Guarded negation

Vince Bárány[1]⋆, Balder ten Cate[2]⋆⋆ and Luc Segoufin[3]

[1] University of Warsaw
[2] UC Santa Cruz
[3] INRIA and ENS Cachan, LSV

Abstract. We consider restrictions of first-order logic and of fixpoint logic in which all occurrences of negation are required to be guarded by an atomic predicate. In terms of expressive power, the logics in question, called GNFO and GNFP, extend the guarded fragment of first-order logic and guarded least fixpoint logic, respectively. They also extend the recently introduced unary negation fragments of first-order logic and of least fixpoint logic.
We show that the satisfiability problem for GNFO and for GNFP is 2ExpTime-complete, both on arbitrary structures and on finite structures. We also study the complexity of the associated model checking problems. Finally, we show that GNFO and GNFP are not only computationally well behaved, but also model theoretically: we show that GNFO and GNFP have the tree-like model property and that GNFO has the finite model property, and we characterize the expressive power of GNFO in terms of invariance for an appropriate notion of bisimulation.

1 Introduction

Modal logic is well known for its "robust decidability": not only are basic decision problems such as satisfiability, validity and entailment decidable, but the decidability of these problems is preserved under various natural variations and extensions to the syntax and semantics of modal logic (e.g., addition of fixpoint operators, backward modalities, nominals; restriction to finite structures). As observed by Vardi [14], this robust decidability is intimately linked to the fact that modal logic has a combination of three properties, namely (i) the *tree model property* (if a formula has a model, it has a model which is a tree), (ii) *translatability into monadic second-order logic* (MSO), and thereby into tree automata and, (iii) the *finite model property* (every satisfiable modal formula is satisfied in a finite structure). The decidability of satisfiability (on arbitrary structures and on finite structures) follows immediately from these three properties. However, we should note here that the two way μ-calculus (the extension of modal logic with fixpoint operators and backward modalities) lacks the finite model property, and hence the decidability of satisfiability on finite structures for this logic involves a separate (non trivial) argument [5].

The properties (i), (ii) and (iii) described above can be viewed as a semantic explanation for the robust decidability of modal logic. Given that modal logic can be viewed

⋆ First author was supported by ERC Starting Grant Sosna.
⋆⋆ The second author was supported by NSF grant IIS-0905276.

as a syntactic fragment of first-order logic, it is also natural to ask for syntactic explanations: *what syntactic features of modal formulas (viewed as first-order formulas) are responsible for their good behavior? And can we generalize modal logic, preserving these features, while at the same time dropping inessential restrictions inherent in modal logic (such as the fact that it can only describe structures with unary and binary relations)?*

Several answers to these questions have been proposed. The first one is to consider the two variable fragment of first-order logic, which is decidable and has the finite model property [12]. Unfortunately, this observation does not go very far towards explaining the robust decidability of modal logic, since it seems impossible to extend the two variable fragment with a fixpoint mechanism while maintaining decidability [9].

The second proposal is to consider logics with guarded quantifications. The *guarded fragment* of first-order logic (GFO) consists of FO formulas in which all quantifiers are "guarded" by atomic predicates. It was introduced in [1]. It has a natural extension with fixpoint operators (GFP) that extends the two-way μ-calculus [10]. Both GFO and GFP have the tree-like model property (if a formula has a model, it has one of bounded tree width), they can be interpreted into MSO (each formula can be transformed into a tree automaton recognizing tree decompositions of its models of bounded tree width) and GFO has the finite model property [1, 8]. Finite satisfiability of GFP was only recently proved decidable in [2].

The third, and most recent proposal is based on unary negation. Unary negation first-order logic (UNFO) restricts first-order logic by constraining the use of negation to subformulas having at most one free variable (and viewing universal quantification as a defined connective). Unary negation fixpoint (UNFP) is the natural extension of UNFO using monadic fixpoints. Again, UNFO generalizes modal logic, and UNFP generalizes the two-way μ-calculus. Both UNFO and UNFP have the tree-like model property, they can be interpreted into MSO and UNFO has the finite model property [13]. Decidability of finite satisfiability for UNFP was also established in [13].

The three extensions of modal logics presented above are incomparable in terms of expressive power. In particular there are properties expressible in UNFO that are not expressible in GFO and vice-versa. In this paper we unify the unary negation and guarded quantification approaches by introducing guarded negation logics.

Guarded negation first-order logic (GNFO) restricts FO by requiring that all occurrences of negation are of the form $\alpha \wedge \neg\phi$ where the "guard" α is an atomic formula (possibly an equality statement) containing all the free variables of ϕ. We also disallow universal quantification as a primitive connective (though a limited form of universal quantification can be expressed using existential quantification and guarded negation). For instance, GNFO cannot express $x \neq y$ but it can express $R(x, y, z) \wedge x \neq y$. Guarded negation fixpoint (GNFP) extends GNFO with a guarded fixpoint mechanism. In terms of expressive power, GNFO forms a strict extension of both UNFO and GFO.

We show that our guarded negation logics have the same desirable properties as modal logics, unary negation logics and guarded logics. In particular, the satisfiability problem for GNFO and GNFP is decidable, both on arbitrary structures and on finite structures. These problems are all 2ExpTime-complete, even for a fixed finite schema (recall that satisfiability of GFO is in ExpTime when the schema is fixed). We also study the (combined) complexity of the model checking problem of GNFO and GNFP. The problem is $\mathrm{P}^{\mathrm{NP}[O(\log^2 n)]}$-complete for GNFO. In the case of GNFP, it is hard for P^{NP} and contained in $\mathrm{NP}^{\mathrm{NP}} \cap \mathrm{coNP}^{\mathrm{NP}}$. Note that a similar gap between the upper bound and the lower bound exists for GFP and the μ-calculus, where the complexity of model checking is known to lie between PTime and NP $\cap$ coNP [4]. Recall that the model checking problem of GFO is PTime-complete [4]. Our proofs are based on reductions to the model checking problem for UNFO and UNFP. Finally, we show that GNFO and GNFP have the tree-like model property, and that GNFO has the finite model property, and we characterize the expressive power of GNFO in terms of invariance for an appropriate notion of bisimulation.

