Oblivious and Semi-Oblivious Boundedness for Existential Rules

Pierre Bourhis, Michel Leclère, Marie-Laure Mugnier, Sophie Tison, Federico Ulliana, Lily Galois

Oblivious and Semi-Oblivious Boundedness for Existential Rules

Pierre Bourhis[1,3,4], **Michel Leclère**[2,4], **Marie-Laure Mugnier**[2,4], **Sophie Tison**[3,4], **Federico Ulliana**[2,4] and **Lily Galois**[3,4]
[1] CNRS, France
[2] Univ. Montpellier, LIRMM, France
[3] Univ. Lille, CRIStAL, France
[4] Inria, France
{firstname.lastname}@inria.fr

Abstract

We study the notion of boundedness in the context of positive existential rules, that is, whether there exists an upper bound to the depth of the chase procedure, that is independent from the initial instance. By focussing our attention on the oblivious and the semi-oblivious chase variants, we give a characterization of boundedness in terms of FO-rewritability and chase termination. We show that it is decidable to recognize if a set of rules is bounded for several classes and outline the complexity of the problem.

This report contains the paper published at IJCAI 2019 [Bourhis *et al.*, 2019] *and an appendix with full proofs.*

1 Introduction

We consider the setting of ontology-based query answering (OBQA) in which answers to conjunctive queries are logically entailed from a knowledge base constituted of a set of facts (or database instance) and an ontology. Existential rules, also known as Tuple Generating Dependencies (TGDs) in database theory, are an expressive knowledge representation language well studied in the OBQA setting [Calì *et al.*, 2009a; Baget *et al.*, 2011; Calì *et al.*, 2013]. These rules generalize function-free Horn rules (like those of datalog) with existentially quantified variables in the rule heads, which allow one to assert the existence of unknown individuals, and hence to reason in open domains. Beside datalog, existential rules generalize the Semantic Web language RDF Schema, as well as most Description Logics used in the OBQA context, namely Horn description logics, in particular those at the core of the tractable profiles of the ontological language OWL 2.

The two main approaches developed to answer conjunctive queries on existential rules knowledge bases are *materialization* and *query rewriting*. Both can be seen as ways of reducing query answering to a classical database query evaluation problem. Materialization relies on a forward chaining technique, called the *chase*, that consists in expanding the database instance with the facts entailed by rules until fixpoint. In contrast, query rewriting is a backward chaining mechanism that consists in rewriting an input query using relevant rules, so that its answers on the knowledge base are exactly the answers of the rewritten query on the database instance alone. Query answering being undecidable for existential rules, both materialization and query rewriting may not terminate.

This led to intensive research aiming at characterizing decidable and tractable classes of existential rules. Several syntactic restrictions were proposed to ensure chase termination (e.g. weak-acyclicity [Fagin *et al.*, 2005]) or the existence of a (finite) first-order rewriting of a conjunctive query, a property referred as *FO-rewritability* [Calvanese *et al.*, 2007]. Nevertheless, the interactions between chase termination and FO-rewritability have been little investigated so far, and not much is known for existential rules on which both hold. What are the relationships between these two properties?

Answering this question leads us to another fundamental problem, which has been extensively studied for datalog, namely *(uniform) boundedness* [Hillebrand *et al.*, 1995]. Boundedness concerns the recursivity of rules, and asks whether there is an upper bound on the depth of the chase, which is independent from any database instance. The property is key for practical optimization of reasoning as it implies that the ruleset is essentially non-recursive (although syntactic conditions may fail to capture this). It is known that boundedness and FO-rewritability are equivalent in the case of datalog [Ajtai and Gurevich, 1994], but this does not hold for existential rules. In this setting, the notion of boundedness also depends on the chase variant as they all behave differently with respect to termination.

We focus our attention on the *oblivious* and *semi-oblivious* (a.k.a. Skolem) chase [Marnette, 2009]. As a matter of fact, almost all known sufficient conditions for chase termination fall within these chase variants (from the simplest ones: rich-acyclicity [Hernich and Schweikardt, 2007], weak-acyclicity [Fagin *et al.*, 2005] and acyclic-GRD [Baget *et al.*, 2011] to the more general MFA [Grau *et al.*, 2013]), at the exception of the recent work of [Carral *et al.*, 2017] which applies to the restricted chase variant. Importantly, we consider a *breadth-first* version for both variants, which ensures the minimal depth of the chase [Delivorias *et al.*, 2018].[1]

[1]See [Delivorias *et al.*, 2020] for an extended version of this conference paper, to appear in Theory and Practice of Logic Programming (added note w.r.t. IJCAI 2019 paper).

Our main contribution is a characterization of boundedness in terms of chase termination *and* FO-rewritability. This means that a set of rules is bounded if and only if it ensures both chase termination for any instance and FO-rewritability for any conjunctive query. We show this by proving two orthogonal results. The first is a bound on the depth of existential variables when the chase terminates on all instances. The second is a bound on the (breadth-first) rank at which facts using terms of a given depth are inferred.

This connection reveals important differences between the two variants. For the oblivious case we show that, when chase termination holds, FO-rewritability on *full-atomic queries* (queries with a single atom and only answer variables) is equivalent with FO-rewritability. Moreover, for the case of *fully-existential rules* (rules where all head atoms have at least one existential variable), we show that chase termination is equivalent to boundedness and so it implies FO-rewritability. None of these properties hold for the semi-oblivious chase.

Recognizing if a set of existential rules is bounded is undecidable already for datalog [Hillebrand *et al.*, 1995]. However, we show the decidability of the problem for major classes of existential rules as direct corollaries of our characterizations and existing results from the literature. Precisely, the problem is PSpace-complete for linear and sticky rules and in 2Exptime for guarded rules. Finally, we consider the k*-boundedness* problem (i.e., whether the chase terminates in k steps on all instances), which was recently proven decidable for several chase variants, including those investigated here [Delivorias *et al.*, 2018]. We show that deciding if a ruleset is k-bounded is in 2Exptime for the breadth-first (semi-) oblivious chase and co-NExptime-complete for datalog.

Proofs omitted due to space limitations are detailed in the appendix.

2 Preliminary Definitions

We consider a relational vocabulary $\mathcal{V} = (\mathcal{P}, \mathcal{C})$ constituted of a finite set of predicates $\mathcal{P}$ and a finite set of constants $\mathcal{C}$. A term v is a constant of $\mathcal{C}$ or a variable. An atom is of the form $p(v_1 \dots v_k)$ where p is a predicate of arity k and the v_i are terms. We denote by $\mathsf{terms}()$ the set of its terms and extend the notation to sets of atoms. An *embedding* φ from a set of atoms A to a set of atoms A' is a substitution of $\mathsf{terms}(A)$ with $\mathsf{terms}(A')$ such that $\varphi(A) \subseteq A'$. A *homomorphism* is an embedding which is the identity on constants.

An *instance* I is a conjunction of atoms on constants and (globally) existentially quantified variables. It is finite unless otherwise specified. Throughout this paper, we see an instance I as the set of its atoms and call *fact* any atom f that belongs to this set. Given a finite set $\mathcal{P}$ of predicates, the *critical instance* I_a is composed of all facts built on $\mathcal{P}$ and special constant a. Any instance I on $\mathcal{P}$ can be embedded into I_a.

An *existential rule* σ is a closed formula $\forall \bar{x}\bar{y}(B[\bar{x}, \bar{y}] \rightarrow \exists \bar{z} H[\bar{x}, \bar{z}]$ where B and H are sets of atoms built on variables called the *body* and the *head* of the rule, also denoted by $\mathsf{body}(\sigma)$ and $\mathsf{head}(\sigma)$ respectively. The set of variables $\bar{x}$ shared by B and H is called the *frontier* of the rule and is denoted by $\mathsf{fr}(\sigma)$. The set of variables $\bar{z}$ that belong to H only are called *existential variables* and are denoted by $\mathsf{ex}(\sigma)$. Universal quantifiers will often be omitted in the remainder of the paper. A rule such that $\mathsf{ex}(\sigma) = \emptyset$ is called *datalog*. A rule where all head atoms contain at least one existential variable is called *fully-existential* and denoted by *FE-rule*. We say that a rule σ is applicable on an instance I if there is a homomorphism π from $\mathsf{body}(\sigma)$ to I and call the pair (σ, π) a *trigger* of I. Given a trigger (σ, π), we denote by $\pi_{|\mathsf{fr}(\sigma)} \subseteq \pi$ the restriction of π to $\mathsf{fr}(\sigma)$.

A *knowledge base* (KB) is a pair (I, Σ) where I is an instance and Σ a set of existential rules. The chase is a fundamental tool for computing logical consequences from a KB since, when it terminates, it computes a universal model of the KB, i.e., a model that maps by homomorphism to any other model of the KB (with a model being seen here as an instance). In this work, we focus our attention on the breadth-first oblivious (o-chase) and semi-oblivious (so-chase) variants. As discussed in Section 3.3, the breadth-first behavior is particularly interesting when studying boundedness.

Definition 1. *Let* (I, Σ) *be a knowledge base and* $\star \in \{\mathsf{o}, \mathsf{so}\}$ *a chase variant. Then, the* breadth-first $\star$-chase *is defined as follows:* $\star\text{-}\mathsf{chase}^0(I, \Sigma) = I$ *and for all saturation rank* $i \geq 0$

$$\star\text{-}\mathsf{chase}^{i+1}(I, \Sigma) = \star\text{-}\mathsf{chase}^{i}(I, \Sigma) \,\cup \bigcup_{(\sigma,\pi)} \pi_\star(\mathsf{head}(\sigma))$$

where (σ, π) *is any trigger of* $\star\text{-}\mathsf{chase}^i(I, \Sigma)$ *and* $\pi_\star \supseteq \pi$ *a substitution that replaces each existential variable* $z \in \mathsf{ex}(\sigma)$ *with a fresh variable named as follows:*

- $\pi_{\mathsf{o}}(z) = z_{(\sigma,\pi)}$
- $\pi_{\mathsf{so}}(z) = z_{(\sigma,\pi_{|\mathsf{fr}(\sigma)})}$

Then, we define $\star\text{-}\mathsf{chase}(I, \Sigma) = \bigcup_{i\geq 0}^{\infty} \star\text{-}\mathsf{chase}^i(I, \Sigma)$. *The* $\star$*-chase* terminates *on* (I, Σ) *if there is a rank* k *with* $\star\text{-}\mathsf{chase}(I, \Sigma) = \star\text{-}\mathsf{chase}^k(I, \Sigma)$.

Note that for the o-chase fresh variables are named by the trigger from which they have been generated. Instead, for the so-chase the naming only depends on the frontier-restriction of the homomorphism of the trigger. This means that any two triggers having the same rule and agreeing on the image of its frontier variables produce *equal* results, hence only one of them is actually considered by the so-chase. The so-chase is very close to the *Skolem chase*, which relies on a skolemisation of the rules: first, each rule σ is transformed by replacing each occurrence of an existential variable z with a functional term $f_z^\sigma(\mathsf{fr}(\sigma))$ on the frontier of σ; then the o-chase is run on the skolemised rules. At each saturation rank, the Skolem chase produces a result isomorphic to that of the so-chase (up to the renaming of each Skolem term by the corresponding fresh variable), hence the forthcoming results on the so-chase also hold for the Skolem chase.

