TPLP **21** (1): 51–79, 2021. © The Author(s), 2020. Published by Cambridge University Press.
doi:10.1017/S1471068420000083 First published online 4 June 2020

Characterizing Boundedness in Chase Variants

STATHIS DELIVORIAS, MICHEL LECLÈRE, MARIE-LAURE MUGNIER and FEDERICO ULLIANA
University of Montpellier, LIRMM, CNRS, INRIA, Montpellier, France
(*e-mails:* jazzcrazysta@yahoo.com, mugnier@lirmm.fr, leclere@lirmm.fr, federico.ulliana@lirmm.fr)

submitted 8 April 2019; accepted 11 May 2020

Abstract

Existential rules are a positive fragment of first-order logic that generalizes function-free Horn rules by allowing existentially quantified variables in rule heads. This family of languages has recently attracted significant interest in the context of ontology-mediated query answering. Forward chaining, also known as the chase, is a fundamental tool for computing universal models of knowledge bases, which consist of existential rules and facts. Several chase variants have been defined, which differ on the way they handle redundancies. A set of existential rules is bounded if it ensures the existence of a bound on the depth of the chase, independently from any set of facts. Deciding if a set of rules is bounded is an undecidable problem for all chase variants. Nevertheless, when computing universal models, knowing that a set of rules is bounded for some chase variant does not help much in practice if the bound remains unknown or even very large. Hence, we investigate the decidability of the k-boundedness problem, which asks whether the depth of the chase for a given set of rules is bounded by an integer k. We identify a general property which, when satisfied by a chase variant, leads to the decidability of k-boundedness. We then show that the main chase variants satisfy this property, namely the oblivious, semi-oblivious (aka Skolem), and restricted chase, as well as their breadth-first versions.

KEYWORDS: knowledge representation and nonmonotonic reasoning, logic programming methodology and applications, databases and semantic web reasoning, theory

1 Introduction

Existential rules (see Calì *et al.* (2008), Baget *et al.* (2009), and Calì *et al.* (2009) for the first papers and Gottlob *et al.* (2012) and Mugnier and Thomazo (2014) for introductory courses) are a positive fragment of first-order logic that generalizes function-free Horn rules, such as the deductive database query language Datalog, and knowledge representation formalisms such as Horn description logics (see Calvanese *et al.* 2005; Baader *et al.* 2005; Krötzsch *et al.* 2007). More specifically, existential rules are of the form $body \rightarrow head$, where *body* and *head* are conjunctions of atoms (without functions), and variables that occur only in the head are existentially quantified. These existentially quantified variables allow one to assert the existence of unknown individuals, a key feature for reasoning on incomplete data with the open-domain assumption.

Existential rules have the same logical form as the general database constraints known as tuple-generating dependencies, which have long been investigated in database theory (Abiteboul *et al.* 1995). Reborn under the names of existential rules, Datalog$^+$ or

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Datalog$^{\exists}$, they have attracted significant interest in the last years as ontological languages, especially for ontology-mediated query answering and ontology-based data integration.

In our setting, a knowledge base (KB) is composed of a set of existential rules, which typically encodes ontological knowledge, and a factbase, a.k.a. instance, which contains factual data. A factbase is a set of atoms built from constants and variables, the variables representing unknown individuals, also called (labeled) *nulls* in databases. The logical translation of a factbase is an existentially closed conjunction of atoms. In this paper, we focus on the forward chaining process for reasoning on KBs, which consists of iteratively extending the factbase with new facts produced by rule applications, until we reach a fixpoint. In the forward chaining process, a rule of the form $body \rightarrow head$ can be applied to a factbase F whenever there is a homomorphism h from $body$ to F. The factbase F is then extended with new atoms obtained by first applying h as a substitution to $head$ and then by renaming each existentially quantified variable (which has no image in h) with a fresh variable, that is, that does not occur in F. Hence, a rule application may produce new (unknown) individuals, that is, nulls. This is illustrated by the following example.

Example 1
Consider the KB $\mathcal{K} = (F, \mathcal{R})$ where $F = \{Human(Alice)\}$ is the factbase and $\mathcal{R}$ is the ruleset containing the single rule $R = \forall x.\ Human(x) \rightarrow \exists y.\ parentOf(y, x) \wedge Human(y)$. Here, R can be applied to F because of the homomorphism $\{x \mapsto Alice\}$. This produces the atoms $parentOf(y_0, Alice)$ and $Human(y_0)$ containing a fresh variable y_0 introduced by safely renaming y in the rule. The factbase resulting from the rule application is therefore $\{Human(Alice), parentOf(y_0, Alice), Human(y_0)\}$, which stands for the logical formula $\exists y_0.\ Human(Alice) \wedge parentOf(y_0, Alice) \wedge Human(y_0)$. Note that R can now be applied again by mapping x to y_0 thereby creating a new individual y_1 and producing the atoms $parentOf(y_1, y_0)$ and $Human(y_1)$. In this example, the forward chaining does not terminate.

Forward chaining with existential rules is also known as the *chase* in databases and a considerable literature has been devoted to its analysis in the context of tuple-generating dependencies (Beeri and Vardi 1984; Fagin *et al.* 2005; Deutsch *et al.* 2008; Marnette 2009; Grahne and Onet 2018). As illustrated by Example 1, the chase may not terminate on a given KB. However, a fundamental property of the chase is that it computes a *universal model* of the KB, that is, a model that homomorphically maps to any other model of the KB (Deutsch *et al.* 2008). This has a major implication in problems like answering ontology-mediated queries, since a Boolean conjunctive query (i.e., an existentially closed conjunction of atoms) is logically entailed by a KB if and only if it homomorphically maps to the result of the chase.

Several variants of the chase have been studied, mainly: the *oblivious* chase (Calì *et al.* 2008), the *Skolem* chase (Marnette 2009), the *semi-oblivious* chase (Marnette 2009), the *restricted* or standard chase (Fagin *et al.* 2005), and the *core* chase (Deutsch *et al.* 2008) (and its variant, the equivalent chase; Rocher 2016).[1] All these chase variants compute logically equivalent results. Nevertheless, they differ on their ability to detect logical

[1] We could also consider the *parsimonious chase*, which was introduced in Leone *et al.* (2012), but it is tailored for answering atomic queries and does not compute a universal model of the KB, hence it is outside the family of chase variants studied here.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

redundancies possibly caused by the presence of nulls. Indeed, a factbase that contains nulls may be logically equivalent to one of its strict subsets, in which case we call it redundant. As deciding whether a factbase is redundant is computationally difficult (in fact, NP-complete; Chandra and Merlin 1977), a given chase variant may choose to detect only specific cases of redundancy. Note that, since redundancies can only be due to nulls, all chase variants output exactly the same results on ground factbases and rules without existential variables (i.e., Datalog rules, also called range-restricted rules; Abiteboul *et al.* 1995). On the other hand, the ability to detect redundancies has a direct impact on the termination of the chase. In short, the oblivious chase blindly performs all possible rule applications and terminates less often than the other variants, while the core chase produces factbases with no redundancies and terminates exactly when the KB admits a finite universal model. The rest of the chase variants lie between these two extremes; they are presented in the next section. The following example aims at illustrating the behavior of the oblivious and core chase variants. For brevity, universal quantifiers will be omitted.

Example 2
Take the ruleset $\mathcal{R} = \{p(x,y) \rightarrow \exists z.p(y,z) \wedge p(z,y)\}$ and the factbase $F = \{p(a,b)\}$. Both chase variants perform the first rule application, which yields $F_1 = \{p(a,b), p(b,z_0), p(z_0,b)\}$. Then, two new applications of R are possible, one for each new atom $p(b,z_0)$ and $p(z_0,b)$. The first application would then add the atoms $p(z_0,z_1)$ and $p(z_1,z_0)$ to F_1, yielding F_2. However, F_2 is logically equivalent to F_1 as there is a homomorphism from F_2 to F_1, which maps z_1 to b and z_0 to itself. A similar thing happens for the second rule application which would create the atoms $p(b,z_2)$ and $p(z_2,b)$. These are again redundant as z_2 can be mapped to z_0. The oblivious chase simply performs both rule applications without testing for redundancy. This also means that it will go on forever, since each new (redundant) atom brings new rule applications. In contrast, the core chase detects that any rule application on F_1 yields an equivalent factbase, and outputs F_1. It is also worth mentioning that in this case the semi-oblivious chase will behave similarly as the oblivious chase, while the restricted chase will behave similarly as the core chase.

