Oblivious and Semi-Oblivious Boundedness for Existential Rules

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

To cite this version:


HAL Id: lirmm-02920624
https://hal-lirmm.ccsd.cnrs.fr/lirmm-02920624
Submitted on 24 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Oblivious and Semi-Oblivious Boundedness for Existential Rules

Pierre Bourhis\textsuperscript{1,3,4}, Michel Leclère\textsuperscript{2,4}, Marie-Laure Mugnier\textsuperscript{2,4}, Sophie Tison\textsuperscript{3,4}, Federico Ulliana\textsuperscript{2,4} and Lily Galois\textsuperscript{3,4}

\textsuperscript{1} CNRS, France
\textsuperscript{2} Univ. Montpellier, LIRMM, France
\textsuperscript{3} Univ. Lille, CRIS\textsc{tal}, France
\textsuperscript{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 [Cali et al., 2009a; Baget et al., 2011; Cali 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].\footnote{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 [Delivoria 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 $V = (\mathcal{P}, C)$ constituted of a finite set of predicates $\mathcal{P}$ and a finite set of constants $C$. A term $t$ is a constant of $C$ or a variable. An atom is of the form $p(v_1 \ldots v_k)$ where $p$ is a predicate of arity $k$ and the $v_i$ are terms. We denote by terms() the set of its terms and extend the notation to sets of atoms. An embedding $\varphi$ from a set of atoms $A$ to a set of atoms $A'$ is a substitution of terms($A$) with 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 $\sigma$ is a closed formula $\forall \bar{x} T[B] \rightarrow \exists \bar{z} H[\bar{x}, \bar{z}]$ where $B$ and $H$ are sets of atoms built on variables called the body and the head of the rule, also denoted by body($\sigma$) and 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 fr($\sigma$). The set of variables $\bar{z}$ that belong to $H$ only are called existential variables and are denoted by ex($\sigma$).

Universal quantifiers will often be omitted in the remainder of the paper. A rule such that $\text{ex}(\sigma) = \emptyset$ 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 $\sigma$ is applicable on an instance $I$ if there is a homomorphism $\pi$ from body($\sigma$) to $I$ and call the pair $\langle \sigma, \pi \rangle$ a trigger of $I$. Given a trigger $\langle \sigma, \pi \rangle$, we denote by $\pi_{\text{fr}(\sigma)} \subseteq \pi$ the restriction of $\pi$ to fr($\sigma$).

A knowledge base (KB) is a pair $(I, \Sigma)$ where $I$ is an instance and $\Sigma$ 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 $(I, \Sigma)$ be a knowledge base and $\ast \in \{o, so\}$ a chase variant. Then, the breadth-first $\ast$-chase is defined as follows: $\ast$-chase$^0(I, \Sigma) = I$ and for all saturation rank $i \geq 0$

$$\ast$$-chase$^{i+1}(I, \Sigma) = \ast$-chase$^i(I, \Sigma) \cup \{ \pi_\sigma(\text{head}(\sigma)) \}_{\langle \sigma, \pi \rangle}$

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

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

Then, we define $\ast$-chase$(I, \Sigma) = \bigcup_{i \geq 0} \ast$-chase$^i(I, \Sigma)$. The $\ast$-chase terminates on $(I, \Sigma)$ if there is a rank $k$ with $\ast$-chase$(I, \Sigma) = \ast$-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 $\sigma$ is transformed by replacing each occurrence of an existential variable $z$ with a functional term $f^\sigma_{\text{fr}(\sigma)}(\sigma, z)$ on the frontier of $\sigma$; 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 o-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, so-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, $\sigma$ is rewritten as $\sigma' = p(x, y) \rightarrow p(x, f^+_x(x))$. The first rule application according to $\sigma'$ produces $p(a, f^+_x(a))$, then the chase halts as the same atom is produced by the next trigger.

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

Note that for the breadth-first chases we consider the above definition implies that rank$(f)$ is the smallest $k$ such that $f \in \star$-chase$^k(I, \Sigma) \setminus \star$-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) $Q$ is a disjunction of CQs with the same free variables, also seen as a set of CQs.

A tuple of constants $(a_1, ..., a_n) \in C^n$ 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, $I \models Q[x_1 \mapsto a_1]$, where $\models$ denotes the classical logical consequence and $Q[x_1 \mapsto a_1]$ is the Boolean query obtained from $Q$ substituting each $x_i$ with $a_i$. A tuple of constants $(a_1, ..., a_n) \in C^n$ is a certain answer to $Q$ on a KB $(I, \Sigma)$ if $I, \Sigma \models Q[x_1 \mapsto a_1]$. This is equivalent to the existence of a saturation rank $k$ such that $\star$-chase$^k(I, \Sigma) \models Q[x_1 \mapsto a_1]$. In other words, the certain answers to $Q$ on $(I, \Sigma)$ are exactly its answers on the possibly infinite instance $\star$-chase$(I, \Sigma)$. The set of (certain) answers to a UCQ $Q$ 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 $\star \in \{o, so\}$ be a chase variant, the $\star$-chase termination class, denoted by $\text{CT}^\star$, contains all rulesets $\Sigma$ such that $\star$-chase$(I, \Sigma)$ terminates for all instances $I$. The $\star$-boundedness class, denoted by $\text{BN}^\star$, contains all bounded rulesets $\Sigma$, i.e., for which there exists an integer $k$ such that $\star$-chase$^k(I, \Sigma) = \star$-chase$(I, \Sigma)$ for all instances $I$. Obviously, $\text{BN}^\star \subseteq \text{CT}^\star$.