Example 1. *Consider the rule* $\sigma = p(x, y) \rightarrow \exists z\, p(x, z)$. *Then* $\mathsf{o}\text{-}\mathsf{chase}(I, \Sigma)$ *with* $I = \{p(a, b)\}$ *and* $\Sigma = \{\sigma\}$ *is infinite - as the chase does not terminate. The atom* $p(a, z_{(\sigma,\pi_1)})$ *with* $\pi_1 = \{x \mapsto a, y \mapsto b\}$ *is first inferred, then* $p(a, z_{(\sigma,\pi_2)})$ *with* $\pi_2 = \{x \mapsto a, y \mapsto z_{(\sigma,\pi_1)}\}$*, and so on. Here, each rule application enables a new trigger. In contrast,* $\mathsf{so}\text{-}\mathsf{chase}(I, \Sigma)$ *is finite, in that only the first rule application will be performed, producing* $p(a, z_{(\sigma,\{x\mapsto a\})})$*, since all triggers map* *the frontier variable x to a. For the Skolem chase, σ is rewritten as $\sigma' = p(x,y) \rightarrow p(x, f_z^\sigma(x))$. The first rule application according to trigger (σ', π_1) produces $p(a, f_z^\sigma(a))$, then the chase halts as the same atom is produced by the next trigger.*

Definition 2. *The* rank *of a fact $f \in \star\text{-chase}(I,\Sigma)$, denoted by $\mathsf{rank}(f)$, is 0 if $f \in I$ and $1 + \max\{\mathsf{rank}(f') | f' \in \pi(\mathsf{body}(\sigma))\}$ if f is produced by the trigger (σ, π). This definition is naturally extended to terms and sets of facts. The* rank *of $\star\text{-chase}(I,\Sigma)$ is the smallest k such that $\star\text{-chase}(I,\Sigma) = \star\text{-chase}^k(I,\Sigma)$ if $\star\text{-chase}(I,\Sigma)$ terminates, and it is infinite otherwise.*

Note that for the breadth-first chases we consider the above definition implies that $\mathsf{rank}(f)$ is the smallest k such that $f \in \star\text{-chase}^k(I,\Sigma) \setminus \star\text{-chase}^{k-1}(I,\Sigma)$.

An *FO-query* $\phi(x_1, ..., x_n)$ is a (function free) first-order formula whose free variables (called answer variables) are exactly $\{x_1, ..., x_n\}$. A *conjunctive query* (CQ) is an FO-query which is an existentially quantified conjunction of atoms. An *atomic* query is a CQ with a single atom. A *full-atomic* query is an atomic query where all terms are free variables. A query is called *Boolean* if it does not have any free variable. As for instances, it will be handful to see CQs as sets of atoms, of course by distinguishing the answer variables. A *union of conjunctive queries* (UCQ) $\mathcal{Q}$ is a disjunction of CQs with the same free variables, also seen as a set of CQs.

A tuple of constants $(a_1, ..., a_n) \in \mathcal{C}^n$ is an *answer* to a CQ $Q(x_1, ..., x_n)$ on an instance I if there is a homomorphism h from Q to I such that $h(x_i) = a_i$ for $1 \leq i \leq n$. Equivalently, $I \models Q[x_i \mapsto a_i]$, where $\models$ denotes the classical logical consequence and $Q[x_i \mapsto a_i]$ is the Boolean query obtained from Q substituting each x_i with a_i. A tuple of constants $(a_1, ..., a_n) \in \mathcal{C}^n$ is a *certain answer* to Q on a KB (I, Σ) if $I, \Sigma \models Q[x_i \mapsto a_i]$. This is equivalent to the existence of a saturation rank k such that $\star\text{-chase}^k(I,\Sigma) \models Q[x_i \mapsto a_i]$. In other words, the certain answers to Q on (I, Σ) are exactly its answers on the possibly infinite instance $\star\text{-chase}(I,\Sigma)$. The set of (certain) answers to a UCQ $\mathcal{Q}$ is the union of the sets of (certain) answers to the CQs it contains.

2.1 Termination vs Boundedness

To begin our study, we need to present the relationships between chase termination and boundedness. Let $\star \in \{\mathsf{o}, \mathsf{so}\}$ be a chase variant, the $\star$-chase termination class, denoted by $\mathsf{CT}^\star$, contains all rulesets Σ such that $\star\text{-chase}(I,\Sigma)$ terminates for all instances I. The $\star$-boundedness class, denoted by $\mathsf{BN}^\star$, contains all bounded rulesets Σ, i.e., for which there exists an integer k such that $\star\text{-chase}^k(I,\Sigma) = \star\text{-chase}(I,\Sigma)$ for all instances I . Obviously, $\mathsf{BN}^\star \subset \mathsf{CT}^\star$.

Example 2. *Let $\sigma_1 = p(x,y) \wedge p(y,z) \rightarrow p(x,z)$ and $\sigma_2 = p(x,y) \wedge p(w,z) \rightarrow p(x,z)$. Because both rules are datalog, $\{\sigma_1\} \in \mathsf{CT}^\star$ and $\{\sigma_2\} \in \mathsf{CT}^\star$. However, $\Sigma = \{\sigma_1\} \notin \mathsf{BN}^\star$, since the rank of $\star\text{-chase}(I,\Sigma)$ depends on I. In contrast, $\{\sigma_2\} \in \mathsf{BN}^\star$ and the bound is $k = 1$. Similarly, $\{\sigma_1, \sigma_2\} \in \mathsf{BN}^\star$. Indeed, σ_2 produces at the first rank all atoms that can be produced by σ_1 at later ranks.*

To get a better understanding of boundedness, it will be useful to decompose each rule of a set thereby distinguishing between its "datalog part" and its "existential part". For instance, a rule of the form $p(x,y) \rightarrow \exists z\, p(x,z) \wedge q(x)$ can be decomposed into a datalog rule $p(x,y) \rightarrow q(x)$ and an FE-rule $p(x,y) \rightarrow \exists z\, p(x,z)$. Let σ be any existential rule of the form $B \rightarrow H_F \wedge H_D$ where B is the set of body atoms, H_F is the set of head atoms with at least one existential variable and H_D are the remaining head atoms. The datalog-fully existential decomposition of σ, denoted by $\mathsf{DF}(\sigma)$, returns a set made of the FE-rule $B \rightarrow H_F$ together with a (single head) datalog rule of the form $B \rightarrow H_D^i$, for each $H_D^i \in H_D$. The definition is then extended to sets $\mathsf{DF}(\Sigma) = \bigcup_{\sigma \in \Sigma} \mathsf{DF}(\sigma)$. This decomposition preserves boundedness and termination of the oblivious chase.[2]

Proposition 1. *$\Sigma \in \mathsf{CT}^\mathsf{o}$ iff $\mathsf{DF}(\Sigma) \in \mathsf{CT}^\mathsf{o}$ and $\Sigma \in \mathsf{BN}^\mathsf{o}$ iff $\mathsf{DF}(\Sigma) \in \mathsf{BN}^\mathsf{o}$.*

3 Upper Bounding the Chase Depth

Our approach consists of defining a notion of *existential depth* for facts, proper to each chase, which is finite on a given instance if and only if the chase terminates on that instance. Then we show that for each chase, the existential depth of all facts produced by the chase for a given ruleset are bounded by those of the critical instance. This means that whenever the chase terminates on the critical instance there is an upper bound to the existential depth of the facts, for all instances. In the next section, with these results in hand, we use FO-rewritability to bound the *rank* at which any fact of a certain existential depth will be inferred. This will give us a characterization of boundedness for the oblivious and so-chase in terms of FO-rewritability and chase termination.

3.1 The Oblivious Case

Intuitively, the notion of existential depth of a term measures the number of fresh variable generation steps that led to the creation of this term.

Definition 3. *The* existential depth *(or simply* depth*) of a term v that belongs to $\mathsf{o}\text{-chase}(I,\Sigma)$ is*

$$\mathsf{depth}_\exists(v) = \begin{cases} 0 & \text{if } v \in \mathsf{terms}(I) \\ 1 + \max\{\ \mathsf{depth}_\exists(v_B)\ \} & \text{otherwise} \end{cases}$$

where v_B is any term in $\pi(\mathsf{body}(\sigma))$ used by a trigger (σ, π) which generates v. The existential depth *of a fact f is the maximum existential depth of its terms. The* existential depth *of $\mathsf{o}\text{-chase}(I,\Sigma)$ is the maximum existential depth of its facts if it is finite and is infinite otherwise.*

To illustrate the definition, consider Example 1. The existential depth of terms in $\mathsf{o}\text{-chase}(I, \{\sigma\})$ is unbounded, which is in line with the non-termination of the o-chase on $(I, \{\sigma\})$. The rule σ_1 in Example 2. shows the difference between rank and existential depth. For any I, the existential depth

[2]This is not true for the so-chase. For instance, for $\Sigma = \{\sigma = p(x,y) \rightarrow \exists z\, p(x,z) \wedge q(x,y)\}$ and $I = \{p(a,b)\}$, $\mathsf{so}\text{-chase}(\Sigma, I)$ is infinite, while $\mathsf{DF}(\Sigma)$ is so-bounded. This is due to the fact that σ has frontier $\{x,y\}$, while the FE-rule $p(x,y) \rightarrow \exists z\, p(x,z)$ in $\mathsf{DF}(\sigma)$ has frontier $\{x\}$. We correct here a wrong claim in Proposition 1 of IJCAI's paper, which has no incidence on the paper's results.

of terms (hence facts) is 0 because σ_1 is datalog, however their rank depends on I. More generally, for any term v and fact f in $\mathsf{o\text{-}chase}(I, \Sigma)$ it holds that $\mathsf{depth}_\exists(v) \leq \mathsf{rank}(v)$ and $\mathsf{depth}_\exists(f) \leq \mathsf{rank}(f)$. Hence, if $\mathsf{o\text{-}chase}(I, \Sigma)$ terminates, its existential depth is finite. Reciprocally, when the existential depth of $\mathsf{o\text{-}chase}(I, \Sigma)$ is finite, so it is the number of its terms, and $\mathsf{o\text{-}chase}(I, \Sigma)$ terminates. We point out that when dealing with sets of *FE-rules* the notions of rank and existential depth coincide, as illustrated by Example 1.

Proposition 2. *If Σ is a set of FE-rules then, for all instance I and term v in* $\mathsf{o\text{-}chase}(I, \Sigma)$*, holds that* $\mathsf{depth}_\exists(v) = \mathsf{rank}(v)$.

It should be clear that, for a given ruleset, the o-chase may have unbounded rank even when it terminates on all instances (see for instance Example 2). Nevertheless, when a ruleset is in CT°, our goal is to show that there exists a bound on the *existential depth* of its terms, which holds for all instances. Aiming at this, we present a lemma stating that existential depth of terms are preserved by embeddings.

Lemma 3. *For any embedding φ from I to I' and any $i \geq 0$, there exists an embedding $\varphi' \supseteq \varphi$ from* $\mathsf{o\text{-}chase}^i(I, \Sigma)$ *to* $\mathsf{o\text{-}chase}^i(I', \Sigma)$ *which preserves the existential depth of terms, i.e., for every term v in* $\mathsf{o\text{-}chase}(I, \Sigma)$ *it holds that* $\mathsf{depth}_\exists(v) = \mathsf{depth}_\exists(\varphi'(v))$.

It is well-known that the o-chase terminates on all instances if and only if it terminates on the critical instance [Marnette, 2009]. We leverage this property to compute a bound on the existential depth under chase termination.

Theorem 4. *When $\Sigma \in \mathsf{CT}^\circ$ there exists a constant k_d such that for every instance I, the existential depth of a term in* $\mathsf{o\text{-}chase}(I, \Sigma)$ *is bounded by k_d.*

Proof. Because $\Sigma \in \mathsf{CT}^\circ$, the o-chase terminates on the critical instance I_a. Let k_d be the largest rank such that $\mathsf{terms}(\mathsf{o\text{-}chase}^{k_d}(I_a, \Sigma)) \setminus \mathsf{terms}(\mathsf{o\text{-}chase}^{k_d - 1}(I_a, \Sigma)) \neq \emptyset$. Every instance I can be embedded into I_a. By Lemma 3 the existential depth of the terms in $\mathsf{o\text{-}chase}(I, \Sigma)$ is bounded by that of $\mathsf{o\text{-}chase}(I_a, \Sigma)$, which is in turn bounded by k_d. □

Chase termination is a necessary condition for boundedness as it bounds the existential depths of the variables generated by the chase - but not the rank (see the datalog case). Interestingly, for FE-rules, chase termination also becomes a sufficient condition for boundedness, because the notion of rank and existential depth coincide (Proposition 2).