Chase variants can be totally ordered with respect to the inclusion of the sets of KBs on which they terminate:

$$\text{Oblivious} < \text{Semi-oblivious} < \text{Restricted} < \text{Core}.$$

Here, $X_1 < X_2$ means that when X_1 halts on a KB, so does X_2, and there are KBs for which the converse is false. Furthermore, the Skolem chase (respectively, the equivalent chase) terminate on the same KBs as the semi-oblivious chase (respectively, the core chase). Of course, none of the chase variants terminates on a KB that does not admit a finite universal model as shown in Example 1. The termination problem, which asks whether for a given set of rules the chase will terminate on any factbase, is undecidable for all these chase variants (Deutsch *et al.* 2008; Baget *et al.* 2010; Gogacz and Marcinkowski 2014).

Inspired by previous work on Datalog (see e.g., Cosmadakis *et al.* 1988; Gaifman *et al.* 1993), we study the related problem of *boundedness*, which asks if, given a set of rules, the chase terminates on any factbase within a *predefined* depth, that is, independent from any

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

factbase. The notion of depth is natural for Datalog programs whose evaluation is defined in a breadth-first manner, but requires further technical definitions when addressing several chase variants, which will be presented later. Hence, we will focus here on breadth-first chase variants. Intuitively speaking, a breadth-first chase is a process that, starting from a factbase $F = F^0$, proceeds as follows: for each step $i > 0$, it (1) computes all new homomorphisms from rule bodies to F^{i-1} and then (2) performs the rule applications associated with these homomorphisms, according to its own redundancy criterion, which yields F^i. The depth of a breadth-first chase on a given KB corresponds to the number of steps needed to terminate, that is, k if the last computed factbase is F^k.

Given a chase variant X, we call a set of rules X-*bounded* if there is a bound k such that, for *every* factbase, the X-chase stops within depth at most k. Of course, since chase variants differ with respect to termination, they also differ with respect to boundedness, and each variant raises a distinct notion of boundedness.

Boundedness implies several nice properties. First, if a set of rules is X-bounded with bound k, then, for any factbase F, the factbase obtained from F after k breadth-first X-chase steps is a universal model of the KB; the converse is also true for X being the core chase (or its variant the equivalent chase). Moreover, boundedness also implies decidability of fundamental static analysis tasks on queries, that is, data-independent problems whose aim is to decide semantic properties of queries that can be exploited to optimize query answering. It ensures the first-order rewritability property (Calvanese *et al*. 2007) also known as finite unification set property (Baget *et al*. 2011): any conjunctive query q can be rewritten using the set of rules $\mathcal{R}$ into a first-order query (and specifically a union of conjunctive queries) Q such that, for any factbase F, the answers to q on $(F, \mathcal{R})$ are exactly the answers to Q on F. Note that the conjunctive query rewriting procedure can be designed in a such a way that it terminates within k breadth-first steps, with k the bound for the core chase (Leclère *et al*. 2016). In turn, the first-order rewritability property ensures the decidability of conjunctive query containment under existential rules, which, given two conjunctive queries, asks if the set of answers to the first query is included in the set of answers to the second query, for any factbase.

The importance of the boundedness problem has been recognized already for rules without existential variables. Indeed, the problem has been first posed and studied for Datalog, where it has been shown to be undecidable (Hillebrand *et al*. 1995; Marcinkowski 1999). Example 3 illustrates some cases of bounded and unbounded rulesets in this setting.

Example 3

Consider the rulesets $\mathcal{R}_1 = \{R\}$ and $\mathcal{R}_2 = \{R, R'\}$, where $R = p(x, y) \wedge p(y, z) \rightarrow p(x, z)$ and $R' = p(x, y) \wedge p(u, z) \rightarrow p(x, z)$. The set $\mathcal{R}_1$ contains a single transitivity rule for the predicate p. This set is clearly unbounded as, for any integer k, there is a factbase $F = \{p(a_i, a_{i+1}) \mid 0 \leq i < 2^k\}$ that requires k chase steps. On the other hand, $\mathcal{R}_2$ also contains a rule that joins individuals on disconnected atoms. In this case, (i) if R generates some facts, then R' generates these same facts as well and (ii) R' needs to be applied only at the first step, for any F, as it does not produce any new atom at a later step. Therefore, $\mathcal{R}_2$ is bounded with the bound $k = 1$. Note that since these examples are in Datalog, the specificities of the chase variants do not play any role.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Example 2 illustrates boundedness beyond Datalog rules. The ruleset $\mathcal{R}$ is not bounded for the oblivious chase, which actually does not even terminate on the given KB. However, it can be checked that the core chase terminates with $\mathcal{R}$ on any factbase after at most one breadth-first step, hence $\mathcal{R}$ is bounded with $k = 1$ for the core chase.

Even though boundedness is undecidable already for Datalog, knowing that a set of rules is bounded for some chase variant does not help much in practice if the bound remains unknown or is very large. Hence, the goal of this paper is to investigate decidability of the *k-boundedness* problem which asks, for a given chase variant and an integer k, whether, for any factbase, the chase halts within depth k. The degree of boundedness of a ruleset (i.e., the smallest k such that this ruleset is k-bounded) can be seen as a measure of the "recursivity" of a ruleset. Given a factbase F, it yields a polynomial bound on the size of the chase output on F (i.e., with respect to the size of F). More precisely, the number of chase atoms produced at depth d is exponential in $|body|^d$, where $|body|$ is the maximal size of a rule body. Similarly, given a conjunctive query q, k-boundedness yields a polynomial bound on the size of a rewriting of q as a union of conjunctive queries. Moreover, since the theoretical bound on the chase output considers all possible factbases, the actual number of breadth-first steps required by the chase on a given factbase is expected to be much smaller in practice.

Our main contribution is to show that k-boundedness is indeed decidable for several main chase variants. It is worth noting that a general approach to derive decidability results for generally undecidable static analysis problems like boundedness is to restrict the rule language. Here, by focusing on the k-boundedness problem, we are able instead to obtain results for the whole existential rule language. Actually, we obtain a strong result by exhibiting a property that a chase variant may enjoy, namely *preservation of ancestry*, and prove that k-boundedness is decidable as soon as this property is satisfied. We then show that it is the case for the oblivious, semi-oblivious and restricted chases, as well as all their breadth-first versions. The decidability of k-boundedness for the core chase remains an open question.

The paper is organized as follows. Section 2 introduces preliminary notions, while Section 3 defines k-boundedness parametrized by the considered chase variant and states some fundamental properties of chase derivations on which we will rely to obtain our results. Section 4 presents the main results.

2 Preliminaries

This section is dedicated to basic notions and properties, including the formalization of the chase variants for which k-boundedness will be investigated.

2.1 Positive existential rules

We consider a first-order setting without functional symbols (except constants, which can be seen as 0-ary functional symbols) nor equality. A vocabulary is a finite set of predicates (usually denoted with the letters p, q, r) and constants (usually denoted with a, b, c). We assume also a countably infinite set of variables (denoted with x, y, z) used in formulas. A *term* is either a variable or a constant. Each predicate p is associated with

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

a positive integer number, called the *arity* of p. An *atom* is of the form $p(e_1, \ldots, e_n)$, where p is a predicate, n is the *arity* of p, and $e_1, \ldots, e_n$ are terms.

A *factbase*, denoted by F, is an existentially quantified conjunction of atoms which is closed, that is, every variable is existentially quantified. We will use the notation $var(F)$, $cnst(F)$, and $term(F)$, to refer to the set of variables, constants, and terms that occur in F. It is very convenient to see factbases simply as sets of atoms. So, for example, $\{p(a, x), q(x, b, c)\}$ can represent the existentially closed conjunction $\exists x.\, p(a, x) \wedge q(x, b, c)$.

A *substitution* σ is a mapping from a set of variables to a set of terms (usually represented as a set of single variable mappings, for example, $\{x \mapsto a, y \mapsto b, \ldots\}$). A *homomorphism* from a set of atoms F to F' is a substitution $h\colon var(F) \to term(F')$ such that $h(F) \subseteq F'$. It is known that a factbase F logically entails a factbase F' if and only if there exists a homomorphism from F' to F seen as atomsets (e.g., from Chandra and Merlin (1977)). An *isomorphism* is a bijective homomorphism. A subset $F' \subseteq F$ is a *retract* of F if there exists a substitution σ that is the identity on the terms of F' such that $\sigma(F) = F'$. In this case, σ is also called a retraction from F to F'. A factbase F is a *core* if none of its strict subsets is a retract.

