On the Boundedness Problem for Two-Variable First-Order Logic

Phokion G. Kolaitis * Martin Otto†

Abstract

A positive first-order formula is bounded *if the sequence of its stages converges to the least fixed point of the formula within a fixed finite number of steps independent of the input structure. The* boundedness problem *for a fragment $\mathcal{L}$ of first-order logic is the following decision problem: given a positive formula in $\mathcal{L}$, is it bounded?*

In this paper, we investigate the boundedness problem for two-variable first-order logic FO^2*. As a general rule,* FO^2 *is a well-behaved fragment of first-order logic, since it possesses the finite-model property and has a decidable satisfiability problem. Nonetheless, our main result asserts that the boundedness problem for* FO^2 *is undecidable, even when restricted to negation-free and equality-free formulas $\varphi(X, x)$ in which x is the only free variable and X is a monadic relation symbol that occurs within the scopes of universal quantifiers only. This undecidability result contrasts sharply with earlier results asserting the decidability of boundedness for monadic Datalog programs, which amounts to the decidability of boundedness for negation-free and equality-free existential first-order formulas $\psi(X, x)$ in which x is the only free variable and X is a monadic relation symbol.*

We demonstrate that our main result has certain applications to circumscription*, the most well-developed formalism of nonmonotonic reasoning. Specifically, using the undecidability of boundedness for* FO^2*, we show that it is an undecidable problem to tell whether the circumscription of a given* FO^2*-formula is equivalent to a first-order formula. In contrast, the circumscription of every* FO^1*-formula is equivalent to a first-order formula.*

1 Introduction and summary of results

The number of distinct variables occurring in formulas of a given logic $\mathcal{L}$ is a natural syntactic resource. By imposing bounds on this resource, one obtains the fragments $\mathcal{L}^k$, $k \geqslant 1$, that consist of all formulas of $\mathcal{L}$ with at most k distinct variables. As regards first-order logic FO over relational vocabularies, Scott [29] was the first to realize that the two-variable fragment FO^2 is well-behaved and possesses desirable properties that are not shared by any other fragment FO^k, $k \geqslant 3$. Specifically, Scott [29] established the decidability of the *satisfiability problem* for FO^2 without equality: given a sentence of FO^2 without equality, does it have a model? After this, Mortimer [27] showed that the satisfiability problem for FO^2 (with equality) is decidable. Moreover, he showed that FO^2 has the *finite model property*, which means that if a sentence of FO^2 has a model, then it has a finite model. Recently, the computational complexity of the satisfiability problem for FO^2 was investigated in [13], where this problem was shown to be NEXPTIME-complete. In contrast, the three-variable fragment FO^3 lacks the finite model property and, to make matters worse, the satisfiability problem for FO^3 is undecidable (see [6, 20]).

Another feature of FO^2 is that it subsumes certain logics that have attracted considerable attention in computer science. In particular, it is well known that *propositional modal logic* can be viewed as a fragment of FO^2 without equality, through a natural translation based on $(\Box\varphi)^*(x) = \forall y(Exy \rightarrow \varphi^*(y))$. Moreover, FO^2 subsumes several *description logics* that have been studied by researchers investigating knowledge-representation problems in artificial intelligence (see [7, 8]). Thus, the decidability of FO^2 provides a uniform explanation for the decidability of propositional modal logic and of several description logics. Nonetheless, it should be noted that FO^2 has rather limited expressive power on finite structures, since it is only a small fragment of first-order logic. For this reason, researchers have examined extensions of FO^2 with constructs that augment its expressive power, while maintaining the character of a two-variable logic. So far,

*Computer Science Department, University of California, Santa Cruz, CA 95064, U.S.A., kolaitis@cse.ucsc.edu. Research partially supported by NSF Grant CCR-9610257.

†Mathematische Grundlagen der Informatik, RWTH Aachen Technical University, D-52056 Aachen, Germany, otto@informatik.rwth-aachen.de. Currently visiting UC Santa Cruz and Stanford University, supported by a DFG grant.

the main such logics studied are the extension C^2 of FO^2 with *counting first-order quantifiers* $\exists^{\geqslant i}$, and the extension FP^2 of FO^2 with a *least fixed point* operator for positive FO^2-formulas. Several powerful description logics can be embedded in C^2; moreover, *μ-calculus* (the extension of propositional modal logic with a least fixed point operator) can be embedded in FP^2. The satisfiability problem for C^2 and FP^2 has been investigated recently by Grädel, Otto and Rosen [14, 15]. Despite the fact that C^2 does not have the finite model property, the satisfiability problem for C^2 turned out to be decidable [14]. In contrast, the satisfiability problem for FP^2 was shown to be undecidable via a reduction from the *tiling problem* for *domino systems* [15].

In this paper, we investigate further the algorithmic properties of FO^2 by focusing on the *boundedness problem* for this logic. Before describing our results, we will provide some background and motivation for this problem. Least fixed point logic FP is the extension of first-order logic with the least fixed points of all positive first-order formulas. Formally, assume that $\varphi(X, x_1 \cdots x_k)$ is a first-order formula with free variables $x_1, \ldots, x_k$, and X is a k-ary relation symbol. If φ is *positive* in X (which means that all occurrences of X in φ are within an even number of negations), then φ has a *least fixed point*, that is, on every structure $\mathbf{A}$ there is a smallest relation S such that $\mathbf{A} \models \forall \mathbf{x}(S\mathbf{x} \leftrightarrow \varphi(S, \mathbf{x}))$. We write $\varphi^\infty(x_1 \cdots x_k)$ for the k-ary query that, given a structure $\mathbf{A}$, returns the least fixed point of φ on $\mathbf{A}$. It is well known that $\varphi^\infty = \bigcup_\xi \varphi^\xi$, where the *stages* φ^ξ of φ, ξ an ordinal, are defined by the transfinite induction $\varphi^\xi(x_1 \cdots x_k) \Longleftrightarrow \varphi(\bigcup_{\eta<\xi} \varphi^\eta, x_1 \cdots x_k)$. Note that for every structure $\mathbf{A}$ there is a smallest ordinal ξ_0, called the *closure ordinal* of φ on $\mathbf{A}$, such that $\varphi^{\xi_0} = \bigcup_{\eta<\xi_0} \varphi^\eta$ (and, consequently, $\varphi^{\xi_0} = \varphi^\infty$) on $\mathbf{A}$; furthermore, if $\mathbf{A}$ is a finite structure, then ξ_0 is a positive integer less than or equal to $|A|^k$, where $|A|$ is the cardinality of the universe A of $\mathbf{A}$. In general, the closure ordinals depend on the structure $\mathbf{A}$, and may be arbitrarily large. We say that φ is *bounded* if there is a positive integer n_0 such that $\varphi^\infty = \varphi^{n_0}$ on every structure (finite or infinite). In other words, a formula is bounded if the sequence of its stages converges to the least fixed point within a fixed finite number of steps independent of the input structure. The concept of boundedness for positive first-order formulas was introduced by Barwise and Moschovakis [4], who established several different criteria for boundedness; in particular, they showed that a formula φ is bounded if and only if its least fixed point φ^∞ is first-order definable. After this, Minker and Nicolas [26], and Maier, Ullman and Vardi [23] considered boundedness in the context of Datalog, which is the database query language of function-free and negation-free logic programs. During the past fifteen years, the study of Datalog has occupied a prominent place in database theory, as Datalog became the principal query language in deductive database systems (see [30, 1]). In terms of expressive power, Datalog is essentially equivalent to the fragment of least fixed point logic obtained by taking the least fixed points of positive existential first-order formulas without equality and without negation. A Datalog program is *bounded* if the recursive predicates in the program can be computed within a fixed finite number of steps independent of the input database; this amounts to the boundedness of the positive existential first-order formula that defines the same recursive predicates. The *boundedness problem* for Datalog asks: given a Datalog program, is it bounded? Equivalently, this is the question of whether recursion can be eliminated in a given Datalog program. Detecting boundedness in Datalog programs has been recognized as a fundamental problem in deductive databases, because the elimination of recursion is the first approach to optimizing such programs (see [2, 17]). Gaifman et al. [12] showed, however, that the boundedness problem for Datalog is undecidable. In view of this negative result, researchers attempted to find large classes of Datalog programs for which boundedness is decidable, and also to identify the sources of undecidability. In particular, Cosmadakis et al. [10] showed that boundedness is indeed decidable for *monadic* Datalog programs, that is, for programs in which every recursive predicate has arity one. This result is optimal with respect to arity, since boundedness is undecidable for Datalog programs with a single binary recursive predicate, as shown by Hillebrand et al. [16]. It should be noted that the absence of equality is of paramount importance in establishing the decidability of boundedness for monadic Datalog programs; indeed, Gaifman et al. [12] showed that boundedness is undecidable for monadic Datalog programs with $\neq$ (see also [16]).