The most difficult result is the decidability of satisfiability on finite structures. For GNFO, we give a reduction to testing whether a union of conjunctive queries is implied by a guarded formula, recently shown decidable in [3]. In the case of GNFP, we make a reduction to the decidability of finite satisfiability of GFP, recently proved in [2].

Related work GNFO and GNFP form decidable extensions of GFO and GFP. Other decidable extensions of GNFO and GNFP have been considered in the past, most notably the *clique-guarded* fragment (and the related *packed* fragment, as well as the weaker *loosely-guarded* fragment) of first-order logic, and of least fixpoint logic [6]. The logics GNFO and GNFP we propose here are incomparable in expressive power to the clique guarded fragments. We leave open the question whether a decidable common generalization exists.

2 Preliminaries

Structures and formulas. We are working on relational structures. We assume given a relational schema τ consisting of a finite set of relation symbols, each having an associated arity. By the *arity of a schema*, we mean the maximal arity of its relations. A *structure* (or *model*) M over a relational schema τ consists of a set $dom(M)$, the *domain* of M, together with an interpretation of each relation symbol $R \in \tau$ as a k-ary relation over $dom(M)$ for k the arity of R according to τ. A structure M is said to be *finite* if $dom(M)$ is finite. If a tuple of elements $\bar{a}$ from $dom(M)$ belongs to the interpretation of a relation symbol R, then we say that $R(\bar{a})$ is a *fact* of M. A tuple (or set) of elements of M is *guarded* if it is a singleton or all its components (elements) occur among those in a fact of M.

We assume familiarity with first-order logic, FO, and least fixpoint logic, LFP, over relational structures. We use classical syntax and semantics for FO and LFP. We write $\phi(\bar{x})$ to denote the fact that the free variables of ϕ are exactly the variables in $\bar{x}$. We also write $M \models \phi(\bar{u})$ or $M, \bar{u} \models \phi(\bar{x})$ for the fact that the tuple $\bar{u}$ of elements of the model M makes the formula $\phi(\bar{x})$ true in M. The *size of a formula* ϕ, denoted by $|\phi|$, is the number of symbols needed to write down the formula.

Conjunctive queries. A conjunctive query (CQ) is a first-order formula of the form $\exists\bar{x}\alpha$ where α is a conjunction of positive atomic formulas (including equalities). A union of conjunctive queries (UCQ) is a disjunction of CQs. A positive-existential query is an FO formula built using disjunction, conjunction and existential quantification only. Every positive-existential query can be transformed in a UCQ at the cost of a possible exponential blow-up. Positive-existential queries belong to GNFO, even to UNFO. The width of a CQ is the number of variables occurring in it, and the width of a UCQ is the maximum width of its CQs. The height of a UCQ is the maximum size of its CQs.

GNFO. We define GNFO, *guarded negation* FO, as the fragment of FO given by the following grammar, where R ranges over predicate symbols, and $\alpha(\bar{x}\bar{y})$ is an atomic formula (possibly an equality statement).

$$\varphi ::= \quad R(\bar{x}) \mid x = y \mid \exists x\varphi \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \alpha(\bar{x}\bar{y}) \wedge \neg\varphi(\bar{y}) \qquad (1)$$

Hence the logic can only negate a subformula if all its free variables are "guarded" by some fact, or if the subformula has at most one free variable (in which case one can use an equality statement of the form $x = x$ or $y = y$ as the guard). For example, $x \neq y$ is not a formula of GNFO but $R(x, y, z) \wedge x \neq y$ is.

We say that a formula of GNFO is in *GN-normal form* if, in its syntax tree, no disjunction is directly below an existential quantifier or a conjunction, and no existential quantifier is directly below a conjunction sign. Every GNFO formula can be brought into GN-normal form, at the cost of an exponential increase in length and linear increase in the number of variables, using the following equivalences as rewrite rules (where x' is a variable not occurring in ψ):

$$\exists x(\phi \vee \psi) \simeq \exists x\phi \vee \exists x\psi,\ \phi \wedge (\psi \vee \chi) \simeq (\phi \wedge \psi) \vee (\phi \wedge \chi),\ (\exists x\phi) \wedge \psi \simeq \exists x'(\phi[x'/x] \wedge \psi)$$

The appeal of the GN-normal form is that it highlights the fact that GNFO formulas can be naturally viewed as being built up from atomic formulas using guarded negation, and unions of conjunctive queries. Indeed, the GNFO formulas in GN-normal form are precisely generated by the following recursive definition:

$$\varphi ::= \quad R(\bar{x}) \mid x = y \mid \alpha(\bar{x}\bar{y}) \wedge \neg\varphi(\bar{y}) \mid q[\varphi_1/U_1, \dots, \varphi_s/U_s] \qquad (2)$$

where q is a UCQ using relation symbols $U_1, \dots, U_s$, and $\varphi_1, \dots, \varphi_s$ are formulas (generated by the same recursive definition) with the appropriate number of free variables corresponding to the relation symbols they replace. Here, $q[\varphi_1/U_1, \dots, \varphi_s/U_s]$ is the result of replacing in q all subformulas of the form $U_i(\bar{x})$ with $i \leq s$ by $\phi_i(\bar{x})$.