An *existential rule* R is a first-order formula of the form $\forall \bar{x}.\, \forall \bar{y}.\, B(\bar{x}, \bar{y}) \to \exists \bar{z}.\, H(\bar{x}, \bar{z})$, where $\bar{x}$, $\bar{y}$, and $\bar{z}$ are disjoint sets of variables, and B and H are conjunctions of atoms called the *body* and the *head* of the rule, respectively. The set of variables $\bar{x}$ is shared by the body and the head of the rule; it is called the *frontier* of the rule, denoted by $fr(R)$. The set $\bar{z}$ is called the set of *existential variables* of the rule. The set $\bar{x} \cup \bar{y}$ is the set of *universally quantified variables* of R. An existential rule R is *Datalog* if it has no existential variables. We call a set of existential rules, denoted by $\mathcal{R}$, a *ruleset*.

In the following, universal quantifiers will be omitted in the examples. Also, it will be sometimes convenient to consider a rule R simply as a pair of sets of atoms (B, H). Furthermore, we will use $body(R)$ to denote B and $head(R)$ to denote H.

Note that an existential rule is generally not equivalent to a clause, even if its head contains a single atom, because of its existential variables. However, any existential rule can be transformed into a set of Horn clauses by a Skolemization operation, which replaces each existentially quantified variable with a Skolem function whose arguments are the frontier variables of the rule. For instance, the rule $R = p(x, w, y) \to \exists z_1.\, \exists z_2.\, q(x, z_1) \wedge t(z_1, z_2, y)$ is rewritten as two rules $p(x, w, y) \to q(x, f_R^{z_1}(x, y))$ and $p(x, w, y) \to t(f_R^{z_1}(x, y), f_R^{z_2}(x, y), y)$. If moreover, all variables in the factbase are replaced by fresh constants (0-ary Skolem terms), the transformation yields a positive logic program which preserves the logical consequence with respect to all formulas where the Skolem functions do not occur. In Section 2.2, we will present a variant of the chase called "Skolem chase" which actually performs the forward chaining on the logical program associated with a set of existential rules.

Let F be a factbase and $R = (B, H)$ an existential rule. If there is a homomorphism π from B to F, then we say that R is *applicable* on F via π. Then, the pair $t = (R, \pi)$ is called a *trigger*, and we will also say that t is applicable on F. We denote by π^s the extension of π which maps all existential variables in H to fresh variables indexed by the trigger t. More precisely, for each existential variable z in H, we define $\pi^s(z) = z_t$. The application of t on F results in the factbase $F \cup \pi^s(H)$, which is called an *immediate derivation* from F through t. This fixed way to name fresh variables ensures that the chase always produces the same atoms when the same trigger is applied on different

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

derivations. It will be useful when comparing forward chaining with the same ruleset but different factbases or chase variants. To simplify notation in cases where R, π, B and H are not specified, $\pi(B)$ is called the *support* of t and is denoted by $support(t)$ and $\pi^s(H)$ is called the *output* of t and is denoted by $output(t)$. Finally, given a trigger (R, π), we denote by $\pi_{|fr(R)}$ the restriction of π to the frontier variables of R.

2.2 Chase variants

The chase is built upon the notion of derivation, which consists of the repeating application of rules from a certain ruleset to a factbase which is evolving with every rule application.

Definition 4 (*Derivation*)
A *derivation* from a KB $(F, \mathcal{R})$ is a (possibly infinite) sequence of pairs $\mathcal{D} = (\emptyset, F_0), (t_1, F_1), (t_2, F_2), \ldots$, where $F_0 = F$ and F_i, for each $i > 0$, is an immediate derivation from F_{i-1} through a new trigger t_i (i.e., $t_i \neq t_j$ for all $i \neq j$).

The sequence of triggers in a derivation is denoted by $triggers(\mathcal{D})$, while the set of atoms inferred by $\mathcal{D}$ is denoted by $F^{\mathcal{D}} = \bigcup_i F_i$. A derivation pair (t_i, F_i) is called an element of a derivation. The k-prefix of a derivation $\mathcal{D}$ is the prefix where the first k triggers have been applied and is denoted by $\mathcal{D}_{|k}$.[2] In this case, we also say that $\mathcal{D}$ is an extension of $\mathcal{D}_{|k}$. Then, an atom A is *produced* by $t_i \in triggers(\mathcal{D})$ if i is the smallest integer such that $A \in F_i \setminus F_{i-1}$.

It is worth noting that if t produces A, then, of course, $A \in output(t)$, but the converse may not hold. Indeed, there may be an atom in the output of a trigger that belongs to the initial factbase or that has been produced earlier. To illustrate, consider $F = \{p(a), q(a), r(a)\}$ and a ruleset $\mathcal{R} = \{R_1, R_2, R_3\}$, where $R_1 : p(x) \rightarrow s(x)$, $R_2 : q(x) \rightarrow s(x)$, and $R_3 : p(x) \rightarrow r(x)$. These rules are applicable on F with the respective triggers $t_i = (R_i, \{x \mapsto a\})$. Consider two derivations $\mathcal{D}_1, \mathcal{D}_2$ from $(F, \mathcal{R})$ where t_i is applied first in $\mathcal{D}_i$, and then the remaining triggers. Although the atom $s(a)$ is in the output of both t_1 and t_2, only t_1 produces $s(a)$ in $\mathcal{D}_1$ and only t_2 produces $s(a)$ in $\mathcal{D}_2$. Furthermore, t_3 does not produce $r(a)$ in any derivation, as this atom belongs to F.

Definition 5 (*Rank and Depth*)
The *rank* of an atom A within a derivation $\mathcal{D}$ is defined as $rank_{\mathcal{D}}(A) = 0$ if $A \in F_0$ and otherwise, let t be the trigger that produces A in $\mathcal{D}$, then $rank_{\mathcal{D}}(A) = 1 + \max\{rank_{\mathcal{D}}(A') \mid A' \in support(t)\}$. This notion is naturally extended to triggers: $rank_{\mathcal{D}}(t) = 1 + \max\{rank_{\mathcal{D}}(A') \mid A' \in support(t)\}$. Then, the *depth* of a derivation is the maximal rank of its atoms if it is finite and infinite otherwise.

Informally speaking, the atom rank does not indicate the number of triggers needed to produce it but rather the number of parallel rule application steps that are needed to produce it. The notion of rank stems from breadth-first derivations but applies to any derivation. Importantly, two derivations may produce the same atom at different ranks, and this already occurs for Datalog KBs, as illustrated by the next example.

[2] Therefore, $\mathcal{D}_{|k}$ comprises the first $k + 1$ elements of $\mathcal{D}$, since derivations start with $(\emptyset, F_0)$.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Example 6
Let $F = \{p(a)\}$ and $\mathcal{R} = \{R_1, R_2, R_3\}$, where $R_1 = p(x) \to q(x)$, $R_2 = q(x) \to r(x)$, $R_3 = p(x) \to r(x)$, and the following derivations:

$$\mathcal{D}_1 = (\emptyset, F), ((R_1, \pi), F_1), ((R_2, \pi), F_2), ((R_3, \pi), F_2),$$

$$\mathcal{D}_2 = (\emptyset, F), ((R_1, \pi), F_1), ((R_3, \pi), F_2), ((R_2, \pi), F_2),$$

where $\pi = \{x \mapsto a\}$. In both derivations, $F_1 = \{p(a), q(a)\}$ and $F_2 = \{p(a), q(a), r(a)\}$. The atom $r(a)$ has rank 2 in $\mathcal{D}_1$, while it has rank 1 in $\mathcal{D}_2$. Note that both derivations are maximal (we will later call them terminating); however, the depth of $\mathcal{D}_1$ is 2, whereas the depth of $\mathcal{D}_2$ is 1.

We define a *chase variant* as a class of derivations representing the possible runs of the chase. The derivations that are proper to each variant are specified by imposing restrictions on which triggers can be applied and when they can be applied. We start by presenting the oblivious (**O**), semi-oblivious (**SO**), restricted (**R**), equivalent chase (**E**), and then move to their breadth-first versions. In short, the oblivious chase applies all possible triggers once, while the semi-oblivious chase does not apply triggers that map a rule frontier in the same way as a previously applied trigger. The restricted chase does not apply a trigger if there is a retraction from the resulting factbase to the current factbase, and the equivalent chase does not apply a trigger if the resulting factbase is (logically) equivalent to the current factbase. The next example illustrates the behavior of these four chase variants.

Example 7
Consider the KBs $\mathcal{K}_1 = (F, \{R_1\})$, $\mathcal{K}_2 = (F, \{R_2\})$, and $\mathcal{K}_3 = (F', \{R_3\})$ built from the factbases $F = \{p(a, a)\}$ and $F' = \{p(a, w)\}$, where w is a variable, and the rules $R_1 = p(x, y) \to \exists z.\ p(x, z)$, $R_2 = p(x, y) \to \exists z.\ p(y, z)$, and $R_3 = p(x, y) \to \exists z. p(x, x) \wedge p(y, z)$.