In effect, the results described in the preceding paragraph settle the boundedness problem for existential first-order formulas. Nonetheless, they leave the boundedness problem for other fragments of first-order logic open. Our goal in this paper is to investigate the boundedness problem for fragments of first-order logic with a fixed number of distinct variables. An inspection of the proofs in Hillebrand et al. [16] reveals that boundedness is undecidable for Datalog programs with five distinct variables; consequently, the boundedness problem for FO^5 is undecidable. On the other hand, the boundedness problem for FO^1 is trivial, since it is easy to verify that every positive FO^1-formula is bounded. The main result of this paper is that the boundedness problem for FO^2 is undecidable. As a matter of fact, we establish the following stronger result.

Main Theorem: *The boundedness problem for* FO^2 *is undecidable, even when restricted to formulas* $\varphi(X, x)$ *that satisfy the following syntactic conditions:* $\varphi(X, x)$ *is a negation-free and equality-free formula of* FO^2 *in which* x *is the only free variable;* X *is a monadic relation symbol; every occurrence of* X *in* $\varphi(X, x)$ *is within the scope of universal quantifiers only.*

The above undecidability result contrasts sharply with the decidability of boundedness for monadic Datalog programs, since the latter amounts to the decidability of boundedness for negation-free and equality-free existential formulas $\psi(X, x)$ in which x is the only free variable and X is a monadic relation symbol.

The undecidability of boundedness for FO^2 is established via a reduction from the tiling problem for domino systems. As mentioned earlier, domino systems were also deployed by Grädel et al. [15] to establish the undecidability of the satisfiability problem for FP^2. From the technical point of view, the reduction presented here is based on a group of new results concerning the expressive power of FP^2. Specifically, first we show that the collection of all pre-wellorderings is axiomatizable by conjunctions of equality-free sentences of FO^2 with sentences of FP^2 that assert the wellfoundedness of compositions of given binary relations; the latter sentences use least fixed points of positive universal formulas with monadic recursive predicates. It should be pointed out that the composition of two binary relations is not definable in FP^2, whereas the wellfoundedness of the composition of any number of binary relations is axiomatizable in FP^2. After this, we show that FP^2 can axiomatize the collection of all pre-wellorderings with a set of representatives from each equivalence class. This makes it possible to reduce the tiling problem for domino systems to the boundedness problem for FO^2. Further technical refinements are then needed to remove all occurrences of negation and thus arrive at the undecidability of boundedness for the restricted fragment of FO^2 described in the statement of the Main Theorem.

Finally, we demonstrate that our main result has certain applications to *circumscription*, the most well-developed formalism of non-monotonic reasoning. Circumscription, introduced by McCarthy [24, 25], is a transformation of first-order formulas $\psi(S)$ to second-order formulas $\psi^*(S)$ asserting that S is a *minimal* predicate satisfying $\psi(S)$. Ideally, one would like to add circumscription axioms to theorem provers and derive the first-order sentences that are true in all minimal models of $\psi(S)$. Since circumscription axioms are inherently second-order in general, a considerable amount of effort has been expended on identifying classes of first-order formulas whose circumscription is equivalent to a first-order formula (see [11] for an up-to-date account of developments in this area). Moreover, it has been established that there is no algorithm for telling whether the circumscription of a given first-order formula is equivalent to a first-order formula; in fact, this problem is undecidable even when restricted to universal first-order formulas [18]. Here, we examine these issues from the perspective of the number of distinct variables in circumscribed first-order formulas. It is quite easy to see that the circumscription of every FO^1-formula is equivalent to a first-order formula. Unfortunately, this nice behavior does not extend to fragments of first-order logic with more than one variable. In fact, using the undecidability of boundedness for FO^2, we show that it is an undecidable problem to tell whether the circumscription of a given FO^2-formula is equivalent to a first-order formula.

2 Preliminaries

Let σ be a relational vocabulary, X a k-ary relation symbol that is not in σ, and $\varphi(X, x_1 \cdots x_k)$ a first-order formula with free variables $x_1, \ldots, x_k$ over the vocabulary $\sigma \cup \{X\}$. If φ is positive in X, then it has a *least fixed point* $\varphi^\infty[\mathbf{A}]$ on every structure $\mathbf{A}$ over σ, that is to say, $\varphi^\infty[\mathbf{A}]$ is the smallest k-ary relation S on $\mathbf{A}$ such that $\mathbf{A} \models \forall \mathbf{x}(S\mathbf{x} \leftrightarrow \varphi(S, \mathbf{x}))$. The least fixed point of such a formula can be computed via a transfinite iteration that defines the *stages* φ^ξ, ξ an ordinal, of φ. Indeed, for every structure $\mathbf{A}$ over σ, let $\varphi^0[\mathbf{A}] = \emptyset$ and $\varphi^\xi[\mathbf{A}] = \{\mathbf{a} \in A^k : \mathbf{A} \models \varphi(\bigcup_{\eta<\xi} \varphi^\eta, \mathbf{a})\}$. It is well known that $\varphi^\infty[\mathbf{A}] = \bigcup_\xi \varphi^\xi[\mathbf{A}]$ (see [28]). Moreover, since φ is positive in X, on every structure $\mathbf{A}$ the sequence of stages is *monotone* (that is, if $\eta < \xi$, then $\varphi^\eta[\mathbf{A}] \subseteq \varphi^\xi[\mathbf{A}]$) and, therefore, eventually constant. It follows that for every structure $\mathbf{A}$ there is a smallest ordinal ξ_0, called the *closure ordinal* of φ on $\mathbf{A}$, such that $\varphi^{\xi_0}[\mathbf{A}] = \bigcup_{\eta<\xi_O} \varphi^\eta[\mathbf{A}]$ (and, consequently, $\varphi^{\xi_0}[\mathbf{A}] = \varphi^\infty[\mathbf{A}]$).

In general, the closure ordinal of φ depends on the given structure; moreover, it may attain arbitrarily large ordinal values, even if the formula φ is syntactically quite simple. To illustrate this point, let $\varphi(X, x)$ be the formula $\forall y(Eyx \to Xy)$, where E is a binary relation symbol in the vocabulary of $\mathbf{A}$. On every structure $\mathbf{A}$, the least fixed point $\varphi^\infty[\mathbf{A}]$ is equal to the *wellfounded part* of E. Consequently, E is wellfounded if and only if $\varphi^\infty[\mathbf{A}]$ is equal to the universe of $\mathbf{A}$.

Moreover, if the structure $\mathbf{A}$ is an ordinal $(\alpha, <)$, then the closure ordinal of φ on $\mathbf{A}$ is equal to α itself.

Definition 2.1.: Let $\varphi(X, x_1 \cdots x_k)$ be a first-order formula that is positive in X, where X is of arity k. We say that φ is *bounded* if the closure ordinals of φ are uniformly bounded by some positive integer on all structures. Equivalently, φ is *bounded* if there is a positive integer n_0 such that $\varphi^\infty[\mathbf{A}] = \varphi^{n_0}[\mathbf{A}]$ for every structure $\mathbf{A}$ (finite or infinite).

Let $\mathcal{L}$ be a collection of first-order formulas. The *boundedness problem for* $\mathcal{L}$ is the following decision problem: given a positive formula $\varphi(X, x_1 \cdots x_k)$ in $\mathcal{L}$, is it bounded? ∎

In the sequel, we will focus on the boundedness problem for two-variable first-order logic FO^2, that is, the fragment of first-order logic in which only two distinct variables x and y are available. Note that a positive first-order formula may be unbounded, while its closure ordinals are uniformly bounded over all finite structures. As an extreme instance of this phenomenon, let $\varphi(X, x)$ be the FO^3-formula $\theta \wedge \forall y(y \prec x \to Xy)$, where θ is a FO^3-sentence asserting that $\prec$ is a linear order without a greatest element. It is obvious that on every finite structure $\mathbf{A}$ the closure ordinal of φ on $\mathbf{A}$ is equal to 0 and $\varphi^\infty[\mathbf{A}] = \emptyset$; in particular, the least fixed point of φ is FO^1-definable on the class of all finite structures, even though it is not definable by any first-order formula on the class of all structures. In contrast to the above, FO^2-boundedness turns out to be quite robust, as demonstrated by the following result.