**Example 2.** Let $\sigma_1 = p(x, y) \land p(y, z) \rightarrow p(x, z)$ and $\sigma_2 = p(x, y) \lor p(u, w) \rightarrow p(x, z)$. Because both rules are datalog, $\text{CT}^o \ni \{\sigma_1\}$ and $\{\sigma_2\} \subseteq \text{CT}^o$. However, $\Sigma = \{\sigma_1\} \notin \text{BN}^o$, since the rank of $\star$-chase$(I, \Sigma)$ depends on $I$. In contrast, $\{\sigma_2\} \subseteq \text{BN}^o$ and the bound is $k = 1$. Similarly, $\{\sigma_1, \sigma_2\} \in \text{BN}^o$. Indeed, $\sigma_2$ produces at the first rank all atoms that can be produced by $\sigma_1$ 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 $p(x, y) \rightarrow \exists z p(x, z) \land q(x)$ 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 $\sigma$ be any existential rule of the form $B \rightarrow H_F \land H_D$ 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 $\sigma$, denoted by $\text{DF}(\sigma)$, returns a set made of the FE-rule $B \rightarrow H_F$ together with a (single head) datalog rule of the form $B \rightarrow H_D'$, for each $H_D' \subseteq H_D$. The definition is then extended to sets $\text{DF}(\Sigma) = \bigcup_{\sigma \in \Sigma} \text{DF}(\sigma)$. This decomposition preserves boundedness and termination of the oblivious chase.²

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

### 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 $\circ$-chase$(I, \Sigma)$ is

$$
depth_3(v) = \begin{cases} 
0 & \text{if } v \in \text{terms}(I) \\
1 + \max\{\text{depth}_3(v_B)\} & \text{otherwise}
\end{cases}
$$

where $v_B$ is any term in $\pi(\text{body}(\sigma))$ used by a trigger $(\sigma, \pi)$ which generates $v$. The existential depth of a fact $f$ is the maximum existential depth of its terms. The existential depth of $\circ$-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 $\circ$-chase$(I, \{\sigma\})$ is unbounded, which is in line with the non-termination of the $\circ$-chase on $(I, \{\sigma\})$. The rule $\sigma_1$ in Example 2 shows the difference between rank and existential depth. For any $I$, the existential depth

²This is not true for the so-chase. For instance, for $\Sigma = \{\sigma = p(x, y) \rightarrow \exists z p(x, z) \land q(x, y)\}$ and $I = \{p(a, b)\}$, so-chase$(\Sigma, I)$ is infinite, while $\text{DF}(\Sigma)$ is so-bounded. This is due to the fact that $\sigma$ has frontier $\{x, y\}$, while the FE-rule $p(x, y) \rightarrow \exists z p(x, z)$ in $\text{DF}(\sigma)$ 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 $\sigma_1$ is datalog, however their rank depends on $I$. More generally, for any term $v$ and fact $f$ in o-chase$(I, \Sigma)$ it holds that $\text{depth}_{\exists}(v) \leq \text{rank}(v)$ and $\text{depth}_{\exists}(f) \leq \text{rank}(f)$. Hence, if o-chase$(I, \Sigma)$ terminates, its existential depth is finite. Reciprocally, when the existential depth of o-chase$(I, \Sigma)$ is finite, so it is the number of its terms, and o-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 $\Sigma$ is a set of FE-rules then, for all instance $I$ and term $v$ in o-chase$(I, \Sigma)$, holds that $\text{depth}_{\exists}(v) = \text{rank}(v)$.

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 is preserved by embeddings.

**Lemma 3.** For any embedding $\varphi$ from $I$ to $I'$ and any $v \in \alpha$.$(I)$ and any $\pi \in \Pi$, there exists an embedding $\varphi' : a \rightarrow \varphi(a)$ from o-chase$(I, \Sigma)$ to o-chase$(I', \Sigma)$ which preserves the existential depth of terms, i.e., for every term $v$ in o-chase$(I, \Sigma)$ it holds that $\text{depth}_{\exists}(v) = \text{depth}_{\exists}(\varphi'(v))$.

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 $\Sigma \in \text{CT}$ there exists a constant $k_\text{d}$ such that for every instance $I$, the existential depth of a term in o-chase$(I, \Sigma)$ is bounded by $k_\text{d}$.