Regarding $\mathcal{K}_1$, the rule R_1 yields infinitely many triggers producing the atoms $p(a, z_0)$, $p(a, z_1), \ldots$ Hence, the oblivious chase does not halt on $\mathcal{K}_1$. Observe that all these triggers map the frontier variable x to the same constant a. Hence, the semi-oblivious chase applies only the first trigger and halts. However, it does not halt on $\mathcal{K}_2$, while the restricted chase does. Here again, R_2 yields infinitely many triggers, producing the atoms $p(a, z_0), p(z_0, z_1), \ldots$; since each of them maps the frontier variables to new existentials, all these triggers are applied by the semi-oblivious chase. However, all generated atoms are redundant with respect to the initial atom $p(a, a)$. More precisely, there is a retraction from the factbase obtained after the first application of R_2 to the initial factbase; hence, the restricted chase halts without producing any atom. On the other hand, the restricted chase does not halt on $\mathcal{K}_3$, while the equivalent chase does. In this case, the first rule application yields the factbase $F'' = \{p(a, w), p(a, a), p(w, z_0)\}$, where w and z_0 are existentially quantified variables. Note that there is no retraction from F'' to F', and hence the restricted chase applies this trigger. The process continues, because from the atom $p(w, z_0)$ there is another trigger application which admits no retraction to F''. This creates a chain of atoms $p(w, z_0), p(z_0, z_1), \ldots$ making the restricted chase not terminating on $\mathcal{K}_3$. The equivalent chase terminates since, despite the fact that there is no retraction from F'' to F', the factbase F'' is actually redundant and logically equivalent to $\{p(a, a)\}$, which is also its core. Since any new trigger would produce equivalent

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

factbases, the equivalent chase does not perform any other rule application and outputs F'' (the core chase, described later, would instead output $\{p(a,a)\}$). Finally, note that $p(a,a)$ is a (finite) universal model for all KBs $\mathcal{K}_1, \mathcal{K}_2$, and $\mathcal{K}_3$.

Definition 8 (*X-applicability*)
Let $\mathcal{D}$ be a finite derivation from $(F, \mathcal{R})$ and t be a trigger which is applicable on $F^{\mathcal{D}}$. Then, t is

1. **O**-*applicable* on $\mathcal{D}$ if it does not belong to $triggers(\mathcal{D})$;
2. **SO**-*applicable* on $\mathcal{D}$ if $t = (R, \pi)$ and there exists no trigger $(R, \pi') \in triggers(\mathcal{D})$ such that $\pi_{|fr(R)} = \pi'_{|fr(R)}$;
3. **R**-*applicable* on $\mathcal{D}$ if there exists no retraction from $F^{\mathcal{D}} \cup output(t)$ to $F^{\mathcal{D}}$;
4. **E**-*applicable* on $\mathcal{D}$ if there exists no homomorphism from $F^{\mathcal{D}} \cup output(t)$ to $F^{\mathcal{D}}$.

Let $X \in \{\mathbf{O}, \mathbf{SO}, \mathbf{R}, \mathbf{E}\}$. A derivation $\mathcal{D}$ from a KB $(F, \mathcal{R})$ where every trigger $t_i \in triggers(\mathcal{D})$ is X-applicable on the prefix $\mathcal{D}_{|i-1}$ of $\mathcal{D}$ is called an *X-derivation.* The class of all the X-derivations is the *X-chase.*

Note that only the definitions of **O**- and **SO**-applicability allow one to extend a derivation with a trigger that does not produce any (new) atom, which are instead ruled out by **R**- and **E**-applicability. The corresponding classes of derivations, that is, chase variants, will be called **O**-chase, **SO**-chase, **R**-chase, and **E**-chase, respectively.

A natural strategy for the chase is to proceed in a breadth-first manner, that is, by applying all possible triggers rank by rank. Breadth-first derivation are defined as follows.

Definition 9 (*Breadth-First Derivations*)
An X-derivation $\mathcal{D}$ is said to be

- *rank-compatible* if $rank(t_i) \leq rank(t_j)$, for all $t_i, t_j \in triggers(\mathcal{D})$ such that $i < j$. Whenever $rank(t_i) < rank(t_{i+1})$ the index i is called a *rank mark*, it means that the first i triggers of $\mathcal{D}$ contain all and only the triggers of rank up to $rank(t_i)$ of $\mathcal{D}$.
- *breadth-first* if it is rank-compatible and, for every rank mark i, the prefix $\mathcal{D}_{|i}$ cannot be extended to an X-derivation of the same depth.

Let $X \in \{\mathbf{O}, \mathbf{SO}, \mathbf{R}, \mathbf{E}\}$ be a chase variant. The **bf-**X-chase variant is the subclass of the X-chase comprised exclusively by breadth-first X-derivations. As we will outline in the next section, breadth-first derivations not only represent a natural way of reasoning on a KB, but for some chase variants, they also behave better than the other derivations with regard to termination.

Example 10
Consider the KB from Example 6: $\mathcal{D}_2$ is breadth-first, while $\mathcal{D}_1$ is not rank-compatible, as the rank of the third trigger is strictly smaller than the rank of the second trigger.

The notion of termination relies on *fairness*, that is, on the fact that, according to the chase variant, all applicable triggers have either been applied at some point or became redundant. Of course fair derivations may be infinite. However, a (finite) fair derivation produces a (finite) universal model of the KB.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Definition 11 (*Fairness and Termination*)
An X-derivation $\mathcal{D}$ is *fair* if whenever a trigger t is X-applicable on $\mathcal{D}_{|i}$, there exists a $k > i$ such that

- either $t_k = t$,
- or t is not X-applicable on $\mathcal{D}_{|k}$.

An X-derivation is *terminating* if it is both fair and finite.

Note that an X-derivation that is not terminating might be finite, but in this case it cannot be fair. To conclude the section, let us now link the previous chase variants to some other known chase variants. The semi-oblivious and Skolem chases, both defined in Marnette (2009), lead to similar derivations. The Skolem chase consists in running the oblivious chase on the Skolemized KB. This is obtained by transforming the set of existential rules into a logical program, as described in Section 2.1, and by replacing every variable of the factbase with a fresh constant. As already mentioned, this in turn corresponds to the classical forward chaining procedure for positive logic programs. The semi-oblivious and Skolem chase yield isomorphic results, in the sense that they generate exactly the same sets of atoms, up to a bijective renaming of nulls by Skolem terms. Therefore, we chose to focus on just one of them.

The core chase (Deutsch *et al.* 2008) and the breadth-first equivalent chase (Rocher 2016) are two variants with a similar behavior, which terminate on a KB if and only if this factbase has a finite universal model. The core chase proceeds in a breadth-first manner and, at each step, performs in parallel all rule applications according to the restricted chase criterion, and then computes a core of the resulting factbase. Hence, the core chase may remove at some step atoms that were introduced at an earlier step. After i breadth-first steps, the equivalent chase and the core chase yield logically equivalent factbases, and they terminate on the same inputs. This follows from the fact that computing the core after each rule application or after a sequence of rule applications gives isomorphic results, and that two finite factbases are logically equivalent if and only if their respective cores are isomorphic. However, it will be convenient to handle the equivalent chase from a formal point of view because it does not remove any atoms produced by a derivation.

3 *k*-Boundedness

In this section, after defining the notion of k-boundedness, we will present a set of properties of breadth-first derivations relevant to the chase variants that will be studied. This will allow us, on the one hand, to present some fundamental results that are key to the decidability of k-boundedness and, on the other hand, to provide a better understanding of boundedness itself.

3.1 k-Boundedness

As already mentioned, the concept of boundedness was first introduced for Datalog programs. A Datalog program is said to be *bounded* if the number of breadth-first steps of a bottom-up evaluation of the program is bounded independently from any factbase (this notion being more precisely called *uniform boundedness* to distinguish it from the

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

notion of *program boundedness* that restricts the set of predicates that may occur in the factbase) (Gaifman *et al.* 1993; Abiteboul 1989; Guessarian and Peixoto 1994). We apply this concept to the more general language of existential rules to define k-boundedness and parametrize it by the considered chase variants. Since every chase behaves differently with respect to termination, every chase gives rise to a distinct notion of k-boundedness. Indeed, a ruleset may be k-bounded for one chase variant but not k-bounded for another variant which employs a weaker applicability condition. This is illustrated by the following example.