Theorem 2.2.: *Let φ be an FO^2-formula that is positive in X, and either $\varphi = \varphi(X, x, y)$ with X binary, or $\varphi = \varphi(X, x)$ with X monadic. Then the following statements are equivalent:*

(1) *φ is bounded.*

(2) *φ is* finitely *bounded, that is, there is a positive integer n_0 such that $\varphi^\infty[\mathbf{A}] = \varphi^{n_0}[\mathbf{A}]$ for every finite structure $\mathbf{A}$.*

(3) *φ^∞ is FO^2-definable on finite structures, that is, there is an FO^2-formula ψ such that for every finite structure $\mathbf{A}$*

$$\mathbf{A} \models \varphi^\infty \leftrightarrow \psi.$$

Proof: Let $\varphi = \varphi(X, x, y)$ be positive in X, where X is binary. It is obvious that (1) $\Longrightarrow$ (2). The implication (2) $\Longrightarrow$ (3) follows from the fact that if $\varphi(X, x, y)$ is an FO^2 formula, then for every $n \geqslant 1$ there is an FO^2-formula $\psi^n(x, y)$ that defines on every structure (finite or infinite) the n-th stage φ^n of $\varphi(x, y, X)$. As a matter of fact, this is an instance of a more general result that holds for every k-variable fragment FO^k, $k \geqslant 1$, of first-order logic and for fixed points that do not involve first-order parameters (see [19]).

It remains to show that (3) $\Longrightarrow$ (1). Assume that $\psi(x, y)$ is an FO^2-formula that defines the least fixed point of $\varphi(X, x, y)$ on finite structures. For every $n \geqslant 1$, let $\psi^n(x, y)$ be an FO^2-formula that defines the n-th stage φ^n of φ on every structure. In particular, for every $n \geqslant 1$, we have that $\models_{\mathrm{FIN}} \forall x \forall y(\psi^n(x, y) \to \psi(x, y))$, where $\models_{\mathrm{FIN}}$ denotes validity over finite structures. It follows that for every $n \geqslant 1$ we have that $\models \forall x \forall y(\psi^n(x, y) \to \psi(x, y))$, because each of the above is an FO^2-sentence and, as shown by Mortimer [27], FO^2 has the finite model property. Our next goal is to show that there is a positive integer n_0 such that $\models \forall x \forall y(\psi(x, y) \to \psi^{n_0}(x, y))$.

Towards a contradiction, assume that for every $n \geqslant 1$ there is a structure $\mathbf{A}_n$ such that $\mathbf{A}_n \models \exists x \exists y(\psi(x, y) \wedge \neg \psi^n(x, y))$. Let c and d be two new constant symbols and let Σ be the following set of first-order sentences:

$$\Sigma = \{\psi(c, d)\} \cup \{\neg \psi^n(c, d) : n \geqslant 1\}.$$

Since $\models \forall x \forall y(\psi^n(x, y) \to \psi^{n+1}(x, y))$ for every $n \geqslant 1$, our hypothesis implies that Σ is a finitely satisfiable set of first-order sentences. By the compactness theorem, Σ has a model; as a matter of fact, it has an ω-saturated model $\mathbf{B}$, since every structure has an ω-saturated elementary extension (see [9, page 216]). It is known, however, that on ω-saturated models the closure ordinal of every positive first-order formula is at most ω (see [28, pages 24–25]), which implies that

$$\mathbf{B} \models \forall x \forall y(\varphi^\infty(x, y) \leftrightarrow \bigvee_{n \geqslant 1} \psi^n(x, y)).$$

It follows that $\mathbf{B} \models \psi(c^{\mathbf{B}}, d^{\mathbf{B}}) \wedge \neg \varphi^\infty(c^{\mathbf{B}}, d^{\mathbf{B}})$, where $c^{\mathbf{B}}$ and $d^{\mathbf{B}}$ are the interpretations of the constant symbols c and d on $\mathbf{B}$. Thus, there is a binary relation P on the structure $\mathbf{B}$ such that P is a fixed point of φ, but P does not contain the relation defined by the formula ψ on $\mathbf{B}$. Consequently, the sentence χ

$$\forall x \forall y(\varphi(P, x, y) \leftrightarrow P(x, y)) \wedge \exists x \exists y(\psi(x, y) \wedge \neg P(x, y))$$

has a model. Note that χ is an FO^2-sentence since both φ and ψ are FO^2-formulas. By Mortimer's theorem [27], χ has a finite model. This, however, is a contradiction, since on finite structures $\psi(x, y)$ defines the least fixed point of φ.

We have just shown the existence of a positive integer n_0 such that $\models \forall x \forall y(\psi(x, y) \to \psi^{n_0}(x, y))$. On the other hand, $\models \forall x \forall y(\psi^{n_0}(x, y) \to \psi^{n_0+1}(x, y))$ and $\models \forall x \forall y(\psi^{n_0+1}(x, y) \to \psi(x, y))$. Therefore, we conclude that

$$\models \forall x \forall y(\psi^{n_0}(x, y) \leftrightarrow \psi^{n_0+1}(x, y)),$$

which implies that on every structure $\mathbf{A}$ (finite or infinite) $\varphi^\infty[\mathbf{A}] = \varphi^{n_0}[\mathbf{A}]$.

The case of $\varphi = \varphi(X, x)$ with monadic X is completely analogous. ∎

As mentioned in the introduction, Barwise and Moschovakis [4] showed that an arbitrary positive first-order formula is bounded if and only if its least fixed point is first-order definable on all structures (finite and infinite). Moreover, Ajtai and Gurevich [3] showed that a Datalog program is bounded if and only if it defines a first-order query on finite structures. In effect, this means that a negation-free and equality-free existential first-order formula φ is bounded if and only if its least fixed point φ^∞ is first-order definable on finite structures. It remains an interesting open problem to determine whether a result similar to the Ajtai-Gurevich theorem holds for FO^2, that is, whether the following is true: a positive FO^2-formula is bounded if and only if its least fixed point is first-order definable on finite structures. The preceding Theorem 2.2 confirms this under the additional assumption that the least fixed-point is FO^2-definable on finite structures.

3 Two-variable least fixed point logic

Least fixed point logic FP extends the syntax and semantics of FO to render least fixed points φ^∞ of positive first-order formulas $\varphi(X, x_1 \cdots x_k)$ definable by new formulas $\mathrm{LFP}_{X,x_1\cdots x_k}\varphi$. Suppose that $\varphi(X, x)$ is a positive in X formula of FO^2 such that x is the only free variable of φ and X is a monadic relation symbol. Then $\mathrm{LFP}_{X,x}\varphi$ is a formula in two-variable fixed point logic FP^2 as considered in [15]. Thus, $\forall x \mathrm{LFP}_{X,x}\big(\forall y(Eyx \to Xy)\big)$ is an FP^2-formula asserting that E is a wellfounded relation. In this section, we investigate the expressive power of FP^2 and obtain certain results that will be used in the study of the boundedness problem for FO^2.

The *composition* of two binary relations E_1 and E_2 is the binary relation $E_1 \circ E_2 = \big\{(x, y) \mid \exists z\big(E_1xz \wedge E_2zy\big)\big\}$. By iterating this operation, we obtain compositions $E_1 \circ \cdots \circ E_k$ of k binary relations $E_1, \ldots, E_k$. Note that the composition $E_1 \circ E_2$ of two binary relations is *not* definable in FP^2 (this can easily be shown using two-pebble games). Nonetheless, FP^2 can express the wellfoundedness of the composition of any number of binary relations. Indeed, let $\mathsf{WF}[E_1 \circ \ldots \circ E_k]$ stand for the assertion that the binary relation $E = E_1 \circ \ldots \circ E_k$ is wellfounded. If $\mathbf{A}$ is a structure over a vocabulary containing $E_1, \ldots, E_k$, then $\mathsf{WF}[E_1 \circ \ldots \circ E_k]$ holds if and only if $\varphi^\infty[\mathbf{A}]$ is equal to the universe of $\mathbf{A}$, where $\varphi(X, x)$ is the positive in X, universal, and equality-free FO^2-formula

$$\forall y\big(E_kyx \to \forall x\big(E_{k-1}xy \to \cdots \to \forall x\big(E_1xy \to Xx\big)\cdots\big)\big).$$

Thus, $\mathsf{WF}[E_1 \circ \ldots \circ E_k]$ is expressible by the FP^2-sentence $\forall x \mathrm{LFP}_{X,x}\varphi(X, x)$.

Actually φ^∞, or membership in the wellfounded part of a composition $E_1 \circ \ldots \circ E_k$, is even definable in the fragment of FP^2 corresponding to the (multi-)modal μ-calculus L_μ, provided that modal operators $\Box_{E^{-1}}$ for the converses E^{-1} of $E = E_1, \ldots, E_k$ are available. Indeed, φ^∞ is easily seen to be equivalent with the L_μ-formula $\mu_X \Box_{E_k^{-1}} \cdots \Box_{E_1^{-1}} X$.