**Proof.** Because $\Sigma \in \text{CT}$, the o-chase terminates on the critical instance $I_c$. Let $k_d$ be the largest rank such that $\text{terms}(\alpha(I_c, \Sigma)) \setminus \text{terms}(\alpha(I_c, \Sigma))$ is empty. Every instance $I$ can be embedded into $I_c$. By Lemma 3 the existential depth of the terms in o-chase$(I, \Sigma)$ is bounded by that of o-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 $\Sigma$ a set of FE-rules, $\Sigma \in \text{CT}$ iff $\Sigma \in \text{BN}$. 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 $\Sigma = \{ \sigma_1 : p(x, y) \rightarrow \exists z \ r(z, y) \}$, $\sigma_2 : r(x, y) \rightarrow \exists z \ s(y, z) \}$. Then, so-chase$^2(I, \Sigma) = I \cup \{ r(z, (\sigma_1, \pi), b) \cup s(b, z, (\sigma_2, \pi)) \}$ with $\pi = \{ y \rightarrow b \}$. Also, so-chase$^2(I, \Sigma)$ for each fresh variable $v$ in $I$, the maximum frontier depth of $v$'s facts is $\text{fr}(\sigma)$. The frontier depth of so-chase$^2(I, \Sigma)$ is defined as the maximum frontier depth of its facts if it is finite and is infinite otherwise.

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 so-chase$(I, \Sigma)$ is

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

where $v_B$ is any term in $\pi(\text{fr}(\sigma))$ used by a trigger $(\sigma, \pi)$ which generates $v$. Accordingly, the frontier depth of a fact $f$ is the maximum frontier depth of its terms. The frontier depth of so-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, $\text{depth}_{\exists}^\pi(v) \leq \text{depth}_{\exists}(v)$ for all $v$ in o-chase$(I, \Sigma)$. The following example illustrates the difference between the two notions of (existential) depth.

**Example 4.** Let $\Sigma = \{ \sigma = p(x, y, z) \rightarrow \exists z \ p(y, x, z) \}$. Starting from $I = \{ p(a, b, c) \}$, the o-chase generates an infinite number of fresh variables $v$ with increasing depth$_{\exists}(v)$.

The rank of the so-chase is instead 2 and for each fresh variable $v$, depth$_{\exists}^\pi(v) = 1$ as all triggers map fr$(\sigma)$ to terms$(I)$.

It is worth noting that not only the oblivious notion of existential 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 $\varphi$ from $I$ to $I'$ and any $i \geq 0$, there exists an embedding $\varphi' \supseteq \varphi$ from so-chase$(I, \Sigma)$ to so-chase$(I', \Sigma)$ which preserves the frontier depth of terms.

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

### 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.\footnote{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.} We define the rank of a chase sequence on \((I, \Sigma)\) 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 \((I, \Sigma)\) if and only if all chase sequences for \((I, \Sigma)\) 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 sequences 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 \(\sigma_2\) are performed before the rule \(\sigma_1\).

In the special case of FE-rules, it is not hard to see that all oblivious chase sequences for \((I, \Sigma)\) 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 \(\sigma_1 = p(x, y, t) \land p(y, z, u) \rightarrow \exists v p(x, z, v)\) and \(\sigma_2 = p(x, y, t) \land p(w, z, u) \rightarrow \exists v p(x, z, v)\). The rank of any chase \((I, \Sigma)\) is bounded by 2 for any \(I\), while again performing all applications of \(\sigma_2\) before \(\sigma_1\) 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 \(fr(\sigma)\) 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 \((Q, \Sigma)\) is **FO-rewritable** (resp. UCQ-rewritable) if there is an FO-query (resp. a UCQ) \(Q\) such that, for all \(I\), the certain answers to \(Q\) on \((I, \Sigma)\) are exactly the answers to \(Q\) on \(I\). It is known that FO-rewritability is equivalent to UCQ-rewritability.\footnote{It follows from the (Finite) Homomorphism preservation theorem, a classical result in model theory [Rossman, 2008].} A set of rules \(\Sigma\) is **FO-rewritable** (or equivalently, UCQ-rewritable) if \((Q, \Sigma)\) 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 \(\Sigma\) is FO-rewritable with respect to \(\mathcal{C}\) if \((Q, \Sigma)\) is FO-rewritable for all \(Q \in \mathcal{C}\). We denote by FO-R\(\mathcal{C}\) the corresponding class. We first point out that FO-rewritability with respect to full-atomic queries, denoted by FO-R\(\mathcal{AF}\), is a strictly weaker property than FO-rewritability.

**Proposition 8.** FO-R\(\mathcal{AF}\) ⊆ FO-R

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

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

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

**Proof.** The number of (non-isomorphic) full-atomic queries to be considered is finite, as for Proposition 9. We take for \(k_{\text{AF}}\) 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 \(\Sigma \in \text{FO-R}^{\mathcal{AF}}\) (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...
terminates). Therefore, we give a different property for the so-chase, which requires the power of FO-rewritability.