A formula of GNFO is said to be *of width* k if, when brought into GN-normal form in the way described above, it uses at most k variables (or equivalently, is built up using UCQs q of width at most k). We denote by GNFO^k all GNFO formulas of width k.

GNFO extends GFO and UNFO. GNFO generalizes the logic UNFO, studied in [13], which only allows the negation of formulas having at most one free variable. It also generalizes the *guarded fragment of first-order logic* (GFO). The logic GFO is the fragment of FO defined by the following grammar, where, again, $\alpha(\bar{x}\bar{y}\bar{z})$ is an atomic formula (possibly an equality statement):

$$\varphi ::= R(\bar{x}) \mid x = y \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \neg\varphi \mid \exists\bar{x}\ \alpha(\bar{x}\bar{y}\bar{z}) \wedge \varphi(\bar{x}\bar{y}) \mid \forall\bar{x}\ \alpha(\bar{x}\bar{y}\bar{z}) \rightarrow \varphi(\bar{x}\bar{y})$$

It is straightforward to check that:

Proposition 1. *Every* GFO *sentence is equivalent to a* GNFO *sentence, via a polynomial time transformation.*[4]

Proof. Let φ be any GFO sentence. We may assume that φ does not contain universal quantifiers, by using negation and (guarded) existential quantification instead. Since φ is a sentence, every subformula $\vartheta(\bar{x})$ is in the scope of an (inner-most) guarded existential quantifier $\exists\bar{x}\bar{u}(\alpha_\vartheta(\bar{x}\bar{u})\wedge\cdots)$. We replace in φ each negated subformula $\vartheta(\bar{x}) = \neg\psi(\bar{x})$ by $\alpha_\vartheta(\bar{x}\bar{u}) \wedge \neg\psi(\bar{x})$ to obtain the desired equivalent GNFO-sentence. □

Example 1. The GNFO sentence $\delta = \exists xy\big(E(x,y) \wedge \neg\exists uvw(E(x,u) \wedge E(u,v) \wedge E(v,w) \wedge E(w,y))\big)$ is not equivalent to any GFO sentence or to any UNFO sentence, even on undirected graphs. This is because δ defines a property that is not invariant under guarded bisimulation (which, incidentally, amounts to ordinary bisimulation in case of undirected graphs), as can be easily verified, nor is it invariant under "UN-bisimulation" as befits UNFO formulas, cf. [13].

3 The satisfiability problem for GNFO

We show in this section how to reduce the (finite) satisfiability problem for GNFO to the problem of testing whether a GFO formula entails (on finite structures) a UCQ. The latter problem is also known as the problem of query answering against a GFO theory, and it has been solved in [3]. To streamline the presentation, we will allow the possibility of zero-ary relation symbols.

Lemma 1. *To every* $\varphi(\bar{x}) \in \mathsf{GNFO}[\tau]$ *one can associate in polynomial time a companion formula* $\psi(\bar{x}) \in \mathsf{GNFO}[\tau \uplus \sigma]$ *of the form*

$$\psi(\bar{x}) = \underbrace{S(\bar{x}) \;\wedge\; \bigwedge_j \forall\,\bar{z}\bar{u}.R_j(\bar{z}\bar{u}) \rightarrow q_j(\bar{z})}_{\psi^+} \;\;\wedge\;\; \underbrace{\bigwedge_i \forall\,\bar{z}\bar{u}.T_i(\bar{z}\bar{u}) \rightarrow \neg p_i(\bar{z})}_{\psi^-} \tag{3}$$

where σ *comprises the new relation symbols occurring as* S, R_j *or* T_i, *where the* q_j*'s and* p_i*'s are positive-existential,* $\mathrm{width}(\psi) = \mathrm{width}(\varphi)$ *and such that* $\varphi \leftrightarrow \exists S\,\exists\overline{T}\,\psi$.

Proof. Given a GNFO-formula φ consider an inner-most occurrence of a guarded negation $R(\bar{z}\bar{u}) \wedge \neg q(\bar{z})$ as a subformula of φ. Then $q(\bar{z})$ is necessarily positive existential. Let T be a new predicate symbol of the same arity as R. We substitute $T(\bar{z}\bar{u})$ in the input formula for the subformula $R(\bar{z}\bar{u}) \wedge \neg q(\bar{z})$, and add the following as conjuncts to ψ^+ and ψ^-, according to their kind.

$$\begin{array}{l}\forall\,\bar{z}\bar{u}.T(\bar{z}\bar{u}) \rightarrow \neg q(\bar{z})\\ \forall\,\bar{z}\bar{u}.T(\bar{z}\bar{u}) \rightarrow R(\bar{z}\bar{u})\\ \forall\,\bar{z}\bar{u}.R(\bar{z}\bar{u}) \rightarrow T(\bar{z}\bar{u}) \vee q(\bar{z})\end{array}$$

Inner-most *equality-guarded* negations $z = u \wedge \neg q(z,u)$ are handled in a similar fashion. Again, $q(z,u)$ must be positive-existential. We choose a new unary relation symbol

[4] This is only true for *sentences*, as $\neg R(xy)$ is in GFO but not expressible in GNFO.

T, replace the subformula in question by $z = u \wedge T(z)$, and add $\forall\, z.T(z) \rightarrow \neg q[u/z]$ and $\forall\, z.T(z) \vee q[u/z]$ as conjuncts to the normal form.