If Δ is a monadic predicate, let $\mathsf{WF}[E_1 \circ E_2 \circ \cdots \circ E_k]^\Delta$ stand for the assertion that $E_1 \circ \ldots \circ E_k$ is *wellfounded over* Δ, that is, $\Delta = \varphi^\infty$, where $\varphi(X, x)$ is the FO^2-formula

$$\Delta x \wedge \forall y\big(E_kyx \to \forall x\big(E_{k-1}xy \to \cdots \to \forall x\big(E_1xy \to (\Delta x \to Xx)\big)\cdots\big)\big).$$

The wellfoundedness assertion $\mathsf{WF}[E_1 \circ \cdots \circ E_k]^\Delta$ is clearly also definable in FP^2. Although this assertion concerns the behaviour of the individual relations E_i even outside Δ, it should be noted that the recursive process for checking $\mathsf{WF}[E_1 \circ E_2 \circ \cdots \circ E_k]^\Delta$ takes place within Δ, in the sense that every stage φ^ξ is a subset of Δ.

The following definition is a straightforward generalization of the notion of a *quotient structure* to the setting in which the given equivalence relation need not be a congruence.

Definition 3.1.: If $\mathbf{A}$ is a relational structure and $\approx$ is an equivalence relation over A, then the *weak quotient* of $\mathbf{A}$ with respect to $\approx$, denoted by $\mathbf{A}^\approx$, is the structure whose universe is the quotient set $A/\approx$, with interpretations $R^\approx$ for each predicate R over $\mathbf{A}$ defined as follows: $([a_1], \ldots, [a_r]) \in R^\approx$ if and only if $R \cap [a_1] \times [a_2] \times \cdots \times [a_r] \neq \emptyset$. ∎

A pre-ordering gives rise to a linear ordering on a quotient of the universe with respect to an equivalence relation. As a canonical vocabulary for pre-orderings, we choose $<$ for the strict pre-ordering relation itself, and $\sim$ for the induced congruence, defined by $x \sim y \iff \neg(x < y \vee y < x)$.

Definition 3.2.: A *pre-wellordering* is a wellfounded pre-ordering; equivalently, it is a preordering whose quotient ordering is a wellordering. For a pre-wellordering $(A, <, \sim)$, we say that $\Delta \subseteq A$ is a *system of representatives* if Δ contains at least one element from each $\sim$-class. ∎

Lemma 3.3.: *$(A, <, \sim)$ is a pre-wellordering if and only if the following three conditions are satisfied:*

(i) $\sim \, = \, \not< \cap \not<^{-1}$;

(ii) $\sim$ *is reflexive;*

(iii) $\sim \circ < \circ \sim$ *is wellfounded.*

Consequently, the class of pre-wellorderings is axiomatizable by a conjunction of an equality-free sentence of FO^2 *with a wellfoundedness assertion for a composition of basic relations.*

Proof: For the crucial direction, assume that $\mathbf{A} = (A, <, \sim)$ satisfies (i) – (iii). By (i), $\sim$ is symmetric, and $(A, <, \sim)$ satisfies a trichotomy property: $\forall x \forall y \big(x < y \vee y < x \vee x \sim y\big)$. As $\sim$ is reflexive, (iii) implies in particular that $<$ itself, as well as $< \circ \sim$ and $\sim \circ <$, are wellfounded, as these compositions are included in $\sim \circ < \circ \sim$. This implies that the trichotomy is strict in the sense that the alternative is exclusive. It suffices to show now that $<$ is transitive and satisfies $\sim \circ < \, = \, < \circ \sim \, = \, <$, that is, that $\sim$ is a congruence for $<$. *Transitivity*: Suppose that $a_1 < a_2 < a_3$. By trichotomy, it suffices to exclude the possibilities $a_3 < a_1$ (which would violate wellfoundedness of $<$) and $a_1 \sim a_3$ (which would violate wellfoundedness of $< \circ \sim$ via $a_1 < a_2 \sim a_2 < a_3 \sim a_1$). *Invariance under* $\sim$: Suppose, for instance, that $a_1 < a_2 \sim a_3$. We want to show that $a_1 < a_3$, again by excluding the other possibilities: $a_3 < a_1$ would violate wellfoundedness of $< \circ \sim$, and $a_1 \sim a_3$ would violate wellfoundedness of $\sim \circ < \circ \sim$. ∎

Lemma 3.4.: *The class of pre-wellorderings with a system of representatives is axiomatizable by the conjunction of an equality-free sentence ψ_0 of* FO^2 *with a wellfoundedness assertion for a composition of basic relations over Δ.*

Proof: *(Sketch)* Let $\psi_0(<, \sim, \Delta)$ be the conjunction of the following FO^2-sentences:

(i) $\forall x \forall y \big(x \sim y \leftrightarrow \neg(x < y \vee y < x)\big)$;

(ii) $\forall x \; (x \sim x)$;

(iii) $\forall x \exists y \big(\Delta y \wedge x \sim y\big)$.

We claim that $\psi = \psi_0 \wedge \mathsf{WF}[\sim^3 \circ < \circ \sim^3]$ provides the desired axiomatization.

For the non-trivial direction assume that $\mathbf{A} \models \psi$ and reason in $\mathbf{A}$. Being a subset of $(\sim^3 \circ < \circ \sim^3) \restriction \Delta$, the relation $(\sim \restriction \Delta) \circ (< \restriction \Delta) \circ (\sim \restriction \Delta)$ is wellfounded; therefore, by Lemma 3.3, $< \restriction \Delta$ is a pre-wellordering. Wellfoundedness of $\sim^2 \circ < \circ \sim^2$ over Δ, together with (i) and (iii), implies that every x has a *unique* representative in $\Delta/\sim$, in the sense that each set $\{y \in \Delta \,|\, y \sim x\}$ intersects exactly one $\sim$-class over Δ. Supposing for instance that there were y and y' in Δ such that $y < y'$ and both equivalent to x, then $y < y' \sim x \sim y$ would be a forbidden cycle. Let $\delta(x)$ denote the representative of x in $\Delta/\sim$, and recall that $<$ induces a wellordering on this quotient. We argue that $x \sim x'$ if and only if $\delta(x) = \delta(x')$, whence in particular $\sim$ is an equivalence relation. For the 'if'-part, consider $y \sim y'$ in Δ and let $x \sim y$ and $x' \sim y'$; then, for instance, $x < x'$ is ruled out, because it would give rise to a forbidden loop $y \sim x < x' \sim y' \sim y$ w.r.t. $\sim \circ < \circ \sim^2$ over Δ. For the 'only-if'-part, let $x \sim x'$, $x \sim y$ and $x' \sim y'$ for y and y' in Δ; then, for instance, $y < y'$ is ruled out, as $y' \sim x' \sim x \sim y < y'$ would be a forbidden loop w.r.t. $\sim^3 \circ <$ over Δ. Finally, using an argument concerning forbidden cycles just as above, it can also be shown that $x < x'$ if and only if $\delta(x) < \delta(x')$, ∎

4 Undecidability of boundedness for two-variable first-order logic

The undecidability of boundedness for FO^2 will be established by reducing the tiling problem for domino systems to the boundedness problem for FO^2. We begin by presenting the basic definitions and a minimum amount of background material.

Definition 4.1.: A *domino system* $\mathcal{D} = (D, R_H, R_V)$ consists of a finite set D of domino pieces together with two binary relations R_H and R_V over D, which, intuitively, specify horizontal and vertical adjacencies. A $\mathcal{D}$*-tiling* of $\mathbb{N} \times \mathbb{N}$ is a mapping $(r, s) \mapsto t(r, s)$ from $\mathbb{N}^2$ to D that respects horizontal and vertical adjacencies, i.e., for all natural numbers r and s, we have that $(t(r, s), t(r+1, s)) \in R_H$ and $(t(r, s), t(r, s+1)) \in R_V$. Tilings of $\mathbb{Z} \times \mathbb{Z}$ and tilings of finite grids $n \times m$ are defined analogously. ∎

Theorem 4.2.: (Berger [5]) *The following domino problem is undecidable: given a domino system $\mathcal{D}$, determine whether there is a $\mathcal{D}$-tiling of* $\mathbb{N} \times \mathbb{N}$.

Call a domino system $\mathcal{D}$ *bounded*, if there is a positive integer n such that there is no $\mathcal{D}$-tiling of the square grid $n \times n$. By König's Lemma, the following are equivalent for any domino system $\mathcal{D}$ (see [6]):

(i) D tiles $\mathbb{N} \times \mathbb{N}$.

(ii) D tiles $\mathbb{Z} \times \mathbb{Z}$.

(iii) D is not bounded.

The *boundedness problem for domino systems* asks: given a domino system, is it bounded? It follows that the boundedness problem for domino systems is undecidable.