Example 12
Consider $\mathcal{R} = \{R_1, R_2\}$, where $R_1 = p(x, y) \rightarrow \exists z.\ q(z, x)$ and $R_2 = q(z, x) \rightarrow \exists w.\ p(x, w)$. Then $\mathcal{R}$ is not **O**-bounded but it is **SO**-bounded; more precisely, $\mathcal{R}$ is 2-bounded for the **SO**-chase, which means that all (fair) **SO**-derivations built with $\mathcal{R}$ have depth at most 2. To see that, first note that, since the rule bodies have a single atom, the saturation of any factbase F is included in the union of the saturations of each atom in F. Hence, the depth of a derivation on F is bounded by the maximal depth of a derivation from an atom in F. Then notice that every derivation from an atom can only alternate the application of rule R_1 with R_2. Now, to show that $\mathcal{R}$ is not **O**-bounded, take an initial fact $p(a, b)$ and observe that we can build an infinite **O**-derivation. On the other hand, the **SO**-chase halts after producing the atom $q(z_{t_1}, a)$ at rank 1 (by applying R_1) and then $p(a, w_{t_2})$ at rank 2 (by applying R_2). At this point, R_1 is not **SO**-applicable anymore because it has been already applied by mapping its frontier x to the constant a. Indeed, **SO**-derivations have depth at most two for all possible initial factbases. Finally, $\mathcal{R}$ is 1-bounded for the **R**-chase, that is, all (fair) **R**-derivations built with $\mathcal{R}$ have depth at most 1. Starting from the fact $p(a, b)$ for instance, the **R**-chase does not produce $p(a, w_{t_2})$ at rank 2 because of the retraction $\pi = \{w_{t_2} \mapsto b\}$. It follows that $\mathcal{R}$ is also 1-bounded for the **E**-chase.

Another important dimension to the boundedness problem, studied also for chase termination (Grahne and Onet 2018), is whether the bound on the chase depth is considered for all (fair) derivations or one (fair) derivation, as illustrated by the following example.

Example 13
Consider the ruleset $\mathcal{R}_2 = \{R, R'\}$ of Example 3. Since $\mathcal{R}_2$ is Datalog, for every factbase F, all derivations from $(F, \mathcal{R}_2)$ are terminating. Also, for every integer k, there exists a factbase F and a terminating derivation from $(F, \mathcal{R}_2)$ of depth k, where R is applied in all possible ways before applying R'. Note, however, that rule R' computes in one rank everything that R can compute in many ranks. This means that, for every factbase F, there is also a terminating derivation from $(F, \mathcal{R}_2)$ of depth 1 and the chase can halt earlier if a wise prioritization on rules is chosen.

This leads us to the following definition.

Definition 14 (k-*Boundedness*)
Let X be any chase variant. A ruleset $\mathcal{R}$ is
$\forall$-X-k-*bounded* if, for each factbase F, all fair X-derivations from $(F, \mathcal{R})$ are of depth at most k,
$\exists$-X-k-*bounded* if, for each factbase F, there is a fair X-derivation from $(F, \mathcal{R})$ of depth at most k.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

The $\forall$-X-k-*boundedness* (resp. $\exists$-X-k-*boundedness*) problem takes as input a ruleset $\mathcal{R}$ and an integer k and asks whether $\mathcal{R}$ is $\forall$-X-k-bounded (resp. $\exists$-X-k-bounded).

Note that saying that all fair X-derivations from $(F, \mathcal{R})$ are of depth at most k is equivalent to saying that all X-derivations from $(F, \mathcal{R})$ are of depth at most k. One direction of this property is trivial: if all X-derivations are at most k-deep, then all fair X-derivations are at most k-deep. For the other direction, we show the contrapositive. Assume there is an unfair X-derivation of depth strictly greater than k. Then take its shortest prefix of depth $k+1$ (i.e., which ends with the first trigger of rank $k+1$). We know that it can be extended by applying at least a trigger (but possibly an infinite number of them) so as to get a fair derivation of depth greater than k. Also note that, by definition, a k-deep fair X-derivation is also terminating.

It should be clarified that in this work, we will focus our attention on $\forall$-X-k-boundedness, later referred as X-k-boundedness or simply k-boundedness when we do not need to specify any particular chase variant. However, we will establish a connection between the two versions of the problem that will allow us to transfer our results to $\exists$-X-k-boundedness for the (breadth-first)(semi-)oblivious chase variants (Theorem 22). This leverages on some fundamental properties of breadth-first derivations, that are now presented.

3.1.1 Rank minimality in breadth-first derivations

The notion of the rank of an atom is central for studying k-boundedness, as the problem amounts to deciding if all atoms produced by the derivations of interest have rank bounded by k, independently of the initial instance.

As illustrated by Examples 3, 6, and 13 on Datalog rulesets, if a chase variant does not impose any constraint on the order in which triggers are applied, then the rank of an atom can vary from one derivation to another. Of course, this happens also for rulesets that are not Datalog. Take for example the KB $(F, \{R_1, R_2, R_3\})$, where $F = \{p(a, b, c)\}$, $R_1 = p(x, y, w) \rightarrow \exists z.\ p(y, z, w)$, $R_2 = p(x, y, w) \rightarrow p(y, y, w) \wedge q(w)$, and $R_3 = q(w) \rightarrow t(w)$. For any integer $k \geq 2$, there is an X-derivation generating the atom $t(c)$ at rank k, when $X \in \{\mathbf{O}, \mathbf{SO}, \mathbf{R}, \mathbf{E}\}$. Note that for the **R**- and **E**-chase this derivation of arbitrary depth can also be terminating.

It is therefore natural to ask what is the minimal rank that a given atom can assume in any X-derivation from a given KB and also whether there is a chase variant that allows one to produce all atoms at their minimal ranks, if a prioritization on triggers is assumed. It turns out that breadth-first oblivious derivations set the lower bound for the ranks of atoms. Recall also that this variant uses the weaker form of applicability condition which makes the **bf-O**-chase inferring any atom that can be produced by any other chase derivation of the same depth.

Proposition 15
Let $\mathcal{D}$ be any **bf-O**-derivation from $(F, \mathcal{R})$ and $\mathcal{D}'$ any derivation from $(F, \mathcal{R})$ of lower or equal depth. Then, $F^{\mathcal{D}'} \subseteq F^{\mathcal{D}}$ and $rank_{\mathcal{D}}(A) \leq rank_{\mathcal{D}'}(A)$, for all $A \in F^{\mathcal{D}'}$.

Proof
By induction on the depth m of $\mathcal{D}$. If $m = 0$, then the claim follows as $F^{\mathcal{D}} = F = F^{\mathcal{D}'}$. Assume that $\mathcal{D}$ is of depth m. We denote by $\mathcal{D}''$ the derivation obtained by applying

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

the maximal subsequence of $triggers(\mathcal{D}')$ of depth up to $m-1$, in the given order. Then, let us denote by $\mathcal{D}_{|depth(m-1)}$ the maximal prefix of $\mathcal{D}$ of depth $m-1$. By the inductive hypothesis, $F^{\mathcal{D}''} \subseteq F^{\mathcal{D}_{|depth(m-1)}}$ and $rank_{\mathcal{D}_{|depth(m-1)}}(A) \leq rank_{\mathcal{D}''}(A)$, for all $A \in F^{\mathcal{D}''}$. Each trigger $t \in triggers(\mathcal{D}') \setminus triggers(\mathcal{D}'')$ is **O**-applicable on $F^{\mathcal{D}_{|depth(m-1)}}$, because any atom in $support(t)$ is produced within rank $m-1$, and so $output(t) \subseteq F^{\mathcal{D}}$. Therefore, $F^{\mathcal{D}'} \subseteq F^{\mathcal{D}}$. Finally, for all $A \in F^{\mathcal{D}'}$, either $A \in F^{\mathcal{D}''}$ and then $rank_{\mathcal{D}}(A) = rank_{\mathcal{D}_{|depth(m-1)}}(A) \leq rank_{\mathcal{D}''}(A) = rank_{\mathcal{D}'}(A)$, or $A \in F^{\mathcal{D}'} \setminus F^{\mathcal{D}''}$ and then $rank_{\mathcal{D}'}(A) = m \leq rank_{\mathcal{D}}(A)$. □

It follows in particular that two **bf-O**-derivations of the same depth produce exactly the same set of atoms. More precisely, these derivations use exactly the same set of triggers but possibly taken in different order.[3]

As breadth-first derivations apply all possible triggers at each rank (according to their criteria), it turns out that two breadth-first derivations of the same depth with the same KB yield equivalent sets of atoms. Nevertheless, one derivation can produce more redundant atoms than the other. The following properties show that breadth-first oblivious derivations can be mapped by retractions to breadth-first semi-oblivious and restricted derivations of the same depth. Furthermore, these retractions map atoms to atoms of smaller or equal rank.