Proceeding in this manner from the inside-out we eliminate all guarded negations until the original input formula is reduced to a single positive-existential formula $p(\bar{x})$ (in the extended signature). Finally we replace $p(\bar{x})$ with $S(\bar{x})$ where S is an appropriate new predicate symbol and add $\forall\, \bar{x}.S(\bar{x}) \rightarrow p(\bar{x})$ as conjunct to the normal form, which is thus finalized. It is now easy to verify the correctness of this transformation. $\square$

We may assume wlog. that the positive-existential formulas q_j of (3) are in prenex normal form, i.e. $q_j(\bar{z}) = \exists \bar{u} \xi_j(\bar{z}, \bar{v})$. Also note that each conjunct $\forall\, \bar{z}\bar{u}.T_i(\bar{z}\bar{u}) \rightarrow \neg p_i(\bar{z})$ of (3) is the negation of a positive-existential sentence $\exists\, \bar{z}\bar{u}.T_i(\bar{z}\bar{u}) \wedge p_i(\bar{z})$. Therefore, the entire ψ^- of (3) can be conceived as the negation of a single positive-existential sentence q. This leads us to the following equivalent formula.

$$\underbrace{S(\bar{x}) \wedge \bigwedge_j \Big(\forall\, \bar{z}\bar{u}.R_j(\bar{z}\bar{u}) \rightarrow \exists \bar{v} \xi_j(\bar{z}\bar{v})\Big)}_{\psi^+} \;\wedge\; \neg \underbrace{\bigvee_i \Big(\exists\, \bar{z}\bar{u}.T_i(\bar{z}\bar{u}) \wedge p_i(\bar{z})\Big)}_{q} \qquad (4)$$

Observe next that without affecting satisfiability of (4) we may introduce new atoms guarding the existential quantifiers in ψ^+ thus obtaining a GFO-formula

$$\psi^* = S(\bar{x}) \wedge \bigwedge_j \Big(\forall\, \bar{z}\bar{u}.R_j(\bar{z}\bar{u}) \rightarrow \exists \bar{v} Q_j(\bar{z}\bar{v}) \wedge \xi_j(\bar{z}\bar{v})\Big)$$

where the Q_j's are distinct new relation symbols of appropriate arity. Then, $\psi^* \models \psi^+$ and, conversely, every model of ψ^+ has an expansion modeling ψ^*.

The entire transformation of an input GNFO-formula φ to the equi-satisfiable $\psi^* \wedge \neg q$, with ψ^* in GFO and q positive existential, can be performed in polynomial time and only results in a polynomial blowup in the signature of the latter normal form. In a final transformation step, which may require at most exponential time, the positive-existential sentence q can be converted to an equivalent Boolean UCQ q^*. In general q^* may be comprised of exponentially many CQs each of size at most $|q|$. Summing up all the reduction steps we obtain:

Proposition 2. *For each* $\varphi(\bar{x}) \in \mathsf{GNFO}[\tau]$ *one can compute in exponential time a* GFO*-formula* $\psi^*(\bar{x})$ *and* UCQ $q^* = \bigvee_l Q_l$*, both of signature* $\tau \uplus \{\bar{T}\}$*, such that*

$$\varphi \;\longleftrightarrow\; \exists \bar{T}\, (\, \psi^* \wedge \neg\, q^*\,) \qquad (5)$$

is valid, and that $|\psi^*| = \mathcal{O}(|\varphi|)$ *and* $\mathrm{height}(q^*) = \max_l |Q_l| \leq |\varphi|$.

We now summarize the main results of [3]. Later we will build on key elements of the construction of [3], stated below as Lemmas 2 and 3 and Theorem 4, from which the following Theorem 1 can be directly derived.

Theorem 1 ([3]). *Given a* GFO*-formula* ψ *and a* UCQ q *of height* h *it is decidable in time* $|q| \cdot 2^{(h|\psi|)^{\mathcal{O}(h|\psi|)}}$ *whether or not* $\psi \wedge \neg q$ *is satisfiable; and if* $\psi \wedge \neg q$ *has a model then it has a finite model of size* $2^{(h|\psi|)^{\mathcal{O}(h|\psi|)}}$

By combining Theorem 1 with the estimates of Proposition 2 we derive the complexity of satisfiability for GNFO, as well as its finite model property.

Theorem 2. *1. The satisfiability problem for* GNFO *is 2*EXPTIME*-complete.*

2. Every satisfiable GNFO*-sentence* φ *has a finite model of size* $2^{2^{|\varphi|^{\mathcal{O}(1)}}}$.

The 2EXPTIME lower bound follows immediately from the fact that satisfiability for UNFO is already hard for 2EXPTIME [13, 7]. It holds even if the schema is fixed (recall that when the schema is fixed the complexity of satisfiability for GFO is ExpTime-complete).

4 The satisfiability problem for GNFP

In a nutshell, GNFP is the extension of GNFO with guarded fixpoints. We show here that both satisfiability and finite satisfiability are decidable for GNFP.