In the sequel, we will show that the boundedness of a domino system $\mathcal{D} = (D, R_H, R_V)$ can be linked to the boundedness of a formula $\varphi_{\mathcal{D}}(X, x) \in \mathrm{FO}^2$. The vocabulary of $\varphi_{\mathcal{D}}(X, x)$ will contain binary predicates H, V, $<_H$, $<_V$, $\sim_H$, $\sim_V$, unary predicates P_d for each domino piece $d \in D$, and a unary predicate Δ. The intended roles in the standard encoding of a tiling are as follows: H and V are the horizontal and vertical successor relations of the canonical grid on $\mathbb{N} \times \mathbb{N}$; Δ is the diagonal consisting of the points $(i, i) \in \mathbb{N} \times \mathbb{N}$; $<_H$ and $<_V$ are the pre-wellorderings of horizontal and the vertical precedence, respectively; $\sim_H$ and $\sim_V$ are the associated equivalence relations of belonging to the same column and row, respectively; each P_d represents the placement of the domino piece d on vertices of the grid.

By way of motivating some of the constructions in the main proof, we show that encodings of valid $\mathcal{D}$-tilings can be axiomatized by a conjunction of an FO^2-sentence $\psi_{\mathcal{D}}$ with several assertions of wellfoundedness over Δ.

Lemma 4.3.: *For each domino system $\mathcal{D}$, there are an FO^2-sentence without equality $\psi_{\mathcal{D}}$ and four compositions R_1, R_2, R_3, R_4 of basic relations such that $\psi_{\mathcal{D}} \wedge \bigwedge_{s=1}^4 \mathsf{WF}(R_s)^\Delta$ is satisfiable if and only if $\mathcal{D}$ tiles $\mathbb{N} \times \mathbb{N}$.*

Proof: Put
$$\begin{aligned} R_1 &= (\sim_H)^3 \circ <_H \circ (\sim_H)^3, \\ R_2 &= (\sim_V)^3 \circ <_V \circ (\sim_V)^3, \\ R_3 &= \sim_H \circ (<_H)^2 \circ H^{-1} \circ \sim_H, \\ R_4 &= \sim_V \circ (<_V)^2 \circ V^{-1} \circ \sim_V, \end{aligned}$$
and let $\psi_{\mathcal{D}}$ be the conjunction of the following FO^2-sentences, where $\psi_0(<, \sim, \Delta)$ is as in Lemma 3.4, and $x \approx y$ is used as shorthand for $x \sim_H y \wedge x \sim_V y$:

(i) $\psi_0(<_H, \sim_H, \Delta) \wedge \psi_0(<_V, \sim_V, \Delta)$.

(ii) $\forall x \exists y\, Hxy \ \wedge\ \forall x \exists y\, Vxy$.

(iii) $\forall x \forall y \big(Hxy \rightarrow x <_H y \wedge x \sim_V y\big)$
$\wedge \forall x \forall y \big(Vxy \rightarrow x <_V y \wedge x \sim_H y\big)$.

(iv) $\forall x \forall y \big(\Delta x \wedge x \approx y \rightarrow \Delta y\big)$.

(v) $\forall x \forall y \big(\Delta x \wedge \Delta y \rightarrow (x <_H y \leftrightarrow x <_V y)\big)$.

(vi) the P_d partition the universe, are closed under $\approx$, and respect H and V (in the sense prescribed by R_H and R_V in $\mathcal{D}$).

We claim that $\mathcal{D}$ tiles $\mathbb{N} \times \mathbb{N}$ if and only if the sentence $\psi_{\mathcal{D}} \wedge \bigwedge_{s=1}^4 \mathsf{WF}(R_s)^\Delta$ is satisfiable. For the more challenging direction, let us argue in a model $\mathbf{A}$ of this sentence. Since $\mathsf{WF}(R_1)^\Delta$, $\mathsf{WF}(R_2)^\Delta$, and the statements in (i) hold, Lemma 3.4 implies that $(<_H, \sim_H)$ and $(<_V, \sim_V)$ are pre-wellorderings with Δ as a common system of representatives. Let h and v be the ordinal-valued rank functions with respect to $<_H$ and $<_V$, respectively.

We show that $\mathsf{WF}(R_3)^\Delta$ and $\mathsf{WF}(R_4)^\Delta$ enforce that H-edges and V-edges always link elements to elements in the immediate successor levels with respect to $<_H$ or $<_V$, respectively. Consider $a \in A$, with a representative $d \sim_H a$ in Δ. By (ii) there is some a' such that $(a, a') \in H$, and by (iii), $a <_H a'$. We want to show that $h(a') = h(a) + 1$. Otherwise, there would be some a'' for which $a <_H a'' <_H a'$, whence $(a, a) \in (<_H)^2 \circ H^{-1}$, so that $(d, d) \in R_3$ would be a forbidden loop over Δ. Note that by (iii), also $a \sim_V a'$, i.e. $v(a) = v(a')$.

The equivalence classes with respect to $\approx$ are just the fibres with respect to the mapping that sends $a \in A$ to the pair of its ranks $(h(a), v(a))$. Indeed, $\approx$ is a congruence for Δ, the predicates P_d, $<_H$, $\sim_H$, $<_V$, and $\sim_V$, though not necessarily for H and V. With the concept of a weak quotient $\mathbf{A}^\approx$ from Definition 3.1, we obtain canonical interpretations for H and V over $A/\approx$, though. The ranks h and v carry over to $\mathbf{A}^\approx$ directly, and, by what we have shown about H and V above, $([a], [a']) \in H^\approx$ if and only if $h(a') = h(a) + 1$ and $v(a') = v(a)$; similarly, $([a], [a']) \in V^\approx$ if and only if $h(a') = h(a)$ and $v(a') = v(a) + 1$. Inductively, it follows that for all $r, s \in \mathbb{N}$ there is an $\alpha \in A/\approx$ with $h(\alpha) = r$ and $v(\alpha) = s$; let $\alpha(r, s)$ be the unique such $\approx$-class. It is now straightforward to verify that the following is a valid tiling of $\mathbb{N} \times \mathbb{N}$ with $\mathcal{D}$: for r and s in $\mathbb{N}$, let $t(r, s)$ be that $d \in D$ for which $\alpha(r, s) \in P_d$ over $\mathbf{A}^\approx$. ∎

Note that the conjunct (v) in $\psi_{\mathcal{D}}$ is redundant for the statement of the lemma just proved. It will become important in the sequel, when we shall make further use of $\psi_{\mathcal{D}}$. Also, the use of wellfoundedness assertions over Δ, rather than plain wellfoundedness assertions, is by no means necessary for the above reduction; it will play a crucial role, however, in the proof of the main result. Meanwhile, we may use a simplified version of Lemma 4.3 to derive as a corollary a strengthening of the undecidability result for two-variable least fixed point logic FP^2 of [15]. This result considerably narrows the gap between decidable and undecidable extensions of the (multi-modal) μ-calculus L_μ. On the one hand, a recent result by Vardi [31] asserts that L_μ remains decidable in the presence of converse (or past) modalities $\Box_{E^{-1}}$ along with the usual (future) modalities $\Box_E$; on the other hand, we strengthen the undecidability of FP^2 by showing that even the extension of L_μ with just Boolean assertions about basic relations and their converses is undecidable.

For the present treatment, however, Proposition 4.4 is an aside in the sense that neither the result itself nor its proof will be used in the sequel.

Proposition 4.4.: *The following equality-free fragment of two-variable least fixed point logic* FP^2 *is undecidable: conjunctions of* L_μ*-formulae with universally quantified Boolean equations involving the binary accessibility relations and their converses.*

Proof: We may reduce the existence of a $\mathcal{D}$-tiling of $\mathbb{N} \times \mathbb{N}$ to the satisfiability of the following simplified variant of the sentence considered in the proof of Lemma 4.3:

(i) $\sim_H = \not<_H \cap (\not<_H)^{-1}$ and $\sim_V = \not<_V \cap (\not<_V)^{-1}$;

(ii) $\forall x \forall y (Hxy \rightarrow x <_H y \wedge x \sim_V y)$
$\wedge \forall x \forall y (Vxy \rightarrow x <_V y \wedge x \sim_H y)$;

(iii) $\forall x \exists y\, Hxy \ \wedge\ \forall x \exists y\, Vxy$;

(iv) wellfoundedness of each of the following:
$<_H$, $\sim_H \circ <_H \circ \sim_H$, $(<_H)^2 \circ H^{-1}$,
$<_V$, $\sim_V \circ <_V \circ \sim_V$, $(<_V)^2 \circ V^{-1}$.

If $\mathbf{A}$ satisfies the above, then $(<_H, \sim_H)$ and $(<_V, \sim_V)$ are pre-wellorderings by Lemma 3.3 (observe that, for example, wellfoundedness of $<_H$ and (i) imply reflexivity of $\sim_H$). Also H- and V-edges always point into direct successor levels w.r.t. $<_H$, respectively $<_V$ (cf. the argument above based on wellfoundedness of R_3 and R_4). It is now easy to show, just as above, that the weak quotient $\mathbf{A}^{\approx}$ encodes a valid tiling on $\mathbb{N} \times \mathbb{N}$.