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

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

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

**Theorem 13.** If \( \Sigma \in \text{FO-R} \) then for all instance \( I \) and fact \( f \in \text{o-chase}(I, \Sigma) \) we have that \( \text{rank}(f) \leq \text{depth}_{\text{BN}}(f) \times (k_{\text{BN}} + 1) + k_{\text{BN}} \) with \( k_{\text{BN}} \) 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\(^0\), the notions of FO-R and FO-R\text{AF} coincide.

**Theorem 14.** FO-R\text{AF} \cap CT\(^0\) = BN\(^0\) = FO-R \cap CT\(^0\)

**Proof.** We start by showing that BN\(^0\) \subseteq FO-R \cap CT\(^0\). By definition BN\(^0\) \subseteq CT\(^0\). Then, BN\(^0\) \subseteq FO-R follows from the equivalence between FO-R and the bounded-depth derivation property [Gottlob et al., 2014]. Moreover, by Proposition 8 we have BN\(^0\) \subseteq FO-R\text{AF} \cap CT\(^0\). To conclude the proof, by Theorem 4 and 12 we have that FO-R\text{AF} \cap CT\(^0\) \subseteq BN\(^0\) and again by Proposition 8 follows FO-R \cap CT\(^0\) \subseteq BN\(^0\). \( \square \)

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

**Theorem 15.** BN\(^0\) = FO-R \cap CT\(^{20}\)

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) equivalent to o-chase termination for FE-rules. Intuitively, when a set of rules \( \Sigma \) is decomposed into 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 \( \sigma \) in \( \Sigma \) we assign a special predicate \( p_{\sigma} \) of arity \( \text{fr}(\sigma) \). \( \Sigma(\sigma) \) is obtained from \( \Sigma \) by replacing each rule \( \sigma = B \rightarrow H \) with two rules: a datalog rule \( B \rightarrow p_{\sigma}(\text{fr}(\sigma)) \) and a rule \( p_{\sigma}(\text{fr}(\sigma)) \rightarrow H \). It can be shown that \( \Sigma \in \text{CT}^{20} \) iff \( \Sigma(\sigma) \in \text{CT}^{20} \) and that \( \Sigma \in \text{BN}^{0} \) iff \( \Sigma(\sigma) \in \text{BN}^{0} \). 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\(^0\) and BN\(^{20}\). A notable class of datalog rules with decidable boundedness (more precisely in linear time) is chain datalog [Guexsian and Peixoto, 1994]. We obtain that membership to BN\(^0\), CT\(^0\) and FO-R remains undecidable for FE-rules, while the decidability of membership to BN\(^0\), hence to CT\(^0\), is still open.\(^3\)

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 \( C \in \text{FO-R} \), it holds that: \( C \in \text{BN}^{0} \) iff \( C \in \text{CT}^{0} \), and \( C \in \text{BN}^{0} \) iff \( C \in \text{CT}^{20} \).

This implies that membership to BN\(^0\) and BN\(^{20}\) is PSpace-complete for the two main classes of FO-rewritable existential rules, namely linear and sticky. Indeed, deciding CT\(^0\) and CT\(^{20}\) 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\(^0\) and BN\(^{20}\) for a major class of existential rules, namely guarded. This class is neither CT\(^{20}\) nor FO-R. However, membership to CT\(^0\) and CT\(^{20}\) 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 breath-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 \( \Sigma \) and a (unary encoded) integer \( k \), does it hold that \( +\text{chase}^{k}(I, \Sigma) = +\text{chase}(I, \Sigma) \) for all instance \( I \)?

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

\(^{3}\)See Proposition 18 in the Appendix.
**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 terminates). Theorem 13 suggests that whenever $\Sigma \in \text{FO-R}$ if any such chase generates only terms of bounded $\text{frontier depth}$ on all instances, then $\Sigma$ 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**


[Grau et al., 2013] Bernardo Cuenca Grau, Ian Horrocks, Markus Krötzsch, Clemens Kupke, Despoina Magka,


**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** \(\Sigma \in CT^o\) iff \(DF(\Sigma) \in CT^o\) and \(\Sigma \in BN^o\) iff \(DF(\Sigma) \in BN^o\).

*Proof.* The proposition is immediate since, for any instance \(I\) and chase step \(i\), \(o\)-chase\(^-(I, \Sigma)\) = \(o\)-chase\(^-(I, DF(\Sigma))\). \(\square\)

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

**Proof of Proposition 2** If \(\Sigma\) is a set of FE-rules then, for all instance \(I\) and term \(v\) in \(o\)-chase\(^-(I, \Sigma)\), holds that \(\text{depth}_2(v) = \text{rank}(v)\).

