrm{rel}(\dot{U})}$ and get that $\mathfrak{A} \in \llbracket \Psi^{\mathrm{sup}} \rrbracket_{(<\omega)}$. For the other direction suppose that $\mathfrak{A} \in \llbracket \Psi^{\mathrm{sup}} \rrbracket_{(<\omega)}$ holds. This means by definition that there exists a subset B of A such that $\Psi^{\mathrm{rel}(B)}$ holds on $\mathfrak{A}$. Therefore the structure $\mathfrak{B}$ that is induced by $\mathfrak{A}$ on B satisfies Ψ and we get $\mathfrak{A} \in \llbracket \Psi \rrbracket^{\hookrightarrow}_{(<\omega)}$. $\square$

Proposition 57. *Let* Ψ *be an* $\mathbb{SO}$ τ-*sentence. Then*

$$\llbracket \Psi \rrbracket^{\mathrm{sH}(<\omega)}_{(<\omega)} = \llbracket \Psi^{\mathrm{shom}} \rrbracket_{(<\omega)}. \quad (51)$$

Proof. Let $\mathfrak{A} \in [\![\Psi]\!]_{(<\omega)}^{\mathrm{sH}_{(<\omega)}}$ and let $\mathfrak{A}'$ be a $\tau \uplus \tau' \uplus \{\mathrm{U}\}$-expansion of $\mathfrak{A}$ that satisfies $\Theta_{\mathrm{U}}^{\tau'} \wedge \bigwedge_{\mathrm{P}' \in \tau'} \forall \boldsymbol{x}.\big(\mathrm{P}(\boldsymbol{x}) \Rightarrow \mathrm{P}'(\boldsymbol{x})\big)$. Let $\mathfrak{B}_1$ be the substructure of $\mathfrak{A}'$ that is induced by $\mathrm{U}^{\mathfrak{A}'}$. Let $\mathfrak{B}_2$ be the τ'-reduct of $\mathfrak{B}_1$. We define $\mathfrak{B}$ as the τ-structure that arises from the signature substitution $\tau' \mapsto \tau$ in the structure $\mathfrak{B}_2$. We claim that there exists a surjective homomorphism $h \colon \mathfrak{A} \twoheadrightarrow \mathfrak{B}$ with the following definition: The map is the identity on $\mathrm{U}^{\mathfrak{A}'}$. Each element $x \notin \mathrm{U}^{\mathfrak{A}'}$ is mapped to some element $y \in \mathrm{U}^{\mathfrak{A}'}$ such that $\eta^{\tau'}(x, y)$ holds. Such a mapping exists (but is not unique) according to the sentence $\Theta_{\mathrm{U}}^{\tau'}$. Since $\mathrm{P}^{\mathfrak{A}'} \subseteq \mathrm{P}'^{\mathfrak{A}'}$ holds and by the formula $\eta^{\tau'}(x, y)$ it follows that h is a homomorphism. Note that h is surjective since it is the identity on $\mathrm{U}^{\mathfrak{A}'}$, which is the domain of $\mathfrak{B}$. By our assumption that $\mathfrak{A}$ is from $[\![\Psi]\!]_{(<\omega)}^{\mathrm{sH}_{(<\omega)}}$ it follows that $\mathfrak{B} \in [\![\Psi]\!]_{(<\omega)}$ holds. This is nothing else than saying that $\mathfrak{A} \models \Psi_{\tau \mapsto \tau'}^{\mathrm{rel}(\mathrm{U})}$ holds, which proves the first inclusion.

For the second inclusion, assume $\mathfrak{A} \in [\![\Psi^{\mathrm{shom}}]\!]_{(<\omega)}$ and that there exists a surjective homomorphism $h \colon \mathfrak{A} \twoheadrightarrow \mathfrak{B}$. Define the $\tau \uplus \tau' \uplus \{\mathrm{U}\}$-expansion $\mathfrak{A}'$ of $\mathfrak{A}$ as follows. For all $\mathrm{P} \in \tau$ we define the relation $\mathrm{P}'^{\mathfrak{A}'}$ as the preimage of $\mathrm{P}^{\mathfrak{A}}$ under h. The set $\mathrm{U}^{\mathfrak{A}'}$ is defined as some set of representatives of the kernel classes of h. It is clear that $\mathfrak{A}'$ satisfies $\Theta_{\mathrm{U}}^{\tau'} \wedge \bigwedge_{\mathrm{P}' \in \tau'} \forall \boldsymbol{x}.\big(\mathrm{P}(\boldsymbol{x}) \Rightarrow \mathrm{P}'(\boldsymbol{x})\big)$ and therefore it satisfies $\Psi_{\tau \mapsto \tau'}^{\mathrm{rel}(\mathrm{U})}$ as well. Let $\mathfrak{A}_1$ be the structure that is induced by $\mathfrak{A}'$ on $\mathrm{U}^{\mathfrak{A}'}$. By the definition of $\mathfrak{A}'$ the τ'-reduct of $\mathfrak{A}_1$ is isomorphic to $\mathfrak{B}$ after the signature substitution $\tau' \mapsto \tau$. This implies that $\mathfrak{B}$ satisfies Ψ. Since the surjective homomorphism h and the structure $\mathfrak{B}$ were arbitrary, we get $\mathfrak{A} \in [\![\Psi]\!]_{(<\omega)}^{\mathrm{sH}_{(<\omega)}}$. $\qquad\square$