Clearly (i) and (ii) are Boolean conditions, as desired. The wellfoundedness assertions in (iv) may be turned into L_μ-assertions stating that a distinguished node (corresponding to the origin in $\mathbb{N} \times \mathbb{N}$) is in the wellfounded part of certain compositions of the basic accessibility relations $<_H$, $<_V$, $\sim_H$, $\sim_V$, H, V, and some of their converses.

It only remains to express the $\forall\exists$-assertions (iii) in L_μ. For this observe that it suffices to assert these in restriction to that part of the universe that is reachable along $(H \cup V)$-paths from the origin (our distinguished node); in this restricted form, these assertions become easily L_μ-expressible. For instance, the formula $\neg \mu_X (\Box_H \bot \vee \Diamond_H X \vee \Diamond_V X)$, where $\bot$ is universally false, asserts that $\forall x \exists y Hxy$ holds along all $(H \cup V)$-paths. ∎

An extension of the preceding proof along the lines indicated in [15] establishes that the satisfiability problem for this restricted fragment of FP^2 is actually still Σ^1_1-hard. Also, the finite satisfiability problem for this fragment is undecidable.

Turning to the main result of this section, think of the above four wellfoundedness assertions over Δ as being checked in an inductive process that covers increasing initial segments of the diagonal Δ. It is important to note that, over the standard model, these four processes all proceed in parallel by adding exactly one new element, namely the next in rank with respect to $<_H$, $<_V$, $(<_H)^2 \circ H^{-1}$, or $(<_V)^2 \circ V^{-1}$; moreover, these four ranks actually coincide over the standard model. It follows that wellfoundedness of these four relations may be formally strengthened to wellfoundedness of their union. We are now ready to establish the undecidability of boundedness for FO^2.

Theorem 4.5.: *There is a recursive reduction of the boundedness problem for domino systems to the boundedness problem for* FO^2 *without equality. Consequently, the boundedness problem for* FO^2 *without equality is undecidable.*

Proof: Given a domino system $\mathcal{D}$, consider the following formula:

$$\varphi_{\mathcal{D}}(X, x) := \neg \psi_{\mathcal{D}} \vee \Delta x \wedge \forall y \big((\textstyle\bigvee_{s=1}^4 R_s yx \wedge \Delta y) \rightarrow Xy \big).$$

Here $\psi_{\mathcal{D}}$ is as in Lemma 4.3. Note that $\varphi_{\mathcal{D}}$ is equivalent to a formula of FO^2 without equality in which every occurrence of X is within the scope of universal quantifiers only. Indeed, the formula $\forall y (\bigvee_s R_s yx \wedge \Delta y \rightarrow Xy)$ is logically equivalent to $\bigwedge_s \forall y (R_s yx \wedge \Delta y \rightarrow Xy)$. Each of these conjuncts may be expressed in FO^2 over the basic predicates as indicated in the discussion of wellfoundedness of compositions.

Intuitively, the sequence of stages $(\varphi_{\mathcal{D}})^\xi$ of $\varphi_{\mathcal{D}}(X, x)$ describes the process that checks wellfoundedness over Δ for the relation $R = \bigcup_{s=1}^4 R_s$. The 'static' conditions expressed in $\psi_{\mathcal{D}}$ are built into $\varphi_{\mathcal{D}}$ in such a way that their violation floods the iteration in the first step. Hence, boundedness of φ is the same as boundedness of φ over models of $\psi_{\mathcal{D}}$. We claim that $\varphi_{\mathcal{D}}(X, x)$ is bounded if and only if $\mathcal{D}$ is bounded, which then proves the theorem. One direction is easy. Suppose that $\mathcal{D}$ admits arbitrarily large square tilings, and hence a tiling of the standard grid on $\mathbb{N} \times \mathbb{N}$. Then the canonical encoding of any $\mathcal{D}$-tiling of $\mathbb{N} \times \mathbb{N}$ satisfies $\psi_{\mathcal{D}}$; moreover, for every $n \geqslant 1$, $(\varphi_{\mathcal{D}})^n = \{(j,j) \in \mathbb{N} \times \mathbb{N} \mid j < n\}$; hence, $\varphi_{\mathcal{D}}$ is unbounded.

Conversely, let $n \geqslant 1$ and let $\mathbf{A}$ be such that $(\varphi_{\mathcal{D}})^{n+1}[\mathbf{A}] \neq (\varphi_{\mathcal{D}})^n[\mathbf{A}]$. Let $\mathbf{A}_n$ be the restriction of $\mathbf{A}$ to the subset $A_n := \{a \in A \mid h(a), v(a) \leqslant n\}$, where h and v are the ranks with respect to $<_H$ and $<_V$. We claim that $\mathbf{A}_n$ encodes a $\mathcal{D}$-tiling of $(n+1) \times (n+1)$: the weak quotient structure $(\mathbf{A}_n)^{\approx}$ with respect to the common refinement $\approx$ of $\sim_H$ and $\sim_V$ (see Definition 3.1) is isomorphic to the standard grid on $(n+1) \times (n+1)$ with a valid $\mathcal{D}$-tiling. The isomorphism is induced by the rank mappings (h, v). In other words, $(\varphi_{\mathcal{D}})^{n+1}[\mathbf{A}] \neq (\varphi_{\mathcal{D}})^n[\mathbf{A}]$ implies that

$\mathcal{D}$ tiles $(n+1)\times(n+1)$, so that unboundedness of $\varphi_{\mathcal{D}}$ implies unboundedness of $\mathcal{D}$.

Note that $\mathbf{A}\models\psi_{\mathcal{D}}$, since otherwise $(\varphi_{\mathcal{D}})^1[\mathbf{A}]=A=(\varphi_{\mathcal{D}})^n[\mathbf{A}]$ for all $n\geqslant 1$. But this immediately implies that the relations $P_d^{\approx}$ partition $A_n/\approx$, and also respect $H^{\approx}$ and $V^{\approx}$, as required by $\mathcal{D}$ (cf. conjunct (vi) in $\psi_{\mathcal{D}}$). It remains to show that $(\mathbf{A}_n)^{\approx}$ is well defined (that is, $\approx$ is an equivalence relation on A_n) and that the rank mappings (h,v) induce an isomorphism onto the standard grid on $(n+1)\times(n+1)$.

Let $D_i:=(\varphi_{\mathcal{D}})^i[\mathbf{A}]$. Observe that D_i is contained in Δ and in the wellfounded part of $R\restriction\Delta$. As $R=\bigcup R_s$, clearly $R\restriction\Delta=\bigcup_{s=1}^4(R_s\restriction\Delta)$, whence D_i is in the wellfounded part of the individual $R_s\restriction\Delta$. This implies in particular that the individual R_s are wellfounded over D_i. Let $\langle D_i\rangle^{\sim_H}=\{a\in A\,|\,\exists d\in D_i(a\sim_H d)\}$ and $\langle D_i\rangle^{\sim_V}=\{a\in A\,|\,\exists d\in D_i(a\sim_V d)\}$. We can first show the following:

(i) w.r.t. both $<_H$ and $<_V$, D_i is an initial, pre-wellordered segment of Δ.

(ii) the sets $\langle D_i\rangle^{\sim_H}$ and $\langle D_i\rangle^{\sim_V}$ are initial segments w.r.t. $<_H$ and $<_V$, respectively.

(iii) $<_H$ and $<_V$ are pre-wellorderings over $\langle D_i\rangle^{\sim_H}$ and $\langle D_i\rangle^{\sim_V}$, respectively.

(iv) D_i is closed w.r.t. $\approx$.

For (i), let us show how wellfoundedness of R_1 implies that D_i is an initial segment of Δ w.r.t. $<_H$. Recall that $R_1=(\sim_H)^3\circ<_H\circ(\sim_H)^3$. If $d,d'\in\Delta$, $d'<_H d$ and $d\in D_i$, then $(d',d)\in R_1\restriction\Delta$, which implies that $d'\in D_{i-1}$. The argument for (ii) is similar: if $a\in\langle D_i\rangle^{\sim_H}$ and $a'<_H a$, then there is a $d\in D_i$ such that $a\sim_H d$, and (by $\psi_{\mathcal{D}}$) at least one $d'\in\Delta$ such that $a'\sim_H d'$. Now $d'<_H d$, as otherwise d would have to be on a cycle w.r.t. $R_1\restriction\Delta$: if $d<_H d'$, then $d<_H d'\sim_H a'<_H a\sim_H d$; and if $d\sim_H d'$, then $d\sim_H d'\sim_H a'<_H a\sim d$. It follows that $d'\in D_i$, whence $a'\in\langle D_i\rangle^{\sim_H}$. (iii) is shown just as in Lemma 3.4. It also follows that $D_i\subseteq\Delta$ is a system of representatives, both for $<_H$ on $\langle D_i\rangle^{\sim_H}$, and for $<_V$ on $\langle D_i\rangle^{\sim_V}$. (iv) follows from the fact that, in restriction to $\langle D_i\rangle^{\sim_H}\cap\langle D_i\rangle^{\sim_V}$, both $\sim_H$ and $\sim_V$ are equivalence relations. The form of the recursive part of $\varphi_{\mathcal{D}}(X,x)$ now implies that all stages D_j, for $j\leqslant i$, are closed w.r.t. $\approx$.