Proposition 16 (*Retraction on Breadth-First Derivations*)
Let $\mathcal{D}$ be any derivation from $(F, \mathcal{R})$ and $\mathcal{D}'$ a **bf-**X-derivation from $(F, \mathcal{R})$ of equal or greater depth, with $X \in \{\mathbf{SO}, \mathbf{R}\}$. Then, there exists a retraction h from $F^{\mathcal{D}} \cup F^{\mathcal{D}'}$ to $F^{\mathcal{D}'}$ such that $rank_{\mathcal{D}}(A) \geq rank_{\mathcal{D}'}\big(h(A)\big)$, for every $A \in F^{\mathcal{D}}$.

Proof
We first show the claim for **bf-O**-derivations $\mathcal{D}$. In the last paragraph of the proof, we will use Proposition 15 to extend the claim to arbitrary derivations $\mathcal{D}$.

We use induction on the depth of $\mathcal{D}$. If $\mathcal{D}$ is of depth 0, then the claim follows as h is the identity. Assume that $\mathcal{D}$ is of depth m. Let $\mathcal{D}_{|depth(m-1)}$ and $\mathcal{D}'_{|depth(m-1)}$ be the maximal prefixes of $\mathcal{D}$ and $\mathcal{D}'$, respectively, that are of depth $m-1$. By the inductive hypothesis, there is a retraction h' from $F^{\mathcal{D}_{|depth(m-1)}} \cup F^{\mathcal{D}'_{|depth(m-1)}}$ to $F^{\mathcal{D}'_{|depth(m-1)}}$. Let $t = (R, \pi) \in triggers(\mathcal{D}) \setminus triggers(\mathcal{D}')$ be any trigger of rank m in $\mathcal{D}$. Then the trigger $t' = (R, h' \circ \pi)$ is **O**-applicable on $F^{\mathcal{D}'_{|depth(m-1)}}$.

If t' is in $\mathcal{D}'$, then its output is in $\mathcal{D}'$, and we can extend h' to $output(t)$ with a bijective renaming of the existential variables in $output(t)$, that is, $h'(x_t) = x_{t'}$, for every existential variable x that occurs in R. In this case, the atoms of $output(t)$ are mapped to atoms of same rank in $\mathcal{D}'$.

If t' is not in $\mathcal{D}'$, then it is not **R**-applicable on $F^{\mathcal{D}'_{|depth(m-1)}}$ (note that a trigger which is not **SO**-applicable is not **R**-applicable either). Hence, there is a retraction h'' from $output(t') \cup F^{\mathcal{D}'_{|depth(m-1)}}$ to $F^{\mathcal{D}'_{|depth(m-1)}}$. So we extend h' by mapping the existential variables in $output(t)$ to $h''(output(t'))$, that is, $h'(x_t) = h''(x_{t'})$, for every existential variable x that appears in R.

[3] This does not hold for **bf-SO**-derivations which can produce atoms where fresh variables can be named differently. However, all semi-oblivious derivations of the same depth produce isomorphic results. If a careful naming of the fresh variables upon the frontier variables of the rule applied by the trigger is chosen, then one can make all **bf-SO**-derivations produce the same sets of atoms.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Notice that in both cases, the atoms of rank m in $\mathcal{D}$ are mapped to atoms of equal or lower rank in $\mathcal{D}'$. Since the sets of existential variables produced by all triggers of rank m in $\mathcal{D}$ are disjoint, the union of all extensions of h' for all these triggers is a retraction of $F^{\mathcal{D}_{|depth(m)}} \cup F^{\mathcal{D}'_{|depth(m)}}$ to $F^{\mathcal{D}'_{|depth(m)}}$.

If now $\mathcal{D}$ is any derivation, then we go back to the previous case using Proposition 15, which asserts the existence of a **bf-O**-derivation $\mathcal{D}^*$ of the same depth as $\mathcal{D}$ such that $F^{\mathcal{D}} \subseteq F^{\mathcal{D}^*}$ and $rank_{\mathcal{D}}(A) \geq rank_{\mathcal{D}^*}(A)$, for all $A \in F^{\mathcal{D}}$. Since $F^{\mathcal{D}} \subseteq F^{\mathcal{D}^*}$, the retraction h from $F^{\mathcal{D}^*} \cup F^{\mathcal{D}'}$ to $F^{\mathcal{D}'}$ is also a retraction from $F^{\mathcal{D}} \cup F^{\mathcal{D}'}$ to $F^{\mathcal{D}'}$ and we also have $rank_{\mathcal{D}}(A) \geq rank_{\mathcal{D}'}\big(h(A)\big)$, for every $A \in F^{\mathcal{D}}$. □

Finally, we show that all breadth-first derivations for the oblivious, semi-oblivious, and restricted chase agree on the rank of the atoms that they produce, independently from the variant.

Proposition 17
Let $\mathcal{D}$ be a **bf-**X-derivation from $(F, \mathcal{R})$ and $\mathcal{D}'$ be a **bf-**Y-derivation from $(F, \mathcal{R})$ with $X, Y \in \{\mathbf{O}, \mathbf{SO}, \mathbf{R}\}$. Then $rank_{\mathcal{D}}(A) = rank_{\mathcal{D}'}(A)$, for all $A \in F^{\mathcal{D}} \cap F^{\mathcal{D}'}$.

Proof
By Proposition 15, the property holds for **bf-O**-derivations, because such derivations produce the same atoms at exactly the same rank. To prove the statement it suffices to show that for any **bf-**X-derivation $\mathcal{D}$, with $X \in \{\mathbf{SO}, \mathbf{R}\}$, there is a **bf-O** derivation $\mathcal{D}'$ from $(F, \mathcal{R})$ and $rank_{\mathcal{D}}(A) = rank_{\mathcal{D}'}(A)$ for all $A \in F^{\mathcal{D}} \cap F^{\mathcal{D}'}$. Assume that $\mathcal{D}$ is of depth m. Let $\mathcal{D}'$ be any **bf-O**-derivation of depth at least m. By Proposition 15, $F^{\mathcal{D}'} \subseteq F^{\mathcal{D}}$. We now proceed by contradiction. Let t be the first trigger in $triggers(\mathcal{D})$ that produces an atom A (common to $F^{\mathcal{D}}$ and $F^{\mathcal{D}'}$) such that $rank_{\mathcal{D}}(A) \neq rank_{\mathcal{D}'}(A)$. Suppose that $rank_{\mathcal{D}}(t) = m$. Thus, for all common atoms of rank $i < m$, their ranks coincide between the two derivations. By Proposition 15, $rank_{\mathcal{D}}(A) > rank_{\mathcal{D}'}(A)$ as **bf-O** derivation set the lower bound for the rank of atoms. Moreover, every atom in $support(t)$ is produced within rank $m-1$ in $\mathcal{D}$, so every atom in $support(t)$ is produced within rank $m-1$ also in $\mathcal{D}'$. Let us denote by $\mathcal{D}_{|depth(m-1)}$ and $\mathcal{D}'_{|depth(m-1)}$ the maximal prefix of $\mathcal{D}$ and $\mathcal{D}'$ of depth $m-1$. We conclude that t is X-applicable on $\mathcal{D}_{|depth(m-1)}$ and **O**-applicable on $\mathcal{D}'_{|depth(m-1)}$.

As t does not produce A at rank m in $\mathcal{D}'$, there is a trigger $t' \in triggers(\mathcal{D}')$ producing A at a rank smaller than m, therefore $A \in F^{\mathcal{D}'_{|depth(m-1)}}$. Since A belongs to the output of two different triggers t and t' in $\mathcal{D}'$, and fresh variables are named after the trigger that generates them, we conclude that A is obtained from an atom in the rule head of t (and t') that only uses frontier variables. Therefore, all the terms of A are present in $support(t)$, so also in $F^{\mathcal{D}_{|depth(m-1)}}$.