*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 \(\text{rank}(f) = i\) then \(\text{depth}_2(f) = i\). The property obviously holds for \(i = 0\). Let \(i > 0\) and \(\text{rank}(f) = i\). By definition of rank, \(f\) was produced from at least one fact \(f'\) of rank \(i - 1\). By induction hypothesis, \(\text{depth}_2(f') = i - 1\), hence, by definition of existential depth, \(f'\) contains a term \(t\) with \(\text{depth}_2(t) = i - 1\). Since all rules are FE-rules, \(f\) contains at least one fresh variable (null) \(v\), and, by definition of existential depth, \(\text{depth}_2(v) = 1 + (i - 1) = i\). Hence, \(\text{depth}_2(f) = i\).

Now, let \(t\) be a term with rank \(i\). If \(i = 0\), \(t\) occurs in \(I\) and \(\text{depth}(t) = 0\). Otherwise, \(t\) has been generated in a fact \(f\) of rank \(i\). Since \(\text{depth}_2(f) = \text{rank}(f)\), \(\text{depth}_2(f) = i\) and, by definition of existential depth, all terms generated in \(f\) have existential depth \(i\), in particular \(t\). \(\square\)

**Proof of Lemma 3** For any embedding \(\varphi\) from \(I\) to \(I'\) and any \(i \geq 0\), there exists an embedding \(\varphi' \supseteq \varphi\) from \(o\)-chase\(^-(I, \Sigma)\) to \(o\)-chase\(^-(I', \Sigma)\) which preserves the existential depth of terms, i.e., for every term \(v\) in \(o\)-chase\(^-(I, \Sigma)\) it holds that \(\text{depth}_2(v) = \text{depth}_2(\varphi'(v))\).

*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 \(\varphi' : o\)-chase\(^{n-1}(I, \Sigma) \rightarrow o\)-chase\(^{n-1}(I', \Sigma)\) preserving the existential depth of terms. Let \((\sigma, \pi)\) be any trigger of \(o\)-chase\(^{n-1}(I, \Sigma)\). We know that \(\pi \circ \sigma(\text{body}(\sigma)) \subseteq o\)-chase\(^{n-1}(I', \Sigma)\) and \((\sigma, \pi)\) is a trigger of \(o\)-chase\(^{n-1}(I', \Sigma)\). Also, there exists a bijection \(\rho_n\) from the fresh terms in \(\pi_0(\text{head}(\sigma))\) to the fresh terms in \(\varphi' \circ \pi_0(\text{head}(\sigma))\) precisely defined as \(\rho_n(z_{(\sigma, \pi)}) = z_{(\varphi', \sigma' \circ \pi')}(\sigma)\). Let \(\varphi'' = \varphi' \cup \rho_n\) be the natural extension of \(\varphi'\) to all triggers that are performed to compute \(o\)-chase\(^n(I, \Sigma)\). Of course, for every trigger \((\sigma, \pi)\) and term \(v_B \in \text{terms}(\pi(\text{body}(\sigma)))\) we have that \(\text{depth}_2(v_B) = \text{depth}_2(\varphi''(v_B))\). We want to show that \(\varphi''\) also preserves the existential depth of fresh terms. Consider now the rule application \((\sigma, \varphi'' \circ \pi)\). Let \(z\) be an existential variable of \(\sigma\). Then, \(\text{depth}_2(z_{(\sigma, \pi)}) = 1 + \max\{\text{depth}_2(v_B)\} = 1 + \max\{\text{depth}_2(\varphi''(v_B))\} = \text{depth}_2(z_{(\sigma, \varphi' \circ \pi)})\). \(\square\)

**Proof of Lemma 6** For any embedding \(\varphi\) from \(I\) to \(I'\) and any \(i \geq 0\), there exists an embedding \(\varphi' \supseteq \varphi\) from \(so\)-chase\(^-(I, \Sigma)\) to \(so\)-chase\(^-(I', \Sigma)\) 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 \(\varphi' : so\)-chase\(^{n-1}(I, \Sigma) \rightarrow so\)-chase\(^{n-1}(I', \Sigma)\).
which preserves the frontier depth of values. Let $(\sigma, \pi)$ be any trigger producing a new fact $f \in \text{so-chase}^n(I, \Sigma)$. Then $\varphi' \circ \pi(\text{body}(\sigma)) \subseteq \text{so-chase}^{n-1}(I', \Sigma)$.

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

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

Proof of Theorem 7 When $\Sigma \in \text{CT}_{\text{FO}}$ there exists a constant $k_3$ such that for all instance $I$, the frontier depth of a term in so-chase$^I(I, \Sigma)$ is bounded by $k_3$.

Proof. If $\Sigma$ is in $\text{CT}_{\text{FO}}$, the so-chase terminates on the critical instance. We take for $k_3$ the smallest rank such that so-chase$^{k_3}(I_n, \Sigma) = \text{so-chase}(I_n, \Sigma)$. Every instance can be embedded into the critical instance. Hence, by Lemma 6 the frontier depth of the terms in so-chase$^{k_3}(I_n, \Sigma)$ is bounded by the frontier depth of so-chase$^I(I_n, \Sigma)$, which is itself bounded by $k_3$.