In particular, $\approx$ is an equivalence relation (indeed a congruence for $<_H$, $\sim_H$, $<_V$, $\sim_V$, D_i and the P_d) in restriction to $\langle D_i\rangle^{\sim_H}\cap\langle D_i\rangle^{\sim_V}$. Let $\mathbf{A}^{D_i}$ be the restriction of $\mathbf{A}$ to $\langle D_i\rangle^{\sim_H}\cap\langle D_i\rangle^{\sim_V}$, but with Δ interpreted as D_i:

$$\mathbf{A}^{D_i}=\big[A,\ldots,(P_d)_{d\in D},D_i\big]\restriction\big(\langle D_i\rangle^{\sim_H}\cap\langle D_i\rangle^{\sim_V}\big).$$

We may pass to the weak quotient structure $\big[\mathbf{A}^{D_i}\big]^{\approx}$, and show that in restriction to this weak quotient the rank-mapping (h,v) induces an isomorphism between $H^{\approx}$ and $V^{\approx}$ and the successor relations on a standard grid on some $n_i\times m_i$. In particular, we have to show that, when restricted to $\langle D_i\rangle^{\sim_H}\cap\langle D_i\rangle^{\sim_V}$, H-edges point into the immediate successor levels with respect to $<_H$ (while preserving $\sim_V$); and similarly for V. Consider first some $(a,a')\in H\restriction\langle D_i\rangle^{\sim_H}$. Since $\langle D_i\rangle^{\sim_H}$ is an initial segment with respect to $<_H$, the $<_H$-ranks restricted to $\langle D_i\rangle^{\sim_H}$ are the same as on the universe. It now follows from the wellfoundedness of R_3 over D_i that $h(a')=h(a)+1$ (just as in the proof of Lemma 4.3). Conversely, let $a,a'\in\langle D_i\rangle^{\sim_H}$ be such that $h(a')=h(a)+1$ and $a'\sim_V a$. Let d be an element in D_i with $a\sim_H d$. By $\psi_{\mathcal{D}}$, there is an outgoing H-edge at a to some a'' such that $a<_H a''$ and $a\sim_V a''$. If $a'<_H a''$, then $(a,a)\in(<_H)^2\circ H^{-1}$ and therefore $(d,d)\in R_3\restriction D_i$, contradicting wellfoundedness of R_3 over D_i. Therefore $a'\sim_H a''$, i.e. $a'\approx a''$, and $([a],[a'])\in H^{\approx}$.

Having established the desired isomorphism with a standard grid on some $n_i\times m_i$, we note that the values n_i and m_i are determined as the suprema of the respective ranks. As D_i provides a common system of representatives for the ranks, $n_i=|D_i/\sim_H|$ and $m_i=|D_i/\sim_V|$. But now conjunct (v) in $\psi_{\mathcal{D}}$, $\forall x\forall y\big(\Delta x\wedge\Delta y\rightarrow(x<_H y\leftrightarrow x<_V y)\big)$, implies that $n_i=m_i=|D_i/\approx|$, since $D_i\subseteq\Delta$. Furthermore, as each D_i is closed with respect to $\approx$, $D_{n+1}\neq D_n$ implies $|D_{n+1}/\approx|\geqslant n+1$ (and, for rank-reasons, equality holds). So $\langle D_{n+1}\rangle^{\sim_H}\cap\langle D_{n+1}\rangle^{\sim_V}=A_n$, proving the claim. ∎

The proof of Theorem 4.5 actually shows that the boundedness problem for FO^2 is undecidable, even when restricted to formulas $\varphi(X,x)$ of FO^2 without equality in which x is the only free variable, X is a monadic relation symbol occurring positively, and every occurrence of X is within the scope of universal quantifiers only. However, relation symbols from the underlying vocabulary may occur negatively in the formulas. Our goal is to sharpen Theorem 4.5 by eliminating all occurrences of negation. The obvious approach to eliminating negations in the formula $\varphi_{\mathcal{D}}(X,x)$ is to extend the vocabulary with new predicates for the complements of all relations that have to be used negatively. Consider the introduction of a new predicate $\tilde{P}$ for the complement of some monadic basic predicate P in $\varphi_{\mathcal{D}}(X,x)$ (e.g., one of the the monadic predicates P_d marking the positions where the domino piece d is placed). Recall that $\varphi_{\mathcal{D}}(X,x)$ was of the form $\varphi_{\mathcal{D}}(X,x)=\neg\psi_{\mathcal{D}}\vee\chi(X,x)$, where $\psi_{\mathcal{D}}$ sees to the static conditions (through flooding), while $\chi(X,x)$ governs the induction. To ensure that $\forall x\,(Px\leftrightarrow\neg\tilde{P}x)$, we may pass to

$$\varphi'_{\mathcal{D}}(X,x) = \neg\psi_{\mathcal{D}} \vee \xi_1 \vee \big(\xi_2 \wedge \chi(X,x)\big),$$

where $\xi_1 = \exists x(Px \wedge \tilde{P}x)$ and $\xi_2 = \forall y(Py \vee \tilde{P}y)$. Note that these disjuncts/conjuncts contribute only positive occurrences of P and $\tilde{P}$. Over structures $\mathbf{A}$ on which $(\varphi'_{\mathcal{D}})^1$ is neither the full universe nor empty, $\tilde{P}$ actually has to be the complement of P: $\mathbf{A} \models \neg\xi_1$ as otherwise $(\varphi'_{\mathcal{D}})^1 = A$, and then $\mathbf{A} \models \xi_2$, as otherwise $(\varphi'_{\mathcal{D}})^1 = \emptyset$. Binary predicates in $\varphi_{\mathcal{D}}$ can be treated analogously. Let $\varphi^*_{\mathcal{D}}(X,x)$ be the result of adding corresponding new disjuncts and conjuncts into $\varphi_{\mathcal{D}}(X,x)$ in the above fashion, and then pushing negations to the atomic level and replacing negative occurrences of predicates by positive occurrences of the relation symbols for their complements. Thus, $\varphi^*_{\mathcal{D}}$ becomes positive in all predicates (and still has X in the scope of universal quantifiers only). We claim that $\varphi^*_{\mathcal{D}}(X,x)$ is bounded if and only if $\mathcal{D}$ is bounded. Clearly, $\varphi^*_{\mathcal{D}}(X,x)$ is bounded over the class of all structures if and only if it is bounded over the class of all those structures that satisfy $\bigwedge\neg\xi_1 \wedge \bigwedge\xi_2$. Over this class, however, $\varphi^*_{\mathcal{D}}(X,x)$ is logically equivalent to $\varphi_{\mathcal{D}}(X,x)$. Thus, we have established the following result.

Main Theorem: *The boundedness problem for* FO^2 *is undecidable, even when restricted to formulas $\varphi(X,x)$ that satisfy the following syntactic conditions: $\varphi(X,x)$ is a negation-free and equality-free formula of* FO^2 *in which x is the only free variable; X is a monadic relation symbol; every occurrence of X in $\varphi(X,x)$ is within the scope of universal quantifiers only.*

The results obtained in this section reveal that the boundedness problem is undecidable, even when restricted to a rather small fragment of FO^2. Nonetheless, the following questions remains open.

Question: Is the boundedness problem for positive universal FO^2-formulas decidable? If so, what about the boundedness problem for arbitrary positive universal first-order formulas?

5 On the circumscription of first-order formulas with two variables

Let σ be a relational vocabulary and S a relation symbol that is not a member of σ. If $\psi(S)$ is a first-order formula over the vocabulary $\sigma \cup \{S\}$, then the *circumscription of S in $\psi(S)$* is the following second-order formula $\psi^*(S)$:

$$\psi(S) \wedge (\forall T)[T \subset S \rightarrow \neg\psi(T)],$$

where $T \subset S$ is shorthand for a first-order formula stating that T is a proper subset of S. In non-monotonic reasoning, one is usually interested in first-order assertions about a predicate S that are true in all *minimal* models of some first-order sentence $\psi(S)$. The circumscription $\psi^*(S)$ of S in $\psi(S)$ provides a concise syntactic axiomatization of minimality. Thus, one would like to have a complete deductive system for deriving all first-order consequences of circumscription. In general, this is impossible, because there are first-order formulas $\psi(S)$ such that the set of first-order consequences of their circumscription $\psi^*(S)$ is not recursively enumerable. One way to avoid this difficulty is to concentrate on first-order formulas whose circumscription is equivalent to some set of of first-order formulas. As mentioned in the introduction, however, it is an undecidable problem to tell whether, given a first-order formula $\psi(S)$, the circumscription $\psi^*(S)$ is expressible in first-order logic.