By Proposition 16, there is a retraction h from $F^{\mathcal{D}'_{|depth(m-1)}} \cup F^{\mathcal{D}_{|depth(m-1)}}$ to $F^{\mathcal{D}_{|depth(m-1)}}$. Because all the terms of A appear in $F^{\mathcal{D}_{|depth(m-1)}}$, we obtain $h(A) = A$. Hence, $A \in F^{\mathcal{D}_{|depth(m-1)}}$, so A is not produced by t in $\mathcal{D}$. A contradiction. □

Proposition 17 will be essential to prove decidability of k-boundedness for the breadth-first semi-oblivious and restricted chase. Surprisingly, it does not hold for the breadth-first equivalent chase variant, that is, with $X = Y =$ **bf-E**, as atoms can have different ranks depending on the derivation, as shown by the following example.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Example 18
For two predicates p_1, p_2, we denote by $R_{p_1p_2}$ the unary inclusion rule $p_1(x) \rightarrow p_2(x)$. Let $\mathcal{R} = \{R_{pq}, R_{qp}, R_{qr}, R_{pr}, R_{rs}, R_{rp}\}$ and $F = \{p(z_1), q(z_2)\}$, where z_1 and z_2 are variables. We construct two **bf-E**-derivations $\mathcal{D}_1, \mathcal{D}_2$ which produce the atom $p(z_2)$ at different ranks. Let $h_1 = \{x \mapsto z_1\}$ and $h_2 = \{x \mapsto z_2\}$. We describe the derivations in terms of their triggers.

$$triggers(\mathcal{D}_1) = (R_{pq}, h_1), (R_{qr}, h_2), (R_{pr}, h_1), (R_{rs}, h_2), (R_{rp}, h_2),$$
$$triggers(\mathcal{D}_2) = (R_{qp}, h_2), (R_{qr}, h_2), (R_{rs}, h_2).$$

The first rank of $\mathcal{D}_1$ sees the application of three rules producing the atoms $q(z_1)$, $r(z_2)$, and $r(z_1)$ in the given order. At this point, the trigger (R_{qp}, h_2) producing $p(z_2)$ is **R**-applicable but not **E**-applicable, because the homomorphism $h = \{z_2 \mapsto z_1\}$ from $F^{\mathcal{D}_{1|3}} \cup \{p(z_2)\}$ to $F^{\mathcal{D}_{1|3}}$ (recall the notation $\mathcal{D}_{|i}$ for the application of the first i triggers of $\mathcal{D}$) makes this inference redundant. Rank 1 is thus complete. At rank 2, two rules starting from the atom $r(z_2)$ are applied. The atom $s(z_2)$ is produced first, resulting to h *not* being homomorphism from $F^{\mathcal{D}_{1|4}} \cup \{p(z_2)\}$ to $F^{\mathcal{D}_{1|4}}$. So then, we can apply R_{rp} and produce $p(z_2)$. The derivation $\mathcal{D}_1$ is terminating and the final factbase is $\{p(z_1), q(z_2), q(z_1), r(z_2), r(z_1), s(z_2), p(z_2)\}$. But in $\mathcal{D}_2$, the triggers are applied in a different order, starting from one that produces $p(z_2)$ at rank 1. This derivation also terminates at rank 2 with factbase $\{p(z_1), q(z_2), p(z_2), r(z_2), s(z_2)\}$. Of course the factbases produced by the two derivations are equivalent.

3.1.2 Breadth-first derivations: termination and depth

We have shown that when $X \in \{\mathbf{O}, \mathbf{SO}, \mathbf{R}\}$, a breadth-first X-derivation produces all atoms at their lower rank. Does this imply that, if the X-chase halts, then the **bf-**X-chase halts as well? This actually holds for the oblivious and semi-oblivious chases. Indeed, among all terminating derivations for the oblivious and semi-oblivious chase, the breadth-first are the ones with the smallest depth, as stated by the next proposition, which directly follows from the above results.

Proposition 19
Let $X \in \{\mathbf{O}, \mathbf{SO}\}$. For each terminating X-chase derivation from $(F, \mathcal{R})$, there is a terminating **bf-**X-derivation from $(F, \mathcal{R})$ of smaller or equal depth. Moreover, all terminating **bf-**X-chase derivations from $(F, \mathcal{R})$ have the same depth.

Proof
The result directly follows from Propositions 15 and 17 for the **bf-O** variant. For the **bf-SO** variant, we furthermore argue that all **bf-SO**-derivations of the same depth produce isomorphic results. □

The case of the restricted chase is more complex, since, for a given factbase, some fair derivations may terminate, while others may not. It may happen that all breadth-first derivations terminate (even with depth less than a predefined number k), but there is a fair non-breadth-first derivation that does not terminate. It may also be the case that no breadth-first derivation terminates, but there is a non-breadth-first derivation that terminates (even with predefined depth less than k), as illustrated by the next example.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Example 20
Let $F = \{p(a, b)\}$ and $\mathcal{R} = \{R_1, R_2, R_3\}$ with $R_1 = p(x, y) \to \exists z.\ p(y, z)$, $R_2 = p(x, y) \to \exists z.\ q(y, z)$ and $R_3 = q(y, z) \to p(y, y)$. It is easy to see that a breadth-first **R**-chase derivation in this KB cannot be terminating. However, by applying only R_2 on F and then R_3 on the new atom, we obtain a terminating **R**-chase derivation. Note also that, for any factbase, there is a terminating **R**-chase derivation of depth at most 2.

Hence, in the case of the restricted chase, breadth-first derivations are not necessarily derivations of minimal depth. In contrast, this holds for rank-compatible **R**-chase derivations (which differ from breadth-first derivations because not all active triggers of a given rank are necessarily applied).

Proposition 21
For each terminating **R**-derivation from $(F, \mathcal{R})$, there exists a terminating rank-compatible **R**-derivation from $(F, \mathcal{R})$ of smaller or equal depth.

Proof
Let $\mathcal{D}$ be a terminating **R**-derivation from $(F, \mathcal{R})$. Let $\mathcal{T}$ be a sorting of $\mathit{triggers}(\mathcal{D})$ such that the rank of each element is greater or equal to the rank of its predecessors. Let $\mathcal{D}'$ be the maximal **R**-derivation from $(F, \mathcal{R})$ that only uses the triggers of $\mathcal{T}$, in the given order. Now, if $\mathit{triggers}(\mathcal{D}') = \mathit{triggers}(\mathcal{D})$, then $\mathcal{D}'$ is rank-compatible; moreover, $\mathcal{D}$ is terminating, hence $\mathcal{D}'$ is also terminating, because both derivations produce the same atomsets. Otherwise, $\mathit{triggers}(\mathcal{D}') \neq \mathit{triggers}(\mathcal{D})$ is due to some of triggers occurring in $\mathcal{D}$ that are no longer **R**-applicable in $\mathcal{D}'$. We will show that $\mathcal{D}'$ is a terminating **R**-derivation in this case as well.

Let $t_1 \dots t_m$ be the triggers of $\mathit{triggers}(\mathcal{D}') \backslash \mathit{triggers}(\mathcal{D})$ that were not **R**-applicable when constructing $\mathcal{D}'$, where $t_i = (R_i, \pi_i)$ and $R_i = H_i \to B_i$. So $F^{\mathcal{D}} = F^{\mathcal{D}'} \cup \pi_1^s(H_1) \cup \dots \cup \pi_m^s(H_m)$. We know that for every $i \in \{1, \dots, m\}$, there is a retraction σ_i from $\pi_i^s(H_i) \cup F^{\mathcal{D}'}$ to $F^{\mathcal{D}'}$. Since with $\sigma_1, \dots, \sigma_m$, only new variables are mapped to different terms (and all other variables are mapped to themselves), we can define the substitution $\tau = \bigcup_{i=1}^{m} \sigma_i$ which has the property that $\tau\big(F^{\mathcal{D}'} \cup \pi_1^s(H_1) \cup \dots \cup \pi_m^s(H_m)\big) = F^{\mathcal{D}'}$.

Suppose that there is a new trigger $t = (R, \pi)$ with $R = B \to H$ which is **R**-applicable on $\mathcal{D}'$, thus there is a homomorphism from B to $F^{\mathcal{D}'}$. Then, since $F^{\mathcal{D}'} \subseteq F^{\mathcal{D}}$, there is a homomorphism from B to $F^{\mathcal{D}}$. But because $\mathcal{D}$ is a terminating **R**-derivation, t is not **R**-applicable on $\mathcal{D}$. We conclude that there is a retraction σ from $\pi^s(H) \cup F^{\mathcal{D}}$ to $F^{\mathcal{D}}$. Since the new variables created from $t_1, \dots, t_m, t$ are all different, the composition $\tau \circ \sigma$ is a retraction from $\pi^s(H) \cup F^{\mathcal{D}'}$ to $F^{\mathcal{D}'}$, which leads us to conclude that no such trigger (R, π) is **R**-applicable on $\mathcal{D}'$. This is a contradiction. □

However, note that different terminating rank-compatible **R**-derivations (and so also **bf-R**-derivations) from the same KB may have different depths (see Example 20). Thus, an important difference between the breath-first (semi-)oblivious and restricted chase variants is that although all terminating breadth-first restricted derivations agree on the ranks of the common atoms they produce (Proposition 17), they are not guaranteed to have the same depth.