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

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

When a query is rewritten, some answer variables may be made equal. Hence, we slightly generalize the notion of query $Q(x_1, \ldots, x_k)$ 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 $i$th and $j$th answer variables of $Q$ are the same. Given a class $C$ in $P_Q$, we denote by $z_C$ 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 $p(x, y)$ 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 $Q(x_1, x_2) = p(x_1, x_2)$. 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 $P_1 \lor P_2$ the upper bound of $P_1$ and $P_2$ in this lattice.

We recall that $(Q, \Sigma)$, with $Q$ a CQ, is FO-rewritable iff there is a UCQ-rewriting of $Q$, i.e., a finite set $Q$ of CQs such that, for any instance $I$, the set of certain answers to $Q$ on $(I, \Sigma)$ is exactly the set of answers to the UCQ obtained from $Q$ on $I$. Each CQ in $Q$ can be obtained from $Q$ and $\Sigma$ 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 $s \in \Sigma$ that unifies a non-empty subset of $Q_i$ with a subset of $s$’s head while satisfying conditions concerning existential variables in $s$. In particular, an answer variable of $Q_i$ cannot be unified with an existential variable of $s$. The following property holds: for any $(I, \Sigma)$ and CQ $Q$, a tuple of constants $(a_1, \ldots, a_k)$ is a certain answer to $Q$ on $(I, \Sigma)$ 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, $Q_i \sqsubseteq P_Q$, holds ($P_Q$ is thinner than $P_{Q_i}$).

Now, assume $\Sigma \in \text{FO-R}\text{HD}_\Sigma$ and let $Q$ be a full-atomic query. If $Q$ is not unifiable with a datalog rule from $\Sigma$, 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 $(Q, \Sigma)$ is FO-rewritable, as the union of the UCQ-rewritings of all $(u(Q_h), \Sigma)$ 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 $P_Q \sqsubseteq P_{Q_s}$; then $(Q, \Sigma)$ is FO-rewritable then $(Q_s, \Sigma)$ also is. By hypothesis, each $(Q_h, \Sigma)$ 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 $Q'$, there is a rewriting sequence of the same length leading from $Q_s$ to a query $Q_s'$, such that $P_{Q_s'} = P_Q \lor P_{Q_s}$, where $\lor$ is the upper bound in the partition lattice), and, given $s'$ the substitution of the answer variables in $Q'$ by the answer variables in $Q'_s$, such that $P_{Q_s'} \sqsubseteq P_{Q_s}$, it holds that $s'(Q') = Q'_s$, 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 $(Q, \Sigma)$ is FO-rewritable. The FO-rewritability of $(Q, \Sigma)$ is equivalent to the
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 \((I,\Sigma)\) if and only if there is a query \( Q' \) obtained by a rewriting sequence from \( Q \) of length less than \( b \), with \((a_1,\ldots,a_k)\) is an answer to \( Q' \) on \( I \). We prove that \((Q,\Sigma)\) 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 \((I,\Sigma)\). There is thus a homomorphism \( h \) from \( Q_s \) to \( \star\cdot\text{chase}(I,\Sigma) \) 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 \( \star\cdot\text{chase}(I,\Sigma) \) 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 \( \{1,\ldots,k\} \) associated with the equality of terms in \((a_1,\ldots,a_k)\). We have \( P_{Q^b} \subseteq P_a \) and \( P_{Q_s} \subseteq P_a \). By Lemma (L), there is a query \( Q^b \) obtained from \( Q_s \) with a rewriting sequence of length less than \( b \), such that (1) \( P_{Q^b} = P_{Q^b} \cup P_{Q_s} \), and (2) \( s'(Q^b) = Q^b \), with \( s' \) the substitution associated with \( P_{Q_b} \subseteq P_{Q^b} \). From (1), we have \( P_{Q_s} \subseteq P_a \). Hence, the homomorphism \( h^b \) from \( Q^b \) to \( I \) can be written \( h' \circ s' \), where \( h' \) is a homomorphism from \( Q^b \) to \( I \) mapping its answer tuple to \((a_1,\ldots,a_k)\). The converse direction ("if there is \( Q^b \) obtained by a rewriting sequence from \( Q_s \), of length less than \( b \), with \((a_1,\ldots,a_k)\) is an answer to \( Q^b \) on \( I \), then \((a_1,\ldots,a_k)\) is a certain answer to \( Q_s \) on \((I,\Sigma)\)') holds because of the soundness of query rewriting based on piece-unifiers.