Starting with the work of Lifschitz [21], researchers in non-monotonic reasoning have searched systematically for large classes of first-order formulas on which either circumscription is always expressible in first-order logic or there is an algorithm to tell whether circumscription is expressible in first-order logic. Typically, these classes are obtained by imposing various syntactic restrictions on the quantifier pattern of the formulas or on their quantifier-free part (see [21, 22, 11]). So far, however, no investigation has been carried out on how the number of distinct variables affects the properties of circumscription. Let $\psi(S)$ be a formula of FO^1. Without loss of generality, we may assume that S is a monadic relation symbol and that the vocabulary σ consists entirely of monadic relation symbols. Consequently, the circumscription $\psi^*(S)$ of S in $\psi(S)$ is a formula of monadic second-order logic over a vocabulary of monadic relation symbols. It is well known (and part of the "folklore") that every such formula is equivalent to a first-order formula. Thus, if $\psi(S)$ is a formula of FO^1, then the circumscription $\psi^*(S)$ is expressible in first-order logic. Our final result in this paper reveals that the situation is dramatically different for formulas of FO^2.

Theorem 5.1.: *The following problem is undecidable: given a formula $\psi(S)$ of* FO^2*, is the circumscription $\psi^*(S)$ expressible in first-order logic?*

Proof: *(Sketch)* We will show that the boundedness problem for FO^2 is reducible to the problem at hand. Once this is established, the result will follow immediately from Theorem 4.5.

Suppose we are given a formula $\varphi(S,x)$ of FO^2 over a vocabulary $\sigma \cup \{S\}$ such that x is its only free variable and S is a monadic relation symbol having only positive occurrences. Let $\psi(S)$ be the FO^2-sentence $\forall x(\varphi(S,x) \leftrightarrow Sx)$, which asserts that S is a fixed point of $\varphi(S,x)$. We now claim that the following statements are equivalent:

(1) The formula $\varphi(S,x)$ is bounded.

(2) The circumscription $\psi^*(S)$ is expressible in first-order logic.

Note that $\psi^*(S)$ asserts that S is the least fixed point of $\varphi(S,x)$. It follows that if $\varphi(S,x)$ is bounded, then $\psi^*(S)$ is equivalent to the first-order sentence $\forall x(Sx \leftrightarrow \varphi^{n_0}(x))$, where n_0 is a positive integer such that $\varphi^\infty = \varphi^{n_0}$. For the other direction, assume that $\psi^*(S)$ is equivalent to a set Σ of first-order sentences over the vocabulary $\sigma \cup \{S\}$, but $\varphi(S,x)$ is unbounded. Let c be a new constant symbol and let Σ' be the following set of first-order sentences:

$$\Sigma' = \Sigma \cup \{S(c)\} \cup \{\neg\varphi^n(c) : n \geqslant 1\}.$$

Since $\varphi(S,x)$ was assumed to be unbounded, Σ' is finitely satisfiable. As in the proof of the earlier Theorem 2.2, it follows that Σ' has an ω-saturated model $\mathbf{A}$. Since $\mathbf{A} \models \Sigma$, the interpretation $S^{\mathbf{A}}$ of the relation symbol S on $\mathbf{A}$ is equal to the least fixed point $\varphi^\infty[\mathbf{A}]$ of $\varphi(S,x)$ on A. Moreover, the ω-saturation of $\mathbf{A}$ implies that the closure ordinal of φ on $\mathbf{A}$ is at most ω. Consequently, $\varphi^\infty[\mathbf{A}] = \bigcup_{n\geqslant 1} \varphi^n[\mathbf{A}]$. This, however, contradicts the fact that $\mathbf{A} \models \{S(c)\} \cup \{\neg\varphi^n(c) : n \geqslant 1\}$. ∎

References

[1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of databases.* Addison-Wesley, 1995.

[2] S. Abiteboul and P. C. Kanellakis. Database Theory Column. *SIGACT News*, 20:17–23, 1989.

[3] M. Ajtai and Y. Gurevich. Datalog vs. first-order logic. *Journal of Computer and System Sciences*, 49:562–588, 1994.

[4] J. Barwise and Y. Moschovakis. Global inductive definability. *Journal of Symbolic Logic*, 43(3):521–534, 1978.

[5] R. Berger. *The undecidability of the domino problem*, volume 66 of *Mem. AMS.* American Mathematical Society, 1966.

[6] E. Börger, E. Grädel, and Y. Gurevich. *The Classical Decision Problem.* Springer-Verlag, 1997.

[7] A. Borgida. On the relationship between description logic and predicate logic queries. In N. Adam, B. Bhargava, and Y. Yesha, editors, *Proceedings of the Third International Conference on Information and Knowledge Management, CIKM 94*, pages 219–225, 1994.

[8] A. Borgida. On the relative expressiveness of description logics and predicate logics. *Artificial Intelligence*, 82(1-2):353–67, 1996.

[9] C. C. Chang and H. J. Keisler. *Model Theory.* North Holland, 3rd edition, 1990.

[10] S. S. Cosmadakis, H. Gaifman, P. C. Kanellakis, and M. Y. Vardi. Decidable optimization problems for database logic programs. In *Proc. 20th ACM Symp. on Theory of Computing*, pages 477–490, 1988.

[11] P. Doherty, W. Lukaszewicz, and A. Szalas. Computing circumscription revisited: a reduction algorithm. *Journal of Automated Reasoning*, 18(3):297–336, 1997.

[12] H. Gaifman, H. Mairson, Y. Sagiv, and M. Vardi. Undecidable optimization problems for database logic problems. *Journal of the Association for Computing Machinery*, 40(3):683–713, 1993. An earlier version appeared in *Proc. 2nd IEEE Symp. on Logic in Computer Science*, pages 106–115, 1987.

[13] E. Grädel, P. G. Kolaitis, and M. Y. Vardi. The decision problem for 2-variable first-order logic. *Bulletin of Symbolic Logic*, 3:53–69, 1997.

[14] E. Grädel, M. Otto, and E. Rosen. Two-variable logic with counting is decidable. In *Proceedings of 12th IEEE Symposium on Logic in Computer Science - LICS '97*, 1997.

[15] E. Grädel, M. Otto, and E. Rosen. Undecidability results for two-variable logics. In *Proceedings of the 14th Annual Symposium on Theoretical Aspects of Computer Science STACS '97*, volume 1200 of *Lecture Notes in Computer Science*, pages 249–260. Springer-Verlag, 1997. Journal version to appear in *Archive for Mathematical Logic.*

[16] G. G. Hillebrand, P. C. Kanellakis, H. G. Mairson, and M. Y. Vardi. Undecidable boundedness problems for datalog programs. *Journal of Logic Programming*, 25(2):163–190, 1995.

[17] P. C. Kanellakis. Elements of relational database theory. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 1073–1157. MIT Press/Elsevier, 1990.

[18] P. G. Kolaitis and C. H. Papadimitriou. Some computational aspects of circumscription. *Journal of the ACM*, 37:1–14, 1990.

[19] P. G. Kolaitis and M. Y. Vardi. On the expressive power of variable-confined logics. In *Proc. 11th IEEE Symp. on Logic in Computer Science*, pages 348–359, 1996.

[20] H. R. Lewis. *Unsolvable Classes of Quantificational Formulas.* Addison-Wesley, 1979.

[21] V. Lifschitz. Computing circumscription. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pages 121–127, 1985.

[22] V. Lifschitz. On the satisfiability of circumscription. *Artificial Intelligence*, 28:17–27, 1986.

[23] D. Maier, J. Ullman, and M. Vardi. On the foundations of the universal relation model. *ACM Trans. on Database Systems*, 9:283–308, 1984.

[24] J. McCarthy. Circumscription - a form of nonmonotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.

[25] J. McCarthy. Applications of circumscription in formalizing common sense knowledge. *Artificial Intelligence*, 28:89–116, 1985.

[26] J. Minker and N. Nicolas. On recursive axioms in deductive databases. *Information Systems*, 7(4):1–15, 1982.

[27] M. Mortimer. On languages with two variables. *Zeitschr. f. Math. Logik u. Grundl. d. Math.*, 21:135–140, 1975.

[28] Y. N. Moschovakis. *Elementary Induction on Abstract Structures.* North Holland, 1974.

[29] D. Scott. A decision method for validity of sentences in two variables. *Journal of Symbolic Logic*, 27:377, 1962.

[30] J. D. Ullman. *Database and Knowledge-Base Systems, Volumes I and II.* Computer Science Press, 1989.

[31] M. Vardi. Reasoning about the past with two-way automata. In *Proc. 25th International Colloq. on Automata, Languages, and Programming.* Springer-Verlag, 1998.