GNFP. In order to define GNFP we introduce extra predicate variables, which will serve for computing fixpoints. We denote the predicates given by the relational schema by P, Q, R, S etc. and the predicate variables serving for computing the fixpoints by X, Y, Z etc. However the fixpoint predicates are not permitted to be used as guards. For instance $R(\bar{x}) \wedge \neg Y(\bar{x})$ is allowed but $Y(\bar{x}) \wedge \neg R(\bar{x})$ is not. Formulas of GNFP$[\tau]$, we omit the schema τ when it is clear from the context, pertain to the following syntax where R is any relational symbol in τ, $\alpha(\bar{x}\bar{y})$ is an atomic formula (possibly an equality statement), and $\sigma \subseteq \tau$.

$$\begin{aligned}\phi ::= {} & R(\bar{x}) \mid \mathsf{x{=}y} \mid X(\bar{x}) \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2 \mid \exists x\, \phi \mid \alpha(\bar{x}\bar{y}) \wedge \neg\phi(\bar{x}) \mid \\ & \mu_{Z,\bar{z}}[\, \mathtt{guarded}_\sigma(\bar{z}) \wedge \phi(\bar{Y}, Z, \bar{z})\,](\bar{x}) \quad \mid \quad \nu_{Z,\bar{z}}[\, \mathtt{guarded}_\sigma(\bar{z}) \wedge \phi(\bar{Y}, Z, \bar{z})\,](\bar{x})\end{aligned}$$

Here μ and ν stand for least- and greatest fixpoints, respectively, and it is further required that in the matrix $\phi(\bar{Y}, Z, \bar{z})$ of a fixpoint formula i) the fixpoint variable Z occurs only positively (i.e. always under an even number of negations) and never as a guard; ii) no first-order parameters (i.e., free variables other than those $\bar{z}$ bound by the fixpoint operator) are permitted; and iii) free fixpoint variables $\bar{Y}$ are allowed, to enable nesting of fixpoint declarations. The clause $\mathtt{guarded}_\sigma(\bar{z})$ signifies the requirement that all tuples belonging to a predicate defined by a fixpoint construct must be guarded by a relational atom of the underlying structure. The clause $\mathtt{guarded}_\sigma(\bar{z})$ can be understood in either of two ways: as a syntactic element (keyword) signifying this intended semantics, or as a formula defining guardedness by a disjunction of existentially quantified relational atoms (allowing relation symbols from σ as well as equality) involving all of the variables $\bar{z}$. Note that while the definitions of least- and greatest fixpoint formulas are symmetric, the two operators are *not* each others duals. This is due to the fact that the dualization of ν may introduce negations that are not guarded.

Since GNFP can be seen as a syntactic fragment of least fixed point logic LFP, we omit the definition of the semantics, cf. [11].

The definition of GN-normal form that we gave for GNFO formulas applies to GNFP as well. Formulas of GNFP in GN-normal form can be naturally thought of as being built up from atomic formulas using (i) guarded negation, (ii) unions of conjunctive queries, and (iii) fixpoint operators. As in the case of GNFO, the *width* of a GNFP-formula is the number of variables it contains after being put in GN-normal form and we let GNFP^k denote the set of GNFP-formulas of width k.

GNFP extends GFP and UNFP. Syntactically, GNFP generalizes the logic UNFP studied in [13], which only allows the negation of formulas having at most one free variable, and only unary fixpoints. GNFP also generalizes the *guarded fragment of fixpoint logic* (GFP) [10]. The logic GFP is the fragment of LFP defined as for GNFP but replacing the first-order part of the syntax based on GNFO by the syntax of GFO. It is not immediate from the syntax, but it is easy to check by induction building on Proposition 1 that GNFP generalizes GFP.

Proposition 3. *Every sentence of* GFP *is equivalent to a sentence of* GNFP*, via a polynomial time transformation.*

The aim of this section is to establish the following main result.

Theorem 3. *It is decidable whether a sentence of* GNFP *has a model and whether it has a finite model. Both of these problems are* 2EXPTIME*-complete.*

The proof of Theorem 3 is a reduction to the (finite) satisfiability of GFP: given a formula of GNFP we construct a formula of GFP whose (finite) satisfiability is equivalent to the one of the initial formula and we then apply known results on (finite) satisfiability for GFP [10, 2]. Before we describe the reduction, we start with some useful notation and some preliminary results taken from [3].

Acyclicity and treeification. We say that a conjunctive query is *acyclic* if it is semantically equivalent to a formula of GFO built with only conjunction and existential quantification. For instance the query $\exists yzw\, T(x,y,z) \wedge T(x,w,z) \wedge E(x,y)$ is acyclic because it is equivalent to the guarded formula $\exists yz T(x,y,z) \wedge E(x,y) \wedge (\exists w T(x,w,z))$. It is easy to check that this definition is equivalent to acyclicity (in the hypergraph sense) of the hypergraph induced by the atoms of the query with its variables as vertices.

Definition 1. *Given a schema τ, the τ-*treeification $\Lambda_q^\tau(\bar{x})$ *of a positive existential query $q(\bar{x})$ over τ is the* UCQ *consisting of the disjunction of all those* acyclic *CQs over τ (modulo renaming of bound variables) that imply q and that are minimal (in the sense that removing any atomic formula would render it non-acyclic or not implying q).*

Consider for instance $q(x) = \exists yzw\, E(x,y) \wedge E(y,z) \wedge E(z,w) \wedge E(w,x)$. Then $\Lambda_q^{\{E\}} = E(x,x) \vee \exists y\, E(x,y) \wedge E(y,x)$. Indeed, the only acyclic queries implying $q(x)$ are obtained by identifying some of its variables resulting in either a reflexive edge on x or a pair of inverse edges. If the schema is $\{E, T\}$ where T is a ternary predicate the treeification of $q(x)$ has a number of additional disjuncts corresponding to various triangulations of $q(x)$, such as $\exists yzw\, T(x,y,z) \wedge T(x,w,z) \wedge E(x,y) \wedge E(y,z) \wedge E(z,w) \wedge E(w,x)$. It can be shown that each disjunct in the treeification of any CQ in whatever schema contains at most three times as many atoms as the CQ itself [3] leading to the following observations.[5]

[5] These figures constitute a slight refinement of those offered in [3, Lemma 10].