\[ Q = p(x_1,\ldots,x_n) \] 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 \( p(a_1,\ldots,a_n) \in \text{o-chase}^{\text{AF}}(I,\Sigma) \). Now, the statement of the lemma considers more generally \( f \) with \( \text{terms}(f) \subseteq \text{terms}(I) \) which could contain also existentially quantified variables. Observe however that, for any instance \( I \), let \( \text{freeze()} \) 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 \( \text{o-chase}^i(I,\Sigma) \) to \( \text{o-chase}^i(\text{freeze}(I),\Sigma) \) that preserves the rank of facts. Because of this, let \( f \) and \( I \) be any fact and instance, we have that \( f \in \text{o-chase}^i(I,\Sigma) \) (for any \( i \)) iff \( \text{freeze}(f) \in \text{o-chase}^i(\text{freeze}(I),\Sigma) \). Since \( \text{freeze}(f) \) contains only constants we can conclude.

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

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

\[ \text{We know that the number of full-atomic queries to be considered is finite and by Proposition 9 can be even bounded by the number of non-isomorphic heads of datalog rules. By the properties of breadth-first query rewriting, we know that for any instance } I \text{ and any full-atomic query } Q, \text{ the certain answers to } Q \text{ on } (I,\Sigma) \text{ are exactly the answers to } Q \text{ on } \text{o-chase}^{k_{\text{AF}}}(I,\Sigma). \text{ We can identify an answer } (a_1,\ldots,a_n) \]

\[ \text{Alternatively, we could rely on the bound given by the bounded derivation-depth property (BDDP) [Call et al., 2009b]. A ruleset } \Sigma \text{ satisfies the BDDP if for all Boolean CQ } Q, \text{ there is an integer } k \text{ such that, for all instance } I, \text{ it holds that } I,\Sigma \models Q \text{ iff } \text{o-chase}^i(I,\Sigma) \models Q. \text{ It has been several times remarked that BDDP is equivalent to UCQ-rewritability, hence to FO-rewritability.} \]
Proof of Theorem 12. If $Σ ∈ FO-RAF$ then for all instance $I$ and fact $f ∈ o$-chase$(I, Σ)$ we have that $\text{rank}(f) ≤ \text{depth}_Ω(f) × (k_{AF} + 1) + k_{AF}$ with $k_{AF}$ the bound provided by Lemma 10.

Proof. We first show that since $Σ ∈ FO-RAF$ then for all instance $I$ and term $v ∈ \text{terms}(o$-chase$(I, Σ))$ it holds that $\text{rank}(v) ≤ \text{depth}_Ω(v) × (k_{AF} + 1)$, where we recall that $\text{rank}(v)$ is the rank where $v$ is introduced.

By induction on the existential depth of $v$. If $\text{depth}_Ω(v) = 0$ then $v ∈ \text{terms}(I)$ and thus $\text{rank}(v) = 0$ also. Assume the property holds for $0 ≤ \text{depth}_Ω(v) ≤ n$. We show that it holds for $\text{depth}_Ω(v) = n + 1$. Let $(σ, π)$ be the trigger that generates $v$. Then, for all $v_B ∈ \text{terms}(π(\text{body}(σ)))$, we know that $\text{depth}_Ω(v_B) ≤ n$. By inductive hypothesis, $\text{rank}(v_B) ≤ n × (k_{AF} + 1)$. Hence for all $f_B ∈ \text{π}(\text{body}(σ))$ it holds that $\text{rank}(f_B) ≤ k + k_{AF}$. Thus $\text{rank}(v) ≤ k + k_{AF} + 1 = (n + 1) × (k_{AF} + 1) = \text{depth}_Ω(v) × (k_{AF} + 1)$. To conclude the proof, since any fact $f ∈ o$-chase$(I, Σ)$ contains only terms $v$ with $\text{rank}(v) ≤ \text{depth}_Ω(v) × (k_{AF} + 1)$, we apply Lemma 10 and we obtain $\text{rank}(f) ≤ \max\{\text{depth}_Ω(v) | v ∈ \text{terms}(f)\} × (k_{AF} + 1) + k_{AF}$.

Proof of Theorem 13. If $Σ ∈ FO-R$ then for all instance $I$ and fact $f ∈ \text{so$-$chase}(I, Σ)$ we have that $\text{rank}(f) ≤ \text{depth}_Ω(f) × (k_{FO} + 1) + k_{FO}$ where $k_{FO}$ is the bound provided by Lemma 11.

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

Since any fact $f ∈ \text{so$-$chase}(I, Σ)$ contains only terms $v$ with $\text{rank}(v) ≤ \text{depth}_Ω(v) × (k_{FO} + 1)$, we use again Lemma 11, and obtain $\text{rank}(f) ≤ \text{depth}_Ω(f) × (k_{FO} + 1) + k_{FO}$.

Proof of Theorem 14. $BN^0 = FO-RAF \cap CT^0$.