Corollary 5. *For Σ a set of FE-rules, $\Sigma \in \mathsf{CT}^\circ$ iff $\Sigma \in \mathsf{BN}^\circ$.*

For general existential rules, we will later show that when a restricted form of FO-rewritability holds, one can also provide a bound to the rank of the o-chase (Theorem 14).

3.2 The Semi-Oblivious Case

When applied to the so-chase, the previous notion of existential depth is not preserved by embedding, which hinders the possibility of using the critical instance to bound the existential depth of terms. As illustrated below, this is due to the fact that the so-chase makes equal the result of two distinct triggers agreeing on a rule frontier.

Example 3. *Consider $I = \{p(a, b)\}$, $I' = I \cup \{r(a, b)\}$ and $\Sigma = \{\sigma_1 : p(x, y) \to \exists z\, r(z, y) \;\; \sigma_2 : r(x, y) \to \exists z\, s(y, z)\}$. Then,* $\mathsf{so\text{-}chase}^2(I, \Sigma) = I \cup \{\, r(z_{(\sigma_1,\pi)}, b) \cup s(b, z_{(\sigma_2,\pi)}) \,\}$ *with $\pi = \{y \mapsto b\}$. Also,* $\mathsf{so\text{-}chase}^2(I, \Sigma) \subseteq \mathsf{so\text{-}chase}^1(I', \Sigma)$ *because all triggers applied by the chase from I are already applicable on I'. The application of σ_2 on $r(a, b)$ and $r(z_{(\sigma_1,\pi)}, b)$ gives equal results, hence* $\mathsf{so\text{-}chase}^1(I', \Sigma) = \mathsf{so\text{-}chase}^2(I', \Sigma)$*. In the embeddings from* $\mathsf{so\text{-}chase}^2(I, \Sigma)$ *to* $\mathsf{so\text{-}chase}^2(I', \Sigma)$*, $z_{(\sigma_2,\pi)}$ is mapped to itself, but both occurrences have different existential depth (resp. 2 and 1).*

It is therefore natural to turn to the following notion of depth, which accounts for frontier terms only.

Definition 4. *The* frontier existential depth *(or simply frontier depth) of a term v that belongs to* $\mathsf{so\text{-}chase}(I, \Sigma)$ *is*

$$\mathsf{depth}^{\mathsf{fr}}_\exists(v) = \begin{cases} 0 & \text{if } v \in \mathsf{terms}(I) \\ 1 & \text{if } \mathsf{fr}(\sigma) = \emptyset \\ 1 + \max\{\, \mathsf{depth}^{\mathsf{fr}}_\exists(v_B) \,\} & \text{otherwise} \end{cases}$$

where v_B is any term in $\pi(\mathsf{fr}(\sigma))$ used by a trigger (σ, π) which generates v. Accordingly, the frontier depth *of a fact f is the maximum frontier depth of its terms. The frontier depth of* $\mathsf{so\text{-}chase}(I, \Sigma)$ *is defined as the maximum frontier depth of its facts if it is finite and is infinite otherwise.*

Note that frontier depth coincides with the (usual) depth of terms generated by the Skolem chase.

Clearly, $\mathsf{depth}^{\mathsf{fr}}_\exists(v) \leq \mathsf{depth}_\exists(v)$ for all v in $\mathsf{o\text{-}chase}(I, \Sigma)$. The following example illustrates the difference between the two notions of (existential) depth.

Example 4. *Let $\Sigma = \{\sigma = p(x, y, u) \to \exists z\; p(y, x, z)\}$. Starting from $I = \{p(a, b, c)\}$, the o-chase generates an infinite number of fresh variables v with increasing* $\mathsf{depth}_\exists(v)$*. The rank of the so-chase is instead 2 and for each fresh variable v,* $\mathsf{depth}^{\mathsf{fr}}_\exists(v) = 1$ *as all triggers map* $\mathsf{fr}(\sigma)$ *to* $\mathsf{terms}(I)$.

It is worth noting that not only the oblivious notion of exitential depth is not effective for studying the so-chase, but also that the frontier depth is not well characterizing the behavior of the o-chase either. The crux is that the finiteness of the frontier depth cannot be related with the termination of the o-chase, as illustrated by Example 4. Using such a notion to study the o-chase would impede us, for instance, to establish Corollary 5, which relies on the fact that rank and existential depth coincide for the oblivious-chase (Property 2).

We are now ready to show that the frontier depth is preserved by embeddings. The next lemma and theorem are the counter-parts of Lemma 3 and Theorem 4 for the so-chase.

Lemma 6. *For any embedding φ from I to I' and any $i \geq 0$, there exists an embedding $\varphi' \supseteq \varphi$ from* $\mathsf{so\text{-}chase}^i(I, \Sigma)$ *to* $\mathsf{so\text{-}chase}^i(I', \Sigma)$ *which preserves the frontier depth of terms.*

Theorem 7. *When $\Sigma \in \mathsf{CT}^{\mathsf{so}}$ there exists a constant k_d such that for every instance I, the frontier depth of a term in* $\mathsf{so\text{-}chase}(I, \Sigma)$ *is bounded by k_d.*

3.3 On the Interest of the Breadth-First Chase

We conclude this section with some remarks on the interest of studying boundedness for breadth-first chases. We assume that the reader is familiar with the notion of chase sequence.[3] We define the rank of a chase sequence on (I, Σ) as the maximal rank of its facts if it is finite, and infinite otherwise.

For the (semi-)oblivious chase, it is well-known that there is a terminating chase sequence for (I, Σ) if and only if all chase sequences for (I, Σ) terminate. However, not all terminating chase sequences have the same rank, and the minimal rank is obtained with breadth-first sequences [Delivorias *et al.*, 2018]. This makes the notion of boundedness we consider equivalent to studying whether there exists a bound such that, for all instance, *there exists* a terminating chasing sequence whose rank is within the bound. Hence, it characterizes the fact that the chase can indeed terminate within that bound, if a strategy ensuring a minimal sequence rank is followed. It is therefore natural to consider breadth-first chases which achieve this property, like the (semi-)oblivious chase. Example 2 illustrates this concept and shows that, already for datalog, the rank of some chase sequences may be not bounded, while the rank of all breadth-first sequences is bounded. This happens for instance if all applications of the transitivity rule σ_2 are performed before the rule σ_1.

In the special case of FE-rules, it is not hard to see that all oblivious chase sequences for (I, Σ) have the same rank. However, this does not hold for the semi-oblivious chase. Below, a variation of Example 2, where some dummy variables are introduced, illustrates this point.

Example 5. *Let* $\Sigma = \{\sigma_1, \sigma_2\}$*, with* $\sigma_1 = p(x, y, t) \wedge p(y, z, u) \rightarrow \exists v\, p(x, z, v)$ *and* $\sigma_2 = p(x, y, t) \wedge p(w, z, u) \rightarrow \exists v\ p(x, z, v)$*. The rank of* $\mathsf{so\text{-}chase}(I, \Sigma)$ *is bounded by 2 for any I, while again performing all applications of σ_2 before σ_1 gives derivations of different ranks.*

4 The Impact of First Order Rewritability

We now turn our attention to FO-rewritability and show that it yields a bound on the rank of specific (sets of) facts that share terms with the initial instance I. For the o-chase, we bound the rank of facts that have all their terms in I. For the so-chase, we consider triggers that map a rule frontier to terms of I: we do not bound the rank of facts that allow to fire such triggers, but we show that for each such trigger $t = (\sigma, \pi)$, there is a trigger $t' = (\sigma, \pi')$ that agrees with t on the mapping of $\mathsf{fr}(\sigma)$ and that is fired at a bounded rank. In Section 5, we will leverage these results to show that FO-rewritability yields a bound on the rank of all facts with a certain existential depth. For the o-chase, a restricted version of FO-rewritability is sufficient to get these properties.

We say that a pair (Q, Σ) is *FO-rewritable* (resp. UCQ-rewritable) if there is an FO-query (resp. a UCQ) $\mathcal{Q}$ such that, for all I, the certain answers to Q on (I, Σ) are exactly the answers to $\mathcal{Q}$ on I. It is known that FO-rewritability is equivalent to UCQ-rewritability.[4] A set of rules Σ is *FO-rewritable* (or equivalently, UCQ-rewritable) if (Q, Σ) is FO-rewritable for every CQ Q. We denote by FO-R the class of FO-rewritable rulesets. We will also consider specific classes of CQs. Given a class of CQs $\mathcal{C}$, we say that a ruleset Σ is FO-rewritable with respect to $\mathcal{C}$ if (Q, Σ) is FO-rewritable for all $Q \in \mathcal{C}$. We denote by $\mathsf{FO\text{-}R}^{\mathcal{C}}$ the corresponding class. We first point out that FO-rewritability with respect to full-atomic queries, denoted by $\mathsf{FO\text{-}R}^{\mathsf{AF}}$, is a strictly weaker property than FO-rewritability.

Proposition 8. $\mathsf{FO\text{-}R}^{\mathsf{AF}} \supset \mathsf{FO\text{-}R}$

Proof. The inclusion holds by definition, and to see that it is strict consider $\Sigma = \{\sigma = p(x, x_1), p(x_1, x_2), p(x_2, z) \rightarrow \exists y\ \ p(x, y), p(y, z)\}$. Σ is not FO-rewritable as for the Boolean query $Q = \{p(a, u), p(u, b)\}$, where a and b are constants, (Q, Σ) is not FO-rewritable (we would need an infinite union of Boolean CQs of the form $\{p(a, u_0), ...p(u_{i-1}, u_i), p(u_i, b)\}$, none of these queries being contained in another). However, $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{AF}}$ as (Q, Σ) is FO-rewritable for any $Q \in AF$. Indeed, σ cannot bring any answer to such query (in more technical terms, an existential variable of σ cannot be unified with an answer variable). ☐

Note also that since full-atomic queries have only answer variables, they cannot be rewritten by means of *FE-rules*. Thus, every set of FE-rules is trivially in $\mathsf{FO\text{-}R}^{\mathsf{AF}}$. More interestingly, to check if $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{AF}}$ one can restrict the full-atomic queries of interest to those corresponding to the heads of the datalog rules yielded by the DF-decomposition of Σ.

Proposition 9. *Let Σ be a ruleset and HD_Σ be the full-atomic queries given by heads of the datalog rules in $\mathsf{DF}(\Sigma)$. Then,* $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{AF}}$ *if and only if* $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{HD}_\Sigma}$.

The following lemma upper bounds the rank of all facts with terms in I for sets of rules enjoying FO-rewritability on full-atomic queries.

Lemma 10. *If* $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{AF}}$ *there is a constant k_{AF} such that, for any instance I and fact f such that* $\mathsf{terms}(f) \subseteq \mathsf{terms}(I)$*, when* $f \in \mathsf{o\text{-}chase}(I, \Sigma)$ *it holds that* $\mathsf{rank}(f) \leq k_{\mathrm{AF}}$.

Proof. The number of (non-isomorphic) full-atomic queries to be considered is finite, as for Proposition 9. We take for k_{AF} the maximal number of breadth-first rewriting steps necessary to obtain a UCQ-rewriting of a full-atomic query (we refer here to the breadth-first rewriting based on aggregated piece-unifiers, see [König *et al.*, 2013]). ☐

The previous lemma also holds for the so-chase, however we want to derive a bound on the rank of facts with a certain frontier depth, and for that full-atomic rewritability is not enough. To illustrate, consider $\Sigma = \{\sigma = p(x, y, u), p(y, z, v) \rightarrow \exists w\ p(x, z, w)\}$. Here $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{AF}}$ (the only rewriting of a full-atomic query is the query itself because of the existential variable w). For any instance I, the frontier depth of facts in the so-chase is bounded by 1, however there is no bound on their rank (although the so-chase

[3] A chase sequence is any sequence of triggers satisfying the applicability criterion of the chase variant. For the oblivious chase, the same trigger should not be applied twice. For the semi-oblivious chase a trigger is not applied if a trigger for the same rule assigning the same image for the frontier variables has been applied before.