Lemma 2. *Consider a schema τ having r many predicate symbols of maximal arity w. Let $q(\bar{x})$ be a UCQ of height h over τ. Then $\Lambda_q^\tau(\bar{x})$ has width w, length $r^{\mathcal{O}(h)}(hw)^{\mathcal{O}(hw)}$, height $\mathcal{O}(h)$, and can be constructed in time $|q|r^{\mathcal{O}(h)}(hw)^{\mathcal{O}(hw)}$.*

Guarded bisimulation and covers. Guarded bisimulations [6] are, roughly speaking, strategies for Duplicator in Ehrenfeucht-Fraïssé games played on structures M and N, with positions restricted to $(\bar{a}, \bar{b})$ such that $\bar{a}$ is guarded in M and $\bar{b}$ guarded in N (cf. also Definition 3 below). Guarded bisimilarity implies GFP-indistinguishability [6]. There is an associated notion of guarded unraveling, which, given a structure M, provides an acyclic structure M^* that is guarded bisimilar to M, thus exhibiting the tree-like model property of GFP [6]. Note that M^* is in general infinite and, because it is acyclic, over M^* every conjunctive query is equivalent to its treeification (indeed, this is one justification for the name "treeification"). An analogue of guarded unraveling for finite structures was provided in [3] in the form of a construction, for all n, of a finite companion $M^{(n)}$ of any finite structure M such that $M^{(n)}$ is guarded bisimilar to M and $M^{(n)} \models q \leftrightarrow \Lambda_q^\tau$ for all UCQs q of width at most n.

Definition 2. A guarded bisimilar cover $\pi\colon N \xrightarrow{\sim} M$ *is an onto homomorphism* $\pi\colon N \to M$ *inducing a guarded bisimulation* $\{(\bar{b}, \pi(\bar{b})) \mid \bar{b}$ *guarded in* $N\}$. *The cover is* weakly n-acyclic *if for every homomorphism $h : Q \to N$ from a structure Q on at most n elements into N there exists an acyclic substructure M_0 of M (not necessarily induced) with $\pi(h(Q)) \subseteq M_0$.*

The following lemma shows how covers relate to treeification.

Lemma 3. *Let $\pi : N \xrightarrow{\sim} M$ be a weakly n-acyclic guarded bisimilar cover of τ-structures, and let $q(\bar{x})$ be a* UCQ *of width at most n. Then for every* guarded tuple $\bar{b}$ *(of not necessarily distinct elements) in N we have $N \models q(\bar{b}) \leftrightarrow \Lambda_q^\tau(\bar{b})$.*

Theorem 4 (Rosati cover [3]). *For all $n \in \mathbb{N}$ every relational structure M of schema τ admits a weakly n-acyclic guarded bisimilar cover $\pi\colon M^{(n)} \xrightarrow{\sim} M$. If M is finite then $|M^{(n)}| = |M|^{w^{\mathcal{O}(n)}}$, where w is the arity of τ, and $M^{(n)}$ can be effectively constructed. We call $M^{(n)}$ the n-*th Rosati cover *of M.*

Say that a relation $Z \subseteq M^r$ is *guarded* in M if every tuple $\bar{a} \in Z$ is guarded in M. The following is an immediate consequence of the definitions.

Fact 1 *Consider a weakly n-acyclic cover $\pi : N \xrightarrow{\sim} M$ and guarded predicates $Z_1, \ldots, Z_k$ over M. Then $\pi : (N, W_1, \ldots, W_k) \xrightarrow{\sim} (M, Z_1, \ldots, Z_k)$ is a weakly n-acyclic cover, where $W_i = \pi^{-1}(Z_i) = \bigcup\{\pi^{-1}(\bar{a}) \mid \bar{a} \in Z_i\}$ for each $1 \leq i \leq k$.*

Reduction to (finite) satisfiability for GFP. Let φ be any given GNFP sentence. As a first step, we compute its GN-normal form $\tilde{\varphi}$. Note that $\tilde{\varphi}$ has the following dimensions: $|\tilde{\varphi}| = 2^{\mathcal{O}(|\varphi|)}$, $\mathrm{width}(\tilde{\varphi}) = \mathcal{O}(|\varphi|)$, and $\tilde{\varphi}$ is built up using only UCQs of height at most $|\varphi|$ (as well as guarded negations and fixpoint operators) as in (2).

Next, essentially, our reduction transforms all UCQs occurring in $\tilde{\varphi}$ to their treeification. For every $k \geq 1$, and for every relational schema τ consisting of at most k-ary relations, we define a translation η from GNFP$^k[\tau]$ sentences in GN-normal form to GFP$^k[\tau \uplus \{C_k\}]$ sentences, where C_k is a new symbol of arity k, by structural recursion, using the following rules.

$$\eta(\, R(\bar{x})\,) = R(\bar{x}) \tag{a}$$
$$\eta(\, Z(\bar{x})\,) = Z(\bar{x}) \tag{b}$$
$$\eta(\, \alpha(\bar{x}\bar{y}) \wedge \neg\psi(\bar{x})\,) = \alpha(\bar{x}\bar{y}) \wedge \neg\, \eta(\psi(\bar{x})) \tag{c}$$
$$\eta(\, \mu_{Z,\bar{z}}[\,\mathtt{guarded}_\tau(\bar{z}) \wedge \psi(\bar{Y}, Z, \bar{z})\,]\,) = \mu_{Z,\bar{z}}[\,\mathtt{guarded}_\tau(\bar{z}) \wedge \eta(\psi(\bar{Y}, Z, \bar{z}))\,] \tag{d}$$
$$\eta(\, \nu_{Z,\bar{z}}[\,\mathtt{guarded}_\tau(\bar{z}) \wedge \psi(\bar{Y}, Z, \bar{z})\,]\,) = \nu_{Z,\bar{z}}[\,\mathtt{guarded}_\tau(\bar{z}) \wedge \eta(\psi(\bar{Y}, Z, \bar{z}))\,] \tag{d'}$$
$$\eta(\, q[\phi_1/U_1, \ldots, \phi_s/U_s]\,) = \Lambda_q^{\tau \uplus \{U_1,\ldots,U_s,C_k\}}[\eta(\phi_1)/U_1, \ldots, \eta(\phi_s)/U_s] \tag{e}$$