Putting everything together, we obtain a characterization of $\exists$-X-k-boundedness in terms of $\forall$-X-k-boundedness for the oblivious and semi-oblivious chases, as well as their breadth-first variants.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Table 1. *Chase variants with respect to heredity/preservation of ancestry*

	O	**bf-O**	**SO**	**bf-SO**	**R**	**bf-R**	**E**	**bf-E**
Heredity	✓	✓	✓	✗	✓	✗	✗	✗
Preservation of ancestry	✓	✓	✓	✓	✓	✓	✗	✗

Proposition 22
Let $X \in \{\mathbf{O}, \mathbf{SO}\}$ and let $\mathcal{R}$ be a ruleset. Then the following statements are equivalent:

1. $\mathcal{R}$ is $\exists$-X-k-bounded;
2. $\mathcal{R}$ is $\exists$-**bf**-X-k-bounded;
3. $\mathcal{R}$ is $\forall$-**bf**-X-k-bounded;

Proof
Follows from Proposition 19. □

In the next section, we focus on $\forall$-k-boundedness and show that it is decidable for the oblivious and semi-oblivious chase as well as their breadth-first versions. By Proposition 22, this will in turn imply decidability of $\exists$-k-boundedness for those chase variants.

4 Decidability of k-boundedness in chase variants

Our approach to the study of k-boundedness is based on identifying a property of chase variants that implies decidability of the problem, namely *preservation of ancestry*. This property is satisfied by the oblivious, semi-oblivious, and restricted chase, as well as their breadth-first variants, which implies decidability of k-boundedness for all these variants. The interest of this approach is that we abstract away from a particular variant thereby providing a proof schema for decidability that can be applied to all chases preserving ancestry. Moreover, we identify a stronger property, namely *heredity*, which is not enjoyed by the breadth-first semi-oblivious and breadth-first restricted chases. While preservation of ancestry is sufficient for decidability of k-boundedness, it remains interesting to consider heredity, as it leads to simpler proofs. Table 1 summarizes the results obtained concerning these properties.

To decide if a ruleset $\mathcal{R}$ is k-bounded, we consider the dual problem of determining if there is a factbase and a derivation of depth $k+1$ constituting a counterexample to the property. Our method relies on the construction of a special factbase that allows us to reproduce such a derivation when the chase enjoys either preservation of ancestry or heredity. We show that the size of such a factbase depends on k and $\mathcal{R}$ only. Therefore, by testing the chase depth over a representative set of bounded-size factbases, we obtain decidability of the problem.

4.1 Preservation of ancestry

Preservation of ancestry is a notion built on the ancestors set of an atom produced by a derivation.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Definition 23 (*Chase Graph and Ancestors*)
Let $\mathcal{D}$ be a derivation from $(F, \mathcal{R})$. The *chase graph* of $\mathcal{D}$ is a (possibly infinite) directed graph $G = (V, E)$, where the nodes are the atoms in $F^{\mathcal{D}}$, and there is an edge from A' to A labeled with t whenever A is produced by $t \in \mathit{triggers}(\mathcal{D})$ and $A' \in \mathit{support}(t)$. Moreover, we say that an atom A' is an *ancestor* of A, if there is a non-empty path from A' to A in the chase graph. The set of ancestors of an atom A is denoted by $\mathit{Anc}_{\mathcal{D}}(A)$, while $\mathit{Anc}^k_{\mathcal{D}}(A)$ is the subset of ancestors whose rank is exactly k. If $A \in F$, then $\mathit{Anc}_{\mathcal{D}}(A) = \emptyset$. Both notations are extended to sets of atoms.

Note that an edge is labeled by the *first* trigger of the derivation that produces the atom. Also, the rank of an atom is equal to the maximum length of a path to this atom in the chase graph, and the depth of the derivation is equal to the maximum length of a path in the chase graph. Note that $\mathit{Anc}^0_{\mathcal{D}}(A)$ are the ancestors of A that belong to the initial factbase, which we call *prime ancestors*. Importantly, the set of prime ancestors of an atom is finite and its size depends on the rank of the atom itself, as well as on the maximum size (i.e., the number of atoms) of the body of a rule, hence it can be bounded.

Proposition 24
Let $\mathcal{D}$ be a derivation from $(F, \mathcal{R})$ and $A \in F^{\mathcal{D}}$ an atom of rank $k > 0$. Then $|\mathit{Anc}^0_{\mathcal{D}}(A)| \leq b^k$, where $b = \mathit{max}\{\ |\mathit{body}(R)| : R \in \mathcal{R}\}$.

Proof
Let (R, π) be the trigger that produces A. Then $\mathit{Anc}^0_{\mathcal{D}}(A)$ is equal to the union of $\{\pi(B_i)\} \cup \mathit{Anc}^0_{\mathcal{D}}(\pi(B_i))$ for all atoms $B_i \in \mathit{body}(R)$. The proof follows by a simple induction on k. □

We denote by $\mathcal{D}_{|G}$ the *restriction of a derivation* $\mathcal{D}$ from $(F, \mathcal{R})$ with respect to $G \subseteq F$, which is the maximal derivation from $(G, \mathcal{R})$ that only uses the triggers of $\mathit{triggers}(\mathcal{D})$, in the given order. As stated below, it can be easily verified that prime ancestors are enough to infer a certain atom as well as all its ancestors.

Proposition 25
Let $\mathcal{D}$ be a derivation from $(F, \mathcal{R})$ and $A \in F^{\mathcal{D}} \setminus F$. Let $\mathcal{D}' = \mathcal{D}_{|\mathit{Anc}^0_{\mathcal{D}}(A)}$. Then $\mathit{Anc}_{\mathcal{D}}(A) \cup \{A\} \subseteq F^{\mathcal{D}'}$.

Proof
By induction on the rank of A in $\mathcal{D}$. Let $t \in \mathit{triggers}(\mathcal{D})$ be the trigger that produces A. If $\mathit{rank}_{\mathcal{D}}(A) = 1$, then $\mathit{Anc}^0_{\mathcal{D}}(A) = \mathit{support}(t)$, hence $t \in \mathit{triggers}(\mathcal{D}')$ and the claim follows. Assume now that $\mathit{rank}_{\mathcal{D}}(A) = n + 1$. Then $\mathit{Anc}_{\mathcal{D}}(A) = \mathit{Anc}_{\mathcal{D}}(\mathit{support}(t)) \cup \mathit{support}(t)$ and $\mathit{Anc}^0_{\mathcal{D}}(A) = \mathit{Anc}^0_{\mathcal{D}}(\mathit{support}(t))$. Since $\mathit{support}(t)$ contains only atoms of rank up to n, by the inductive hypothesis, $\mathit{Anc}_{\mathcal{D}}(\mathit{support}(t)) \cup \mathit{support}(t) \subseteq F^{\mathcal{D}'}$, hence $t \in \mathit{triggers}(\mathcal{D}')$ and $A \in F^{\mathcal{D}'}$. □

Two issues have to be addressed at this point if one wants to exploit the restriction of a derivation to decide k-boundedness. The first is that $\mathcal{D}$ and $\mathcal{D}_{|\mathit{Anc}^0_{\mathcal{D}}(A)}$ could belong to different chase variants. The second is that $\mathcal{D}$ and $\mathcal{D}_{|\mathit{Anc}^0_{\mathcal{D}}(A)}$ could disagree on the rank of atoms. Both properties are not immediate, and this is especially true when $\mathcal{D}$

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

is breadth-first as the restriction of a derivation could, on the one hand, break rank exhaustiveness and, on the other hand, increase the rank of some atoms.

The following examples illustrate that for the breadth-first semi-oblivious and breadth-first restricted chase the restriction of a derivation can lead to a different chase variant. The examples show that there may be triggers that were not applicable starting from F but become applicable by starting from the prime ancestors of an atom.

Example 26 (*Breadth-First Semi-oblivious Chase*)
Let $F = \{p(a,b), r(a,c)\}$ and $\mathcal{R} = \{R_1 = p(x,y) \to r(x,y),\ R_2 = r(x,y) \to \exists z.\ q(x,z),\ R_3 = r(x,y) \to t(y)\}$. Let $\mathcal{D}$ be the (terminating) **bf-SO**-chase derivation of depth 2 from F whose sequence of associated triggers is (R_1,π_1), (R_3,π_2), (R_2,π_2), (R_3,π_1) with $\pi_1 = \{x \mapsto a, y \mapsto b\}$ and $\pi_2 = \{x \mapsto a, y \mapsto c\}$, which produces $r(a,b), t(c), q(a, z_{(R_2,\pi_2)}), t(b)$; the trigger (R_2,π_1) is then **O**-applicable but not **SO**-applicable, as it maps the frontier variable x to a, like the trigger (R_2,π_2). Let $F' = \{p(a,b)\}$. The restriction of $\mathcal{D}$ induced by F' includes only $(R_1,\pi_1), (