Proof. $⇒$ If $Σ ∈ BN^0$ there is $k$ such that, for all $I$, $\text{o$-$chase}(I, Σ) = \text{o$-$chase}(I, Σ)$. Hence, there is $k$ such that for all $I$ and all Boolean CQ $Q$ we have $Σ, I ⊨ Q$ (i.e., $\text{o$-$chase}(I, Σ) ⊨ | Q |$ iff $\text{o$-$chase}(I, Σ) ⊨ | Q |$ iff $I ⊨ Q^k$, 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, $BN^0 ⊆ CT^0$ follows by definition.

$⇐$ By Theorem 4 and 12.

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

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 $BN^{or}$, CT$^{or}$ and FO-R remains undecidable for FE-rules.

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

The proposition leads directly to the undecidability of $BN^{or}$ and CT$^{or}$ 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$^{or}$ if and only if $ψ(Σ)$ is in CT$^{or}$. Since every datalog set is CT$^{or}$, $ψ(Σ)$ is also CT$^{or}$. Now, consider the (undecidable) problem of whether $Σ$ is (uniformly) bounded. We have that $Σ$ is bounded iff $ψ(Σ)$ is bounded, which amounts to asking if $ψ(Σ)$ is FO-R (as we already know it is in CT$^{or}$).

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

Obviously, $ψ$ is injective (Point (1)).

To prove the point (2), we show that for each fact $f = p(v_1, ..., v_n) ∈ \text{so$-$chase}(I, Σ)$ generated by a trigger $(σ, π)$ it holds $p^+(v_1, ..., v_n, z_σ(ψ(σ), π(σ))) ∈ \text{so$-$chase}(Ψ(I), Ψ(Σ))$, and vice-versa.

We focus on the direction $⇒$ as the direction $⇐$ is similar. By induction on the rank $k$ of the so-chase. If $k = 0$ then by definition $f ∈ I$ implies $ψ(f) ∈ ψ(I)$. Assume that the property holds for $0 ≤ i ≤ n$. 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 $f_B = p(x_1, ..., x_k) ∈ \text{body}(σ)$ we know that $π(f_B) ∈ \text{so$-$chase}(I, Σ)$. Hence, by induction $ψ(π(f_B)) = p^+(v_1, ..., v_n, z^+)$ is in $\text{so$-$chase}(Ψ(I), Ψ(Σ))$ where $z^+ = z_σ(ψ(f_B), f_B)$ if $f_B ∈ I$, or $z^+ = z_σ(ψ(f_B), f_B)$. This follows from any KB $Ψ(I)$ being generated by a trigger $(σ', π')$. Then, $ψ(σ, π) = ψ(σ, π) ∪ f_B$ is applicable and produces $ψ(f)$.

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 $ψ(V)$ to an instance $I$ on $V$ such that
so-chase\( (P, \psi(\Sigma)) \) and so-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 \( \phi \) assigns to each atom \( \alpha = p^+(v_1, \ldots, v_n, z) \) on \( \psi(V) \) 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 \( f = p^+(v_1, \ldots, v_n, z) \in \text{so-chase}(I, \psi(\Sigma)) \) generated by a trigger \( (\psi(\sigma), \pi) \), it holds that \( p(v_1, \ldots, v_n) \in \text{so-chase}(I, \Sigma) \). 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 \( f_B = p^+(v_1, \ldots, v_k, z) \in \text{body}(\psi(\sigma)) \) we know that \( \pi(f_B) \in \text{so-chase}(I', \psi(\Sigma)) \). Hence, by induction \( \phi(\pi(f_B)) \in \text{so-chase}(I, \Sigma) \). Then, the trigger \( (\sigma, \pi') \) (where \( \pi' \) is the appropriate restriction of \( \pi \) 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 \( V \). Let us note that \( \phi(\psi(I)) = I \). By what precedes so-chase\( (I, \Sigma) \) and so-chase\( (\psi(I), \psi(\Sigma)) \) have the same rank. Now, let an instance \( I' \) on \( \psi(V) \) and \( I = \phi(I') \); by what precedes the rank of so-chase\( (I, \Sigma) \) is at least the rank of so-chase\( (I', \psi(\Sigma)) \). Furthermore, by embedding \( \psi(I) \) in \( I' \), we can by using similar arguments prove that the rank of so-chase\( (I', \psi(\Sigma)) \) is at least the rank of so-chase\( (I, \Sigma) \). So, so-chase\( (I', \psi(\Sigma)) \) and so-chase\( (I, \Sigma) \) have the same rank.

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. \( P^n_0 \)) is their respective distinguished 0-ary predicate. As they are non-recursive, \( Q_1 \) (resp. \( Q_2 \)) is \( k_1 \)-bounded (resp. \( k_2 \)) 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 \( (Q_1, Q_2) \). We 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. \( P_0 \rightarrow P_1 \)) 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 \( P_0 \) can not be derived from \( (I, Q_2) \). As \( P_0 \) is not generated by the \( Q_2 \) part, \( P_i \) 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 \( Q_2 \).