where in (e) q is a UCQ of signature $\{U_1, \ldots, U_s\}$ disjoint from $\tau \uplus \{\bar{Y}, C_k\}$ and $\phi_1, \ldots, \phi_s \in$ GNFP$^k[\tau \uplus \{\bar{Y}\}]$, where $\bar{Y}$ enumerates the free fixpoint variables occurring in any of the ϕ_i's, each of which is of a form (a)–(d), and such that $q[\phi_1/U_1, \ldots, \phi_s/U_s]$ is a subformula of $\tilde{\varphi}$. In particular the ϕ_i define guarded relations. It can be readily seen that all formulas in GN-normal form can be decomposed as in (a)–(e) and we have the following bounds and on the translation η.

Lemma 4. *For all* GNFP^k*-formula φ of GN-normal form $\tilde{\varphi}$, $|\eta(\tilde{\varphi})| = 2^{(k|\varphi|)^{\mathcal{O}(1)}}$ and $\eta(\tilde{\varphi})$ can be computed within this time bound, and its width remains k.*

The following key lemma generalizes Lemma 3 and attests to the correctness of our reduction. It is proved by structural induction on formulas, while relying on Fact 1 and Lemma 3 to deal with the cases (d) and (e), respectively.

Lemma 5. *Let $\pi : N \xrightarrow{\sim} M$ be a weakly k-acyclic guarded bisimilar cover of $\tau \uplus \{C_k\}$-structures and $\phi(\bar{Y}, \bar{x}) \in$ GNFP$^k[\tau]$ a formula in GN-normal form with free fixpoint variables $\bar{Y}$. Then for every assignment of guarded relations $\bar{W}$ on N to $\bar{Y}$ such that $W_i = \pi^{-1}(\pi(W_i))$ for all $1 \leq i \leq |\bar{Y}|$ and for every guarded tuple $\bar{b}$ in N we have:*

$$(N, \bar{W}) \models \eta(\phi)(\bar{b}) \leftrightarrow \phi(\bar{b})\ .$$

Theorem 5. *A* GNFP^k*-sentence $\tilde{\varphi}$ in GN-normal form is satisfiable (in the finite) if, and only if, $\eta(\tilde{\varphi}) \in$* GFP^k *is satisfiable (in the finite).*

Proof. It is easy to see that for every model M of $\tilde{\varphi}$ its expansion (M, C_k) is a model of $\eta(\tilde{\varphi})$, where each C_k is the complete k-ary relation on M.

Conversely, consider some M a model of $\eta(\tilde{\varphi})$ and its k-th Rosati cover $M^{(k)}$, equally a model of $\eta(\tilde{\varphi})$. Lemma 5 proves that $M^{(k)}$ is, in fact, a model of $\tilde{\varphi}$, and we know from Theorem 4 that if M is finite then so is $M^{(k)}$. □

Both satisfiability [10] and finite satisfiability [2] of GFP sentences have been shown decidable in time $2^{\mathcal{O}(nw^w)}$, where n is the length of the input formula and w is its width. Starting with a GNFP^k sentence φ whose GN-normal form is $\tilde{\varphi}$, we get from Lemma 4 that $|\eta(\tilde{\varphi})| = 2^{(k|\varphi|)^{\mathcal{O}(1)}}$ and that $\eta(\tilde{\varphi})$ is computable within that same time bound, but its width remains k. Theorem 3 now follows from these bounds via Theorem 5.

5 Additional results

Model checking In this section we study the combined complexity of the model checking problem, where the input consists of a sentence and a structure and the goal is to decide whether the sentence is true on the structure. It was shown in [13] that the model checking problem for UNFO is $\mathrm{P}^{\mathrm{NP}[O(\log^2 n)]}$-complete, and that the model checking problem for UNFP is in $\mathrm{NP}^{\mathrm{NP}} \cap \mathrm{coNP}^{\mathrm{NP}}$ and P^{NP}-hard. We show that the upper-bounds also apply to GNFO and GNFP. The proof is a reduction to formulas with unary negations by constructing an incidence structure.

Theorem 6. *The model checking problem for* GNFO *is* $\mathrm{P}^{\mathrm{NP}[O(\log^2 n)]}$*-complete.* *For* GNFP *it is in* $\mathrm{NP}^{\mathrm{NP}} \cap \mathrm{coNP}^{\mathrm{NP}}$ *and hard for* P^{NP}.

Expressive power We develop an appropriate notion of bisimulation for GNFO and GNFP, and use it to characterize the expressive power of GNFO.

Recall that a tuple of elements of a structure M is said to be *guarded* if there is a fact of M in which all elements from the tuple occur. We denote by *guarded*(M) the set of all guarded tuples of M. If M and N are structures and $\bar{a}$ and $\bar{b}$ are tuples of elements from $dom(M)$ and $dom(N)$, respectively, then we say that $M, \bar{a}$ and $N, \bar{b}$ are *locally isomorphic* if there is a partial isomorphism $f : M \to N$ such that $f(\bar{a}) = \bar{b}$.