[4] It follows from the (Finite) Homomorphism preservation theorem, a classical result in model theory [Rossman, 2008].

terminates). Therefore, we give a different property for the so-chase, which requires the power of FO-rewritability.

Lemma 11. *If $\Sigma \in \mathsf{FO\text{-}R}$ there is a constant k_{FO} such that, for any instance I and any trigger (σ,π) from $\mathsf{so\text{-}chase}(I,\Sigma)$ with $\pi(\mathsf{fr}(\sigma)) \subseteq \mathsf{terms}(I)$, there is also a trigger (σ,π') from $\mathsf{so\text{-}chase}(I,\Sigma)$ such that $\pi'_{|\mathsf{fr}(\sigma)} = \pi_{|\mathsf{fr}(\sigma)}$ and $\mathsf{rank}(f) \leq k_{\mathrm{FO}}$ for all $f \in \pi'(\mathsf{body}(\sigma))$.*

Proof. Similar to the proof of Lemma 10 but considering CQs of the form $Q_{\mathsf{body}(\sigma)}$ whose atoms correspond to the atoms of $\mathsf{body}(\sigma)$, for $\sigma \in \Sigma$, and all variables are existentially quantified except for those in $\mathsf{fr}(\sigma)$. The number of such queries is bounded by the cardinal of Σ. We take for k_{FO} the maximal number of breadth-first rewriting steps necessary to obtain a UCQ-rewriting from any $Q_{\mathsf{body}(\sigma)}$ query. The proof actually shows that FO-rewritability with respect to rule body queries is sufficient to derive the lemma. □

5 Boundedness: Linking Depth and Rank

We can finally establish a connection between the rank and depth of a fact when the chase is run on FO-rewritable sets of rules. This will immediately lead us to a characterization of boundedness for the oblivious and semi-oblivious chases.

Theorem 12. *If $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{AF}}$ then for all instance I and fact $f \in \mathsf{o\text{-}chase}(I,\Sigma)$ we have that $\mathsf{rank}(f) \leq \mathsf{depth}_{\exists}(f) \times (k_{\mathrm{AF}}+1)+k_{\mathrm{AF}}$ with k_{AF} the bound provided by Lemma 10.*

Theorem 13. *If $\Sigma \in \mathsf{FO\text{-}R}$ then for all instance I and fact $f \in \mathsf{so\text{-}chase}(I,\Sigma)$ we have that $\mathsf{rank}(f) \leq \mathsf{depth}^{\mathsf{fr}}_{\exists}(f) \times (k_{\mathrm{FO}}+1)+k_{\mathrm{FO}}$ with k_{FO} the bound provided by Lemma 11.*

For the o-chase, boundedness is exactly termination and FO-rewritability on full-atomic queries. Furthermore, for rulesets in CT^{o}, the notions of FO-R and $\mathsf{FO\text{-}R}^{\mathsf{AF}}$ coincide.

Theorem 14. $\mathsf{FO\text{-}R}^{\mathsf{AF}} \cap \mathsf{CT}^{\mathsf{o}} = \mathsf{BN}^{\mathsf{o}} = \mathsf{FO\text{-}R} \cap \mathsf{CT}^{\mathsf{o}}$

Proof. We start by showing that $\mathsf{BN}^{\mathsf{o}} \subseteq \mathsf{FO\text{-}R} \cap \mathsf{CT}^{\mathsf{o}}$. By definition $\mathsf{BN}^{\mathsf{o}} \subseteq \mathsf{CT}^{\mathsf{o}}$. Then, $\mathsf{BN}^{\mathsf{o}} \subseteq \mathsf{FO\text{-}R}$ follows from the equivalence between FO-R and the bounded-depth derivation property [Gottlob *et al.*, 2014]. Moreover, by Proposition 8 we have $\mathsf{BN}^{\mathsf{o}} \subseteq \mathsf{FO\text{-}R}^{\mathsf{AF}} \cap \mathsf{CT}^{\mathsf{o}}$. To conclude the proof, by Theorem 4 and 12 we have that $\mathsf{FO\text{-}R}^{\mathsf{AF}} \cap \mathsf{CT}^{\mathsf{o}} \subseteq \mathsf{BN}^{\mathsf{o}}$ and again by Proposition 8 follows $\mathsf{FO\text{-}R} \cap \mathsf{CT}^{\mathsf{o}} \subseteq \mathsf{BN}^{\mathsf{o}}$. □

For the so-chase, boundedness can be characterized again as termination and FO-rewritability by Theorem 7 and 13.

Theorem 15. $\mathsf{BN}^{\mathsf{so}} = \mathsf{FO\text{-}R} \cap \mathsf{CT}^{\mathsf{so}}$

Summing up, we have the following differences between boundedness for o-chase and so-chase. o-chase-boundedness requires $i)$ o-chase termination and full-atomic-rewritability and $ii)$ is equivalent to o-chase termination for FE-rules. Intuitively, when a set of rules Σ is decomposed into $\mathsf{DF}(\Sigma)$, the fully-existential part may cause non-termination of the o-chase, while the datalog part may cause non-FO-rewritability. Furthermore, the fully-atomic queries possibly leading to infinite rewritings in this case correspond to the heads of the datalog rules. Note however that this restricted form of FO-rewritability has still to be verified with respect to the whole set of rules. In contrast, so-chase-boundedness $i)$ requires a stronger form of FO-rewritability and $ii)$ FE-rules do not behave differently from general existential rules for this chase. Intuitively, for the so-chase, any existential rule (even an FE-rule) has an "underlying" datalog rule. This is illustrated by the following transformation. To each rule σ in Σ we assign a special predicate p_σ of arity $|\mathsf{fr}(\sigma)|$. $\Psi(\Sigma)$ is obtained from Σ by replacing each rule $\sigma = B \to H$ with two rules: a datalog rule $B \to p_\sigma(\mathsf{fr}(\sigma))$ and a rule $p_\sigma(\mathsf{fr}(\sigma)) \to H$. It can be shown that $\Sigma \in \mathsf{CT}^{\mathsf{so}}$ iff $\Psi(\Sigma) \in \mathsf{CT}^{\mathsf{o}}$ and that $\Sigma \in \mathsf{BN}^{\mathsf{so}}$ iff $\Psi(\Sigma) \in \mathsf{BN}^{\mathsf{o}}$. This may also provide an alternative path to study so-chase boundedness by reducing it to o-chase boundedness.

6 Decidability and Complexity

From the undecidability of (uniform) boundedness of datalog [Hillebrand *et al.*, 1995], we immediately obtain the undecidability of membership to BN^{o} and $\mathsf{BN}^{\mathsf{so}}$. A notable class of datalog rules with decidable boundedness (more precisely in linear time) is chain datalog [Guessarian and Peixoto, 1994]. We obtain that membership to $\mathsf{BN}^{\mathsf{so}}$, $\mathsf{CT}^{\mathsf{so}}$ and FO-R remains undecidable for *FE-rules*, while the decidability of membership to BN^{o}, hence to CT^{o}, is still open.[5]

Importantly, new decidability and complexity results about boundedness for specific existential rules studied in the literature can be obtained as direct corollaries of our results. This is in particular the case for classes known to be FO-rewritable.

Corollary 16. *For any class of existential rules $\mathcal{C} \in \mathsf{FO\text{-}R}$, it holds that: $\mathcal{C} \in \mathsf{BN}^{\mathsf{o}}$ iff $\mathcal{C} \in \mathsf{CT}^{\mathsf{o}}$, and $\mathcal{C} \in \mathsf{BN}^{\mathsf{so}}$ iff $\mathcal{C} \in \mathsf{CT}^{\mathsf{so}}$.*

This implies that membership to BN^{o} and $\mathsf{BN}^{\mathsf{so}}$ is PSpace-complete for the two main classes of FO-rewritable existential rules, namely *linear* and *sticky*. Indeed, deciding CT^{o} and $\mathsf{CT}^{\mathsf{so}}$ is PSpace-complete for both [Calautti *et al.*, 2015; Calautti and Pieris, 2019]. We also get an upper bound on the complexity of membership to BN^{o} and $\mathsf{BN}^{\mathsf{so}}$ for a major class of existential rules, namely *guarded*. This class is neither $\mathsf{CT}^{\mathsf{so}}$ nor FO-R. However, membership to CT^{o} and $\mathsf{CT}^{\mathsf{so}}$ for guarded rules is decidable in 2Exptime [Calautti *et al.*, 2015]. Then a careful reduction from [Barceló *et al.*, 2018] allows us to set the result. The paper shows that checking FO-rewritability for a single query under guarded rules is in 2Exptime. This suffices since by Lemma 10 and 11 we need to test only a polynomial number of queries.

We conclude by considering the k-boundedness problem, which asks whether the chase actually halts within k steps. The problem is decidable for the breadth-first (semi-)oblivious chase and any set of existential rules [Delivorias *et al.*, 2018]. Therefore, the k-boundedness question becomes interesting for dealing with fragments of existential rules where boundedness is undecidable. We study here the complexity of the following version of the problem. Given a ruleset Σ and a (unary encoded) integer k, does it hold that $\star\text{-}\mathsf{chase}^k(I,\Sigma) = \star\text{-}\mathsf{chase}(I,\Sigma)$ for all instance I?

Theorem 17. *Deciding k-boundedness is in 2Exptime for existential rules for the o-chase and so-chase; co-*

[5]See Proposition 18 in the Appendix.

NExptime-complete for datalog rules; in co-NExptime on FE-rules for the o-chase*.*

Proof. (Sketch) The upper bound results rely on the decidability arguments from [Delivorias *et al.*, 2018]. Co-NExptime-hardness for datalog is by reduction from the co-NExptime-hard inclusion problem of non-recursive Boolean datalog queries [Benedikt and Gottlob, 2010]. □

7 Outline and Perspectives

In this paper, we have characterized boundedness in terms of FO-rewritability and chase termination, for the oblivious and semi-oblivious chase variants. We conclude with a discussion on the extent of our results to more powerful chase variants (i.e., which terminate at least when the semi-oblivious chase terminate). Theorem 13 suggests that whenever $\Sigma \in \mathsf{FO\text{-}R}$ if any such chase generates only terms of bounded *frontier depth* on all instances, then Σ is bounded. We leave open the question to determine if for other chase variants, like the restricted and the core chases, boundedness is again the intersection of chase termination and FO-rewritability.

Acknowledgements. This work was supported by ANR projects CQFD (ANR-18-CE23-0003), DataCert (ANR-15-CE39-0009), DeLTA (ANR-16- CE40-0007) and the CNRS-Momentum project Managing-Data.

References

[Ajtai and Gurevich, 1994] Miklós Ajtai and Yuri Gurevich. Datalog vs first-order logic. *J. Comput. Syst. Sci.*, 49(3):562–588, 1994.

[Baget *et al.*, 2011] Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On rules with existential variables: Walking the decidability line. *Artif. Intell.*, 175(9-10):1620–1654, 2011.

[Barceló *et al.*, 2018] Pablo Barceló, Gerald Berger, Carsten Lutz, and Andreas Pieris. First-order rewritability of frontier-guarded ontology-mediated queries. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 1707–1713, 2018.

[Benedikt and Gottlob, 2010] Michael Benedikt and Georg Gottlob. The impact of virtual views on containment. *PVLDB*, 3(1):297–308, 2010.

[Bourhis *et al.*, 2019] Pierre Bourhis, Michel Leclère, Marie-Laure Mugnier, Sophie Tison, Federico Ulliana, and Lily Gallois. Oblivious and semi-oblivious boundedness for existential rules. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 1581–1587. ijcai.org, 2019.

[Calautti and Pieris, 2019] Marco Calautti and Andreas Pieris. Oblivious chase termination: The sticky case. In *22nd International Conference on Database Theory, ICDT 2019 (to appear)*, 2019.

[Calautti *et al.*, 2015] Marco Calautti, Georg Gottlob, and Andreas Pieris. Chase termination for guarded existential rules. In *Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, pages 91–103, 2015.

[Calì *et al.*, 2009a] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. Datalog extensions for tractable query answering over ontologies. In *Semantic Web Information Management - A Model-Based Perspective*, pages 249–279. 2009.

[Calì *et al.*, 2009b] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A General Datalog-Based Framework for Tractable Query Answering over Ontologies. In *Proceedings of the Twenty-Eigth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2009*, pages 77–86. ACM, 2009.

[Calì *et al.*, 2013] Andrea Calì, Georg Gottlob, and Michael Kifer. Taming the infinite chase: Query answering under expressive relational constraints. *J. Artif. Intell. Res.*, 48:115–174, 2013.

[Calvanese *et al.*, 2007] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. Autom. Reasoning*, 39(3):385–429, 2007.

[Carral *et al.*, 2017] David Carral, Irina Dragoste, and Markus Krötzsch. Detecting chase (non)termination for existential rules with disjunctions. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 922–928, 2017.

[Delivorias *et al.*, 2018] Stathis Delivorias, Michel Leclère, Marie-Laure Mugnier, and Federico Ulliana. On the k-boundedness for existential rules. In *International Joint Conference on Rules and Reasoning*, pages 48–64. Springer, 2018.

[Delivorias *et al.*, 2020] Stathis Delivorias, Michel Leclère, Marie-Laure Mugnier, and Federico Ulliana. Characterizing boundedness in chase variants. *CoRR*, abs/2004.10030, 2020.

[Fagin *et al.*, 2005] Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: semantics and query answering. *Theor. Comput. Sci.*, 336(1):89–124, 2005.

[Gallois, 2019] Lily Gallois. *Dialog between chase approach and string rewriting system approach*. Theses, Université de Lille, December 2019. In French.

[Gottlob *et al.*, 2014] Georg Gottlob, Stanislav Kikot, Roman Kontchakov, Vladimir Podolskii, Thomas Schwentick, and Michael Zakharyaschev. The price of query rewriting in ontology-based data access. *Artificial Intelligence*, 213:42–59, 2014.

[Grau *et al.*, 2013] Bernardo Cuenca Grau, Ian Horrocks, Markus Krötzsch, Clemens Kupke, Despoina Magka, Boris Motik, and Zhe Wang. Acyclicity notions for existential rules and their application to query answering in ontologies. *J. Artif. Intell. Res.*, 47:741–808, 2013.

[Guessarian and Peixoto, 1994] Irène Guessarian and Marcos Veloso Peixoto. About boundedness for some datalog and datalog$^{\text{neg}}$ programs. *J. Log. Comput.*, 4(4):375–403, 1994.

[Hernich and Schweikardt, 2007] André Hernich and Nicole Schweikardt. Cwa-solutions for data exchange settings with target dependencies. In *Proceedings of the Twenty-Sixth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 11-13, 2007, Beijing, China*, pages 113–122, 2007.

[Hillebrand *et al.*, 1995] Gerd G Hillebrand, Paris C Kanellakis, Harry G Mairson, and Moshe Y Vardi. Undecidable boundedness problems for datalog programs. *The Journal of logic programming*, 25(2):163–190, 1995.

[König *et al.*, 2013] Mélanie König, Michel Leclère, Marie-Laure Mugnier, and Michaël Thomazo. On the exploration of the query rewriting space with existential rules. In *Web Reasoning and Rule Systems - 7th International Conference, RR 2013, Mannheim, Germany, July 27-29, 2013. Proceedings*, pages 123–137, 2013.

[König *et al.*, 2015] Mélanie König, Michel Leclère, Marie-Laure Mugnier, and Michaël Thomazo. Sound, complete and minimal ucq-rewriting for existential rules. *Semantic Web*, 6(5):451–475, 2015.

[Marnette, 2009] Bruno Marnette. Generalized schema-mappings: from termination to tractability. In *Proceedings of the Twenty-Eigth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2009, June 19 - July 1, 2009, Providence, Rhode Island, USA*, pages 13–22, 2009.

[Rossman, 2008] Benjamin Rossman. Homomorphism preservation theorems. *J. ACM*, 55(3):15:1–15:53, 2008.

Appendix

This appendix contains the proofs that were omitted in the paper due to space limitation. Note that the proofs of Theorem 4 and Proposition 8 are provided in the paper, hence not recalled below.

Proof of Proposition 1 $\Sigma \in \mathsf{CT}^{\mathsf{o}}$ *iff* $\mathsf{DF}(\Sigma) \in \mathsf{CT}^{\mathsf{o}}$ *and* $\Sigma \in \mathsf{BN}^{\mathsf{o}}$ *iff* $\mathsf{DF}(\Sigma) \in \mathsf{BN}^{\mathsf{o}}$.

Proof. The proposition is immediate since, for any instance I and chase step i, $\mathsf{o\text{-}chase}^i(I, \Sigma) = \mathsf{o\text{-}chase}^i(I, \mathsf{DF}(\Sigma))$. □

Remarks. For the so-chase, only one direction holds true: if $\Sigma \in \mathsf{CT}^{\mathsf{so}}$ then $\mathsf{DF}(\Sigma) \in \mathsf{CT}^{\mathsf{so}}$ and if $\Sigma \in \mathsf{BN}^{\mathsf{so}}$ then $\mathsf{DF}(\Sigma) \in \mathsf{BN}^{\mathsf{so}}$. Note that the decomposition has no incidence on the FO-rewritability of Σ since Σ and $\mathsf{DF}(\Sigma)$ are logically equivalent.

Proof of Proposition 2 *If Σ is a set of FE-rules then, for all instance I and term v in* $\mathsf{o\text{-}chase}(I, \Sigma)$*, holds that* $\mathsf{depth}_\exists(v) = \mathsf{rank}(v)$.

Note that this proposition could also be stated for facts instead of terms.

Proof. By a straightforward induction on the rank of facts, we show that, for all $i \geq 0$ and fact f, if $\mathsf{rank}(f) = i$ then $\mathsf{depth}_\exists(f) = i$. The property obviously holds for $i = 0$. Let $i > 0$ and $\mathsf{rank}(f) = i$. By definition of rank, f was produced from at least one fact f' of rank $i - 1$. By induction hypothesis, $\mathsf{depth}_\exists(f') = i - 1$, hence, by definition of existential depth, f' contains a term t with $\mathsf{depth}_\exists(t) = i - 1$. Since all rules are FE-rules, f contains at least one fresh variable (null) v, and, by definition of existential depth, $\mathsf{depth}_\exists(v) = 1 + (i - 1) = i$. Hence, $\mathsf{depth}_\exists(f) = i$.

Now, let t be a term with rank i. If $i = 0$, t occurs in I and $\mathsf{depth}_\exists(t) = 0$. Otherwise, t has been generated in a fact f of rank i. Since $\mathsf{depth}_\exists(f) = \mathsf{rank}(f)$, $\mathsf{depth}_\exists(f) = i$ and, by definition of existential depth, all terms generated in f have existential depth i, in particular t. □

Proof of Lemma 3 *For any embedding φ from I to I' and any $i \geq 0$, there exists an embedding $\varphi' \supseteq \varphi$ from* $\mathsf{o\text{-}chase}^i(I, \Sigma)$ *to* $\mathsf{o\text{-}chase}^i(I', \Sigma)$ *which preserves the existential depth of terms, i.e., for every term v in* $\mathsf{o\text{-}chase}(I, \Sigma)$ *it holds that* $\mathsf{depth}_\exists(v) = \mathsf{depth}_\exists(\varphi'(v))$.

Proof. By induction on i. If $i = 0$ then all terms have existential depth 0 in I and I', then for $\varphi' = \varphi$ the thesis follows. Assume the property holds for $0 \leq i < n$. Let $i = n$. By inductive hypothesis there exists an embedding $\varphi' : \mathsf{o\text{-}chase}^{n-1}(I, \Sigma) \rightarrow \mathsf{o\text{-}chase}^{n-1}(I', \Sigma)$ preserving the existential depth of terms. Let (σ, π) be any trigger of $\mathsf{o\text{-}chase}^{n-1}(I, \Sigma)$. We know that $\varphi' \circ \pi(\mathsf{body}(\sigma)) \subseteq \mathsf{o\text{-}chase}^{n-1}(I', \Sigma)$ and $(\sigma, \varphi' \circ \pi)$ is a trigger of $\mathsf{o\text{-}chase}^{n-1}(I', \Sigma)$. Also, there exists a bijection ρ_n from the fresh terms in $\pi_{\mathsf{o}}(\mathsf{head}(\sigma))$ to the fresh terms in $\varphi' \circ \pi_{\mathsf{o}}(\mathsf{head}(\sigma))$ precisely defined as $\rho_n(z_{(\sigma,\pi)}) = z_{(\sigma,\varphi'\circ\pi)}$. Let $\varphi'' = \varphi' \uplus \bigcup \rho_n$ be the natural extension of φ' to all triggers that are performed to compute $\mathsf{o\text{-}chase}^n(I, \Sigma)$. Of course, for every trigger (σ, π) and term $v_B \in \mathsf{terms}(\pi(\mathsf{body}(\sigma)))$ we have that $\mathsf{depth}_\exists(v_B) = \mathsf{depth}_\exists(\varphi''(v_B))$. We want to show that φ'' also preserves the existential depth of fresh terms. Consider now the rule application $(\sigma, \varphi'' \circ \pi)$. Let z be an existential variable of σ. Then, $\mathsf{depth}_\exists(z_{(\sigma,\pi)}) = 1 + \max\{\mathsf{depth}_\exists(v_B)\} = 1 + \max\{\mathsf{depth}_\exists(\varphi''(v_B))\} = \mathsf{depth}_\exists(z_{(\sigma,\varphi''\circ\pi)})$. □

Proof of Lemma 6 *For any embedding φ from I to I' and any $i \geq 0$, there exists an embedding $\varphi' \supseteq \varphi$ from* $\mathsf{so\text{-}chase}^i(I, \Sigma)$ *to* $\mathsf{so\text{-}chase}^i(I', \Sigma)$ *which preserves the frontier depth of terms.*

Proof. By induction on i. If $i = 0$ then all values have frontier depth 0 in I and I', then for $\varphi' = \varphi$ the thesis follows. Assume the property holds for $0 \leq i < n$. Let $i = n$. By inductive hypothesis, we know that there exists an embedding $\varphi' : \mathsf{so\text{-}chase}^{n-1}(I, \Sigma) \rightarrow \mathsf{so\text{-}chase}^{n-1}(I', \Sigma)$

which preserves the frontier depth of values. Let (σ,π) be any trigger producing a new fact $f \in \mathsf{so\text{-}chase}^n(I,\Sigma)$. Then $\varphi' \circ \pi(\mathsf{body}(\sigma)) \subseteq \mathsf{so\text{-}chase}^{n-1}(I',\Sigma)$.

Consider first the case where $\mathsf{fr}(\sigma) = \emptyset$. In this case $\mathsf{depth}^{\mathsf{fr}}_{\exists}(f) = 1$ and any term $v \in \mathsf{terms}(f)$ is a fresh term $v = z_{(\sigma,\emptyset)}$ generated from an existential variable $z \in \mathsf{ex}(\sigma)$. Thus $f \in \mathsf{so\text{-}chase}^n(I',\Sigma)$ as well and the embedding φ' is the identity on the terms of f. Also, f has frontier depth 1 in $\mathsf{so\text{-}chase}^n(I',\Sigma)$.

Now, if $\mathsf{fr}(\sigma) \neq \emptyset$ we again distinguish two cases. If for all triggers of the form (σ,π') applied to compute $\mathsf{so\text{-}chase}^{n-1}(I',\Sigma)$ we have that $\varphi \circ \pi_{|\mathsf{fr}(\sigma)} \neq \pi'_{|\mathsf{fr}(\sigma)}$ then the trigger $(\sigma, \varphi' \circ \pi)$ has not yet been applied in $\mathsf{o\text{-}chase}^{n-1}(I',\Sigma)$. So, we define $\varphi'' \supseteq \varphi'$ to be such that $\varphi''(z_{(\sigma,\pi_{|\mathsf{fr}(\sigma)})}) = z_{(\sigma,\varphi' \circ \pi_{|\mathsf{fr}(\sigma)})}$ for every $z \in \mathsf{fr}(\sigma)$. Otherwise, there is a trigger (σ,π') such that $\varphi \circ \pi_{|\mathsf{fr}(\sigma)} = \pi'_{|\mathsf{fr}(\sigma)}$ applied to compute $\mathsf{so\text{-}chase}^{n-1}(I',\Sigma)$ which makes $(\sigma,\varphi' \circ \pi)$ producing the same result as (σ,π'). In this case we define $\varphi'' \supseteq \varphi'$ to be such that $\varphi''(z_{(\sigma,\pi_{|\mathsf{fr}(\sigma)})}) = z_{(\sigma,\pi'_{|\mathsf{fr}(\sigma)})}$ for all $z \in \mathsf{fr}(\sigma)$. To conclude, we have that $\mathsf{depth}^{\mathsf{fr}}_{\exists}(z_{(\sigma,\pi_{|\mathsf{fr}(\sigma)})}) = 1 + \max\{\mathsf{depth}^{\mathsf{fr}}_{\exists}(v) \mid v \in \pi(\mathsf{fr}(\sigma))\} = 1 + \max\{\mathsf{depth}^{\mathsf{fr}}_{\exists}(\varphi''(v)) \mid v \in \pi(\mathsf{fr}(\sigma))\} = \mathsf{depth}^{\mathsf{fr}}_{\exists}(\varphi''(z_{(\sigma,\pi_{|\mathsf{fr}(\sigma)})}))$. □

Proof of Theorem 7 *When $\Sigma \in \mathsf{CT}^{\mathsf{so}}$ there exists a constant k_d such that for all instance I, the frontier depth of a term in* $\mathsf{so\text{-}chase}(I,\Sigma)$ *is bounded by k_d.*

Proof. If Σ is in $\mathsf{CT}^{\mathsf{so}}$, the so-chase terminates on the critical instance. We take for k_d the smallest rank such that $\mathsf{so\text{-}chase}^{k_d}(I_a,\Sigma) = \mathsf{so\text{-}chase}(I_a,\Sigma)$. Every instance can be embedded into the critical instance. Hence, by Lemma 6 the frontier depth of the terms in $\mathsf{so\text{-}chase}(I,\Sigma)$ is bounded by the frontier depth of $\mathsf{so\text{-}chase}(I_a,\Sigma)$, which is itself bounded by k_d. □

Proof of Proposition 9 *Let Σ be a ruleset and HD_Σ be the full-atomic queries given by heads of the datalog rules in* $\mathsf{DF}(\Sigma)$. *Then,* $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{AF}}$ *if and only if* $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{HD}_\Sigma}$.

Proof. Since HD_Σ is a particular set of full-atomic queries, $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{AF}}$ implies $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{HD}_\Sigma}$. For the other direction, first note that Σ and $\mathsf{DF}(\Sigma)$ are equivalent sets of rules, hence they behave similarly with respect to first-order rewritability. Specifically, for any CQ Q and set of rules Σ, (Q,Σ) is FO-rewritable iff $(Q,\mathsf{DF}(\Sigma))$ is FO-rewritable. Hence, we conveniently consider in the following that Σ is in the form of $\mathsf{DF}(\Sigma)$.

When a query is rewritten, some answer variables may be made equal. Hence, we slightly generalize the notion of query $Q(x_1,...,x_k)$ by allowing to equate some answer variables, which is represented by assigning to Q the partition P_Q on $\{1,\ldots,k\}$ associated with answer variable equality, i.e., i and j are in the same class of P_Q iff the *ith* and *jth* answer variables of Q are the same. Given a class C in P_Q, we denote by x_C the answer variable associated with C. Then, the full-atomic query given by an atom has exactly the same arity as this atom, for instance the query associated with $p(x,x,y)$ is $Q(x_{\{1,2\}},x_{\{1,2\}},x_3) = p(x_{\{1,2\}},x_{\{1,2\}},x_3)$, with $P_Q = \{\{1,2\},\{3\}\}$, and not a query of the form $Q(x_1,x_2) = p(x_1,x_1,x_2)$. Given partitions P_1 and P_2 on $\{1,\ldots,k\}$, we note $P_1 \sqsubseteq P_2$ if P_1 is thinner than P_2, i.e., for each class $C \in P_1$, there is a class $C' \in P_2$ with $C \subseteq C'$. The $\sqsubseteq$ relation organizes the set of partitions of $\{1,\ldots,k\}$ into a lattice. As usual, we denote by $P_1 \vee P_2$ the upper bound of P_1 and P_2 in this lattice.

We recall that (Q,Σ), with Q a CQ, is FO-rewritable iff there is a UCQ-rewriting of Q, i.e., a finite set $\mathcal{Q}$ of CQs such that, for any instance I, the set of certain answers to Q on (I,Σ) is exactly the set of answers to the UCQ obtained from $\mathcal{Q}$ on I.[6] Each CQ in $\mathcal{Q}$ can be obtained from Q and Σ by a finite rewriting sequence, based on so-called piece-unifiers (see e.g., [König *et al.*, 2015] for definitions). More precisely, each query Q_{i+1} in a rewriting sequence is obtained from the preceding query Q_i, and a piece-unifier u of Q_i with a rule $\sigma \in \Sigma$ that unifies a non-empty subset of Q_i with a subset of σ's head while satisfying conditions concerning existential variables in σ. In particular, an answer variable of Q_i cannot be unified with an existential variable of σ. The following property holds: for any (I,Σ) and CQ Q, a tuple of constants $(a_1,\ldots,a_k)$ is a certain answer to Q on (I,Σ) iff there is a finite rewriting sequence from Q to a CQ Q' such that $(a_1,\ldots,a_k)$ is an answer to Q' on I. Also note that for any CQ Q_i obtained from a CQ Q by a rewriting sequence, $P_Q \sqsubseteq P_{Q_i}$ holds (P_Q is thinner than P_{Q_i}).

Now, assume $\Sigma \in \mathsf{FO\text{-}R}^{\mathsf{HD}_\Sigma}$ and let Q be a full-atomic query. If Q is not unifiable with a datalog rule from $\mathsf{DF}(\Sigma)$, its UCQ-rewriting is Q itself, because a full-atomic query is not unifiable with an FE rule. Otherwise, let Q_h be the full-atomic query associated with any datalog rule head unifiable with Q by a unifier u. One has $u(Q) = u(Q_h)$. If all such $(u(Q_h),\Sigma)$ are FO-rewritable, we obtain that (Q,Σ) is FO-rewritable, as the union of the UCQ-rewritings of all $u(Q_h)$ yields a suitable UCQ-rewriting of Q. We will show the following property (P1): let Q and Q_s be two full-atomic queries with the same predicate such that $P_Q \sqsubseteq P_{Q_s}$; if (Q,Σ) is FO-rewritable then (Q_s,Σ) also is. By hypothesis, each (Q_h,Σ) is FO-rewritable, hence (P1) implies that, for any substitution u, $(u(Q_h),\Sigma)$ is also FO-rewritable, which will conclude the proof.

It remains to prove (P1). We prove a preliminary lemma (L): let Q and Q_s be two full-atomic queries with the same predicate such that $P_Q \sqsubseteq P_{Q_s}$; then, for any rewriting sequence of length l leading from Q to a query Q^l, there is a rewriting sequence of the same length leading from Q_s to a query Q^l_s, such that $P_{Q^l_s} = P_{Q_l} \vee P_{Q_s}$ (where $\vee$ is the upper bound in the partition lattice), and, given s' the substitution of the answer variables in Q^l by the answer variables in Q^l_s associated with $P_{Q^l} \sqsubseteq P_{Q^l_s}$, it holds that $s'(Q^l) = Q^l_s$, up to a bijective renaming of non-answer variables. Let us now prove (P1). Let Q and Q_s be two full-atomic queries on the same predicate of arity k such that $P_Q \sqsubseteq P_{Q_s}$ and (Q,Σ) is FO-rewritable. The FO-rewritability of (Q,Σ) is equivalent to the

[6]As already noticed, the equivalence between the rewritability into a union of CQs and first-order rewritability follows from the (Finite) Homomorphism Preservation Theorem [Rossman, 2008].

following statement: there is an integer b such that for any instance I and any tuple of constants $(a_1, \ldots, a_k)$, it holds that $(a_1, \ldots, a_k)$ is a certain answer to Q on (I, Σ) if and only if there is a query Q^b obtained by a rewriting sequence from Q of length less than b, with $(a_1, \ldots, a_k)$ is an answer to Q^b on I. We prove that (Q_s, Σ) is FO-rewritable by such a statement. Let I be any instance and $(a_1, \ldots, a_k)$ be a certain answer to Q_s on (I, Σ). There is thus a homomorphism h from Q_s to $\star$-$\mathsf{chase}(I, \Sigma)$ that maps its answer variable tuple to $(a_1, \ldots, a_k)$. Given s the homomorphism from Q to Q_s, it holds that $h \circ s$ is a homomorphism from Q to $\star$-$\mathsf{chase}(I, \Sigma)$ that maps its answer variable tuple to $(a_1, \ldots, a_k)$. Since Q is FO-rewritable, there is a query Q^b obtained by a rewriting sequence of length less than b such that $(a_1, \ldots, a_k)$ is an answer to Q^b on I. Let h^b be a homomorphism from Q^b to I yielding this answer. Let P_a be the partition on $\{1, ..., k\}$ associated with the equality of terms in $(a_1, \ldots, a_k)$. We have $P_{Q^b} \sqsubseteq P_a$ and $P_{Q_s} \sqsubseteq P_a$. By Lemma (L), there is a query Q_s^b obtained from Q_s with a rewriting sequence of length less than b, such that (1) $P_{Q_s^b} = P_{Q^b} \vee P_{Q_s}$, and (2) $s'(Q^b) = Q_s^b$, with s' the substitution associated with $P_{Q_b} \sqsubseteq P_{Q_s^b}$. From (1), we have $P_{Q_s^b} \sqsubseteq P_a$. Hence, the homomorphism h^b from Q^b to I can be written $h' \circ s'$, where h' is a homomorphism from Q_s^b to I mapping its answer tuple to $(a_1, \ldots, a_k)$. The converse direction ("if there is Q_s^b obtained by a rewriting sequence from Q_s, of length less than b, with $(a_1, \ldots, a_k)$ is an answer to Q_s^b on I, then $(a_1, \ldots, a_k)$ is a certain answer to Q_s on (I, Σ)") holds because of the soundness of query rewriting based on piece-unifiers.

$\square$

Proof of Lemma 10 *If $\Sigma \in \mathsf{FO\text{-}R^{AF}}$ there is a constant k_{AF} such that, for any instance I and fact f such that $\mathsf{terms}(f) \subseteq \mathsf{terms}(I)$, when $f \in \mathsf{o\text{-}chase}(I, \Sigma)$ it holds that $\mathsf{rank}(f) \leq k_{\mathrm{AF}}$.*

Proof. Assume $\Sigma \in \mathsf{FO\text{-}R^{AF}}$. We take for k_{AF} the maximal number of breadth-first rewriting steps necessary to obtain a UCQ-rewriting of a full-atomic query (we refer here to the breadth-first rewriting based on aggregated piece-unifiers, see [König *et al.*, 2013]; this query rewriting technique ensures the following property: for any (I, Σ) and any CQ Q, for any k, the set of answers to Q on $\mathsf{o\text{-}chase}^k(I, \Sigma)$ is equal to the set of answers to Q_k on I, where Q_k is the UCQ-rewriting of Q with Σ obtained by k breadth-first rewriting steps).[7]

We know that the number of full-atomic queries to be considered is finite and by Proposition 9 can be even bounded by the number of non-isomorphic heads of datalog rules. By the properties of breadth-first query rewriting, we know that for any instance I and any full-atomic query Q, the certain answers to Q on (I, Σ) are exactly the answers to Q on $\mathsf{o\text{-}chase}^{k_{\mathrm{AF}}}(I, \Sigma)$. We can identify an answer $(a_1, \ldots, a_n)$ to $Q = p(x_1, \ldots, x_n)$ with the fact (having only constants) $p(a_1, \ldots, a_n)$. Then we have that $(a_1, \ldots, a_n)$ is a certain answer to Q iff $p(a_1, \ldots, a_n) \in \mathsf{o\text{-}chase}^{k_{\mathrm{AF}}}(I, \Sigma)$. Now, the statement of the lemma considers more generally facts f with $\mathsf{terms}(f) \subseteq \mathsf{terms}(I)$ which could contain also existentially quantified variables. Observe however that, for any instance I, let $\mathsf{freeze}()$ be a bijective renaming of the variables of I by constant values that do not appear in I, then, for all $i \geq 0$, there is an isomorphism from $\mathsf{o\text{-}chase}^i(I, \Sigma)$ to $\mathsf{o\text{-}chase}^i(\mathsf{freeze}(I), \Sigma)$ that preserves the rank of facts. Because of this, let f and I be any fact and instance, we have that $f \in \mathsf{o\text{-}chase}^i(I, \Sigma)$ (for any i) iff $\mathsf{freeze}(f) \in \mathsf{o\text{-}chase}^i(\mathsf{freeze}(I), \Sigma)$. Since $\mathsf{freeze}(f)$ contains only constants we can conclude. $\square$

[7] Alternatively, we could rely on the bound given by the bounded derivation-depth property (BDDP) [Calì *et al.*, 2009b]. A ruleset Σ satisfies the BDDP if for all Boolean CQ Q, there is an integer k such that, for all instance I, it holds that $I, \Sigma \models Q$ iff $\mathsf{o\text{-}chase}^k(I, \Sigma) \models Q$. It has been several times remarked that BDDP is equivalent to UCQ-rewritability, hence to FO-rewritability.

Proof of Lemma 11 *If $\Sigma \in \mathsf{FO\text{-}R}$ there is a constant k_{FO} such that, for any instance I and any trigger (σ, π) from $\mathsf{so\text{-}chase}(I, \Sigma)$ with $\pi(\mathsf{fr}(\sigma)) \subseteq \mathsf{terms}(I)$, there is also a trigger (σ, π') from $\mathsf{so\text{-}chase}(I, \Sigma)$ such that $\pi'_{|\mathsf{fr}(\sigma)} = \pi_{|\mathsf{fr}(\sigma)}$ and $\mathsf{rank}(f) \leq k_{\mathrm{FO}}$ for all $f \in \pi'(\mathsf{body}(\sigma))$.*

Proof. Assume $\Sigma \in \mathsf{FO\text{-}R}$. We denote by $Q_{\mathsf{body}(\sigma)}$ any conjunctive query whose atoms correspond to the atoms of $\mathsf{body}(\sigma)$, for $\sigma \in \Sigma$, and all variables are existentially quantified except for those in $\mathsf{fr}(\sigma)$, which are the answer variables. We know that the number of such queries is bounded by the cardinal of Σ. We take for k_{FO} the maximal number of breadth-first rewriting steps necessary to obtain a UCQ-rewriting from any $Q_{\mathsf{body}(\sigma)}$ query.

By the properties of breadth-first query rewriting (based on aggregated piece-unifiers [König *et al.*, 2013]), we know that, for any instance I, the certain answers to $Q_{\mathsf{body}(\sigma)}$ on (I, Σ) are exactly the answers to $Q_{\mathsf{body}(\sigma)}$ on $\mathsf{o\text{-}chase}^{k_{\mathrm{FO}}}(I, \Sigma)$ and therefore on $\mathsf{so\text{-}chase}^{k_{\mathrm{FO}}}(I, \Sigma)$. By definition of query-answer, there is a homomorphism π' from $Q_{\mathsf{body}(\sigma)}$ to $\mathsf{so\text{-}chase}^{k_{\mathrm{FO}}}(I, \Sigma)$ which maps each answer variable (recall, originated from a frontier variable) to a constant of I and each existentially quantified variable to a term of $\mathsf{so\text{-}chase}^{k_{\mathrm{FO}}}(I, \Sigma)$.

Let (σ, π) any trigger from $\mathsf{so\text{-}chase}(I, \Sigma)$ with $\pi(\mathsf{fr}(\sigma)) \subseteq \mathsf{terms}(I)$. Assume first that all values in I are constants. According to the previous observation, we know that there is π' from $Q_{\mathsf{body}(\sigma)}$ to $\mathsf{so\text{-}chase}^{k_{\mathrm{FO}}}(I, \Sigma)$ with $\pi'_{|\mathsf{fr}(\sigma)} = \pi_{|\mathsf{fr}(\sigma)}$ and therefore a trigger (σ, π') as desired. Now, for the case where I also contains existentially quantified variables, let I be an instance and let $\mathsf{freeze}()$ be a bijective renaming of the variables of I by constant values that do not appear in I then, for all $i \geq 0$, there is an isomorphism from $\mathsf{so\text{-}chase}^i(I, \Sigma)$ to $\mathsf{so\text{-}chase}^i(\mathsf{freeze}(I), \Sigma)$ that preserves the rank of facts. This implies that if I is any instance then for any f and i holds $f \in \mathsf{so\text{-}chase}^i(I, \Sigma)$ iff $\mathsf{freeze}(f) \in \mathsf{so\text{-}chase}^i(\mathsf{freeze}(I), \Sigma)$ and we conclude. $\square$

A closer look at our proof actually shows that FO-rewritability with respect to queries associated with rule bodies is sufficient to derive Lemma 11.

Proof of Theorem 12 *If $\Sigma \in$ FO-R$^{\mathsf{AF}}$ then for all instance I and fact $f \in \mathsf{o\text{-}chase}(I, \Sigma)$ we have that $\mathsf{rank}(f) \leq \mathsf{depth}_\exists(f) \times (k_{\mathrm{AF}} + 1) + k_{\mathrm{AF}}$ with k_{AF} the bound provided by Lemma 10.*

Proof. We first show that since $\Sigma \in$ FO-R$^{\mathsf{AF}}$ then for all instance I and term $v \in \mathsf{terms}(\mathsf{o\text{-}chase}(I, \Sigma))$ it holds that $\mathsf{rank}(v) \leq \mathsf{depth}_\exists(v) \times (k_{\mathrm{AF}} + 1)$, where we recall that $\mathsf{rank}(v)$ is the rank where v is introduced.

By induction on the existential depth of v. If $\mathsf{depth}_\exists(v) = 0$ then $v \in \mathsf{terms}(I)$ and thus $\mathsf{rank}(v) = 0$ also. Assume the property holds for $0 \leq \mathsf{depth}_\exists(v) \leq n$. We show that it holds for $\mathsf{depth}_\exists(v) = n + 1$. Let (σ, π) be the trigger that generates v. Then, for all $v_B \in \mathsf{terms}(\pi(\mathsf{body}(\sigma)))$, we know that $\mathsf{depth}_\exists(v_B) \leq n$. By inductive hypothesis, $\mathsf{rank}(v_B) \leq n \times (k_{\mathrm{AF}} + 1)$. Since $\Sigma \in$ FO-R$^{\mathsf{AF}}$, we can apply Lemma 10 using as instance $\mathsf{o\text{-}chase}^k(I, \Sigma)$ where $k = n \times (k_{\mathrm{AF}} + 1)$. Hence for all $f_B \in \pi(\mathsf{body}(\sigma))$ it holds that $\mathsf{rank}(f_B) \leq k + k_{\mathrm{AF}}$. Thus $\mathsf{rank}(v) \leq k + k_{\mathrm{AF}} + 1 = (n+1) \times (k_{\mathrm{AF}} + 1) = \mathsf{depth}_\exists(v) \times (k_{\mathrm{AF}} + 1)$. To conclude the proof, since any fact $f \in \mathsf{o\text{-}chase}(I, \Sigma)$ contains only terms v with $\mathsf{rank}(v) \leq \mathsf{depth}_\exists(v) \times (k_{\mathrm{AF}} + 1)$, we apply Lemma 10 and we obtain $\mathsf{rank}(f) \leq \max\{ \mathsf{depth}_\exists(v) \mid v \in \mathsf{terms}(f) \} \times (k_{\mathrm{AF}} + 1) + k_{\mathrm{AF}}$. ☐

Proof of Theorem 13 *If $\Sigma \in$ FO-R then for all instance I and fact $f \in \mathsf{so\text{-}chase}(I, \Sigma)$ we have that $\mathsf{rank}(f) \leq \mathsf{depth}^{\mathsf{fr}}_\exists(f) \times (k_{\mathrm{FO}} + 1) + k_{\mathrm{FO}}$ where k_{FO} is the bound provided by Lemma 11.*

Proof. We first show that since Σ is FO-rewritable, then for all instance I and term $v \in \mathsf{terms}(\mathsf{so\text{-}chase}(I, \Sigma))$ it holds that $\mathsf{rank}(v) \leq \mathsf{depth}^{\mathsf{fr}}_\exists(v) \times (k_{\mathrm{FO}} + 1)$. By induction on the frontier depth of v. If $\mathsf{depth}^{\mathsf{fr}}_\exists(v) = 0$ then $v \in \mathsf{terms}(I)$ and thus $\mathsf{rank}(v) = 0$ also. Assume the property holds for $0 \leq \mathsf{depth}^{\mathsf{fr}}_\exists(v) \leq n$. We show that it holds for $\mathsf{depth}^{\mathsf{fr}}_\exists(v) = n+1$. Let (σ, π) be the trigger that generates v. By definition of frontier depth, for all $v_{fr} \in \mathsf{terms}(\pi(\mathsf{fr}(\sigma)))$, we know that $\mathsf{depth}^{\mathsf{fr}}_\exists(v_{fr}) \leq n$. By inductive hypothesis, $\mathsf{rank}(v_{fr}) \leq n \times (k_{\mathrm{FO}} + 1)$. Since $\Sigma \in$ FO-R, we can apply Lemma 11 using as instance $\mathsf{so\text{-}chase}^k(I, \Sigma)$ where $k = n \times (k_{\mathrm{FO}} + 1)$. This gives us $\mathsf{rank}(\pi(\mathsf{body}(\sigma))) \leq k + k_{\mathrm{FO}}$. Thus $\mathsf{rank}(v) \leq n \times (k_{\mathrm{FO}} + 1) + k_{\mathrm{FO}} + 1 = (n+1) \times (k_{\mathrm{FO}} + 1) = \mathsf{depth}^{\mathsf{fr}}_\exists(v) \times (k_{\mathrm{FO}} + 1)$.

Since any fact $f \in \mathsf{so\text{-}chase}(I, \Sigma)$ contains only terms v with $\mathsf{rank}(v) \leq \mathsf{depth}^{\mathsf{fr}}_\exists(v) \times (k_{\mathrm{FO}}+1)$, we use again Lemma 11, and obtain $\mathsf{rank}(f) \leq \mathsf{depth}^{\mathsf{fr}}_\exists(f) \times (k_{\mathrm{FO}}+1) + k_{\mathrm{FO}}$. ☐

Proof of Theorem 14 BN$^{\mathsf{o}}$ = FO-R$^{\mathsf{AF}}$ $\cap$ CT$^{\mathsf{o}}$.

Proof. ($\Rightarrow$) If $\Sigma \in$ BN$^{\mathsf{o}}$ there is k such that, for all I, $\mathsf{o\text{-}chase}^k(I, \Sigma) = \mathsf{o\text{-}chase}(I, \Sigma)$. Hence, there is k such that for all I and all Boolean CQ Q we have $\Sigma, I \models Q$ (i.e., $\mathsf{o\text{-}chase}(I, \Sigma) \models Q$ iff $\mathsf{o\text{-}chase}^k(I, \Sigma) \models Q$ iff $I \models Q^k$, where Q^k is the query obtained by k steps of breadth-first rewriting from Q and Σ, which implies the FO-rewritability of Σ. To conclude, BN$^{\mathsf{o}}$ $\subseteq$ CT$^{\mathsf{o}}$ follows by definition.

($\Leftarrow$) By Theorem 4 and 12. ☐

Proof of Theorem 15 BN$^{\mathsf{so}}$ = FO-R $\cap$ CT$^{\mathsf{so}}$.

Proof. Identical to proof of Theorem 14 for the direct sense. The other direction holds by Theorem 7 and 13. ☐

The next proposition leads to conclude that membership to BN$^{\mathsf{so}}$, CT$^{\mathsf{so}}$ and FO-R remains undecidable for FE-rules.

Proposition 18. *There is a translation ψ from any KB (I, Σ) on a vocabulary $\mathcal{V}$, where Σ is a set of existential rules, to a KB $(\psi(I), \psi(\Sigma))$ on a vocabulary $\psi(\mathcal{V})$, where $\psi(\Sigma)$ is a set of FE-rules, such that:*
(1) ψ is injective, and
(2) $\mathsf{so\text{-}chase}(I, \Sigma)$ and $\mathsf{so\text{-}chase}(\psi(I), \psi(\Sigma))$ have the same rank, and
(3) for any instance I' on $\psi(\mathcal{V})$, there is an instance $\psi(I)$ such that $\mathsf{so\text{-}chase}(I', \psi(\Sigma))$ and $\mathsf{so\text{-}chase}(\psi(I), \psi(\Sigma))$ have the same rank.

The proposition leads directly to the undecidability of BN$^{\mathsf{so}}$ and CT$^{\mathsf{so}}$for FE-rules. Concerning the undecidability of FO-R for FE-rules, we take for Σ a set of datalog rules. Then Σ is in CT$^{\mathsf{so}}$ if and only if $\psi(\Sigma)$ is in CT$^{\mathsf{so}}$. Since every datalog set is CT$^{\mathsf{so}}$, $\psi(\Sigma)$ is also CT$^{\mathsf{so}}$. Now, consider the (undecidable) problem of whether Σ is (uniformly) bounded. We have that Σ is bounded iff $\psi(\Sigma)$ is bounded, which amounts to asking if $\psi(\Sigma)$ is FO-R (as we already know it is in CT$^{\mathsf{so}}$).

Proof. (of proposition 18). Take a vocabulary $\mathcal{V} = (\mathcal{P}, \mathcal{C})$ and define the set $\mathcal{P}^+$ where each predicate $p \in \mathcal{P}$ of arity k is replaced by a predicate p^+ of arity $k+1$. Let ψ be a transformation defined as follows. First, $\psi(\mathcal{V}) = (\mathcal{P}^+, \mathcal{C})$. Then, given an atom $\alpha = p(v_1, \ldots, v_k)$ then $\psi(\alpha) = p^+(v_1, \ldots, v_k, z_\alpha)$ where z_α is a fresh variable. Let $\sigma = B_1(\bar{x}, \bar{y}) \ldots B_n(\bar{x}, \bar{y}) \rightarrow \exists \bar{z} H_1(\bar{x}, \bar{z}), \ldots, H_m(\bar{x}, \bar{z})$ be a rule, then $\psi(\sigma) = \psi(B_1), \ldots, \psi(B_n) \rightarrow \exists z_{H_1} \ldots z_{H_m} \psi(H_1) \ldots \psi(H_m)$. Finally, $\psi(I) = \bigcup_{\alpha \in I} \psi(\alpha)$ and $\psi(\Sigma) = \bigcup_{\sigma \in \Sigma} \psi(\sigma)$.

Obvliously, ψ is injective (Point (1)).

To prove the point (2), we show that for each fact $f = p(v_1, \ldots, v_n) \in \mathsf{so\text{-}chase}^i(I, \Sigma)$ generated by a trigger (σ, π) it holds $p^+(v_1, \ldots, v_n, z_{(\psi(\sigma), \pi_{|\mathsf{fr}(\sigma)})}) \in \mathsf{so\text{-}chase}^i(\psi(I), \psi(\Sigma))$, and vice-versa.

We focus on the direction $\Rightarrow$) as the direction $\Leftarrow$) is similar. By induction on the rank i of the so-chase. If $i = 0$ then by definition $f \in I$ implies $\psi(f) \in \psi(I)$. Assume that the property holds for $0 \leq i \leq n$. We show that it holds for rank $n+1$. Let f be any atom of rank $n+1$ produced by the trigger (σ, π). This means that for all body atom $f_B = p(x_1, \ldots, x_k) \in \mathsf{body}(\sigma)$ we know that $\pi(f_B) = p(v_1, \ldots, v_n) \in \mathsf{so\text{-}chase}^n(I, \Sigma)$. Hence, by induction $\psi(\pi(f_B)) = p^+(v_1, \ldots, v_n, z^+) \in \mathsf{so\text{-}chase}^n(\psi(I), \psi(\Sigma))$ where $z^+ = z_{\pi(f_B)}$ if $\pi(f_B) \in I$, or $z^+ = z_{(\psi(\sigma'), \pi'_{|\mathsf{fr}(\sigma')})}$ if $\pi(f_B)$ has been generated by a trigger (σ', π'). Then, the trigger $(\psi(\sigma), \pi \cup \bigcup_{f_B} \{z_{f_B} \mapsto z^+\})$ is applicable and produces $\psi(\pi(f))$. Since f is of rank $n+1$ there does not exist another rule application that could have generated the same atom at a previous rank, and the same holds for its image.

For point (3), we build a transformation ϕ from any instance I' on $\psi(\mathcal{V})$ to an instance I on $\mathcal{V}$ such that

$\mathsf{so\text{-}chase}(I', \psi(\Sigma))$ and $\mathsf{so\text{-}chase}(I, \Sigma)$ have the same rank. [Note that the proof does not follow exactly point (3) here: we consider directly I instead of $\psi(I)$].

The transformation ϕ assigns to each atom $\alpha = p^+(v_1, \ldots, v_k, z)$ on $\psi(\mathcal{V})$ the atom $\phi(\alpha) = p(v_1, \ldots, v_k)$. Let $I = \phi(I') = \bigcup_{\alpha \in I'} \phi(\alpha)$. We show that for each fact $f = p^+(v_1, \ldots, v_n, z) \in \mathsf{so\text{-}chase}^i(I', \psi(\Sigma))$ generated by a trigger $(\psi(\sigma), \pi)$, it holds that $p(v_1, \ldots, v_n) \in \mathsf{so\text{-}chase}^i(I, \Sigma)$. By induction on the rank i of the so-chase. If $i = 0$ then by definition $f \in I'$ implies $\phi(f) \in I$. Assume that the property holds for $0 \leq i \leq n$. We show that it holds for rank $n + 1$. Let f be any atom of rank $n + 1$ produced by the trigger $(\psi(\sigma), \pi)$. This means that for all body atom $f_B = p^+(v_1, \ldots, v_k, z) \in \mathsf{body}(\psi(\sigma))$ we know that $\pi(f_B) \in \mathsf{so\text{-}chase}^n(I', \psi(\Sigma))$. Hence, by induction $\phi(\pi(f_B)) \in \mathsf{so\text{-}chase}^n(I, \Sigma)$. Then, the trigger (σ, π') (where π' is the appropriate restriction of π) is applicable and produces $\phi(\pi(f))$. Since f is of rank $n + 1$ there does not exist another rule application that could have generated the same atom at a previous rank; as the chase is semi-oblivious, and the last component of a predicate never occurs in the frontier, $\phi(\pi(f))$ is also of rank $n + 1$.

So, let an instance I on $\mathcal{V}$. Let us note that $\phi(\psi(I)) = I$. By what precedes $\mathsf{so\text{-}chase}(I, \Sigma)$ and $\mathsf{so\text{-}chase}(\psi(I), \psi(\Sigma))$ have the same rank. Now, let an instance I' on $\psi(\mathcal{V})$ and $I = \phi(I')$: by what precedes the rank of $\mathsf{so\text{-}chase}(I, \Sigma)$ is at least the rank of $\mathsf{so\text{-}chase}(I', \psi(\Sigma))$. Furthermore, by embedding $\psi(I)$ in I', we can by using similar arguments prove that the rank of $\mathsf{so\text{-}chase}(I', \psi(\Sigma))$ is at least the rank of $\mathsf{so\text{-}chase}(I, \Sigma)$. So, $\mathsf{so\text{-}chase}(I', \psi(\Sigma))$ and $\mathsf{so\text{-}chase}(I, \Sigma)$ have the same rank.

□

Proof of Theorem 17 *The* k-boundedness *problem is:*

- *in 2Exptime on general existential rules for the* o-chase *and* so-chase*;*
- *co-NExptime-complete on datalog;*
- *in co-NExptime on FE-rules for the* o-chase.

Proof. The upper bound results mostly come from [Delivorias *et al.*, 2018] [Delivorias *et al.*, 2020]. Indeed, from these papers, a ruleset is k-bounded for the (s)o-chase iff the (s)o-chase stops within k steps for instances of size at most b^{k+1}, with b the maximum number of atoms in a rule body. So, to disprove k-boundedness, it suffices to guess a breadth-first derivation from an instance of size at most b^{k+1} to a fact of rank $k + 1$, which can be done in NExptime for datalog. This gives also the 2Exptime upper bound by checking exhaustively for each of these instances that there is no breadth-first derivation of depth $k + 1$. When the ruleset is fully existential, it can be proven that a rule set is not $k-$bounded for the o-chase iff there exists a non-necessarly breadth-first partial derivation of depth $k + 1$ from the critical instance. So, by guessing such a derivation, we get also a co-NExptime decision procedure on FE-rules for the o-chase. See also [Gallois, 2019] for detailed proofs of the above results.

Co-NExptime-hardness of k-boundedness for datalog rules is proven by reduction from the co-NExptime-hard inclusion problem of non-recursive Boolean datalog queries [Benedikt and Gottlob, 2010]. Let Q_1, Q_2 two non-recursive Boolean datalog queries and P_0 (resp. P_0^2) their respective distinguished 0-ary predicate. As they are non-recursive, Q_1 (resp. Q_2) is k_1-(resp.k_2) bounded with k_1 (resp. k_2) the number of predicates in Q_1 (resp. Q_2). Let $p = max(k_1, k_2) + 2$. Let us note that the size of p encoded in unary is bounded by the size of (Q_1, Q_2). Let us define a new ruleset $Q_1' \cup Q_2'$: Q_1' (resp. Q_2') is obtained from Q_1 (resp. Q_2) by adding 0-ary predicates P_i and rules $P_{i-1} \rightarrow P_i$ (resp. $P_0 \rightarrow P_i$) with $1 \leq i \leq p$. The size of $Q_1' \cup Q_2'$ is linear w.r.t. the size of (Q_1, Q_2). We will prove that $Q_1' \cup Q_2'$ is $p - 1$-bounded iff Q_1 is contained in Q_2

Let us first suppose that Q_1 is contained in Q_2. Let I be any instance. If P_0 can be derived from (I, Q_2), all the P_i are generated in at most k_2+1 steps and so the breadth-first chase for $(I, Q_1' \cup Q_2')$ stops after $max(k_1, k_2+1)$ steps. Otherwise, P_0 can neither be derived from (I, Q_1) and the breadth-first chase for $(I, Q_1' \cup Q_2')$ stops after $max(k_1, k_2)$ steps. So, in both cases, $Q_1' \cup Q_2'$ is $(p - 1)$-bounded.

If Q_1 is not contained in Q_2, there exists I such that P_0 can be derived from (I, Q_1) whereas P_0^2 can not be derived from (I, Q_2). As P_0^2 is not generated by the Q_2' part, P_k will be generated by the Q_1' part, and so the breadth-first chase for $(I, Q_1' \cup Q_2')$ will need at least p steps.

So $Q_1' \cup Q_2'$ is $p-1$-bounded iff Q_1 is contained in Q_2. □