Definition 3. *Let M, N be two structures. A GN-bisimulation (of width $k \geq 1$) is a binary relation $Z \subseteq$ guarded$(M) \times$ guarded(N) such that the following hold for every pair $(\bar{a}, \bar{b}) \in Z$, where $\bar{a} = a_1, \ldots, a_m$ and $\bar{b} = b_1, \ldots, b_n$*

- *$M, \bar{a}$ and $N, \bar{b}$ are locally isomorphic (and in particular, $m = n$)*
- *For every finite set $X \subseteq$ dom(M) (with $|X| \leq k$) there is a partial homomorphism $h : M \to N$ whose domain is X, such that $h(a_i) = b_i$ for all a_i in X, and such that every $\bar{a}' \in$ guarded(M) consisting of elements in the domain of h, the pair $(\bar{a}', h(\bar{a}'))$ belongs to Z.*
- *Likewise in the other direction, where $X \subseteq$ dom(N).*

Note that if X above is restricted to guarded sets then we obtain a definition of guarded bisimulation. We write $M \approx_{GN}^{(k)} N$ if there is a non-empty GN-bisimulation (of width k) between M and N.

Proposition 4. *For $k \geq 1$, if $M \approx_{GN}^{k} N$ then M and N satisfy the same* GNFP^k *sentences. In particular, if $M \approx_{GN} N$ then M and N satisfy the same* GNFP *sentences.*

In fact, GN-bisimulation invariance can be used to *characterize* GNFO.

Theorem 7. GNFO *is the $\approx_{GN}$-invariant fragment of* FO*, and for all $k \geq 1$,* GNFO^k *is the $\approx_{GN}^{k}$-invariant fragment of* FO *(on arbitrary structures).*

Finally, building on Theorem 7 and Theorem 5, we can show the following.

Theorem 8. GNFP *has the tree-like model property.*

6 Discussion

We have provided a logical framework generalizing both GFO and UNFO while preserving their nice properties, in particular decidability of satisfiability. Our results on satisfiability carry over to the validity and entailment problems for GNFO, and likewise for GNFP, as these problems are all reducible to each other. For instance, a GNFO entailment $\phi(\bar{x}\bar{y}) \models \psi(\bar{x}\bar{z})$ holds if, and only if, for a fresh relation R of appropriate arity $\exists\bar{x}\bar{y}\bar{z}(\phi(\bar{x}\bar{y}) \wedge R(\bar{x}\bar{y}\bar{z}) \wedge \neg\psi(\bar{x}\bar{z})))$ is not satisfiable.

Another immediate consequence of our results is that query answering for unions of conjunctive queries with respect to guarded negation fixpoint theories (i.e., the analogue of Theorem 1 replacing GFO by GNFP) is decidable and 2ExpTime-complete. Furthermore, although our definition of GNFO does not include constant symbols, they could be added without affecting the complexity of these problems, relying on the same technique used in [7].

It would be tempting to further generalize by including the two variable fragment of FO (FO^2). Unfortunately this would lead to undecidability. Actually a simple combination of FO^2 with UNFO already yields undecidability as FO^2 can express the fact that a relation correspond to inequality ($\forall x, y \;\; R(x, y) \leftrightarrow x \neq y$) and the extension of UNFO with inequality is undecidable [13]. Similarly, unconstrained universal quantification leads to undecidability, since every subformula of the form $\neg\psi(\bar{x})$ can be trivially guarded using a fresh relation $R(\bar{x})$, adding $\forall\bar{x}\; R\bar{x}$ as a conjunct to the main formula.

References

1. Hajnal Andréka, Johan van Benthem, and István Németi. Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 27:217–274, 1998.
2. Vince Bárány and Mikołaj Bojańczyk. Finite satisfiability for guarded fixpoint logic. Draft available at `arXiv:1104.2262v1 [cs.LO]`, 2011.
3. Vince Bárány, Georg Gottlob, and Martin Otto. Querying the guarded fragment. In *Symposium on Logic In Computer Science (LICS)*, 2010.
4. Dietmar Berwanger and Erich Grädel. Games and model checking for guarded logics. In *Logic for Programming, Artificial Intelligence, and Reasoning (LPAR)*, 2001.
5. Mikołaj Bojańczyk. Two-way alternating automata and finite models. In *Intl. Coll. on Automata, Languages and Programming (ICALP)*, 2002.
6. Erich Grädel. Decision procedures for guarded logics. In *Intl. Conf. on Automated Deduction (CADE)*, volume 1632 of *LNCS*, 1999.
7. Erich Grädel. On the restraining power of guards. *J. Symb. Logic*, 64(4):1719–1742, 1999.
8. Erich Grädel. Why are modal logics so robustly decidable? In *Current Trends in Theoretical Computer Science*, pages 393–408. 2001.
9. Erich Grädel, Martin Otto, and Eric Rosen. Undecidability results on two-variable logics. *Arch. Math. Log.*, 38(4-5):313–354, 1999.
10. Erich Grädel and Igor Walukiewicz. Guarded fixed point logic. *Proc. LICS'99*, IEEE 1999.
11. Leonid Libkin. *Elements of Finite Model Theory*. Springer, 2004.
12. Michael Mortimer. On languages with two variables. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 21(8):135–140, 1975.
13. Balder ten Cate and Luc Segoufin. Unary negation. In *Proc. STACS'11*, 2011.
14. Moshe Y. Vardi. Why is modal logic so robustly decidable? In *Descriptive Complexity and Finite Models*, pages 149–184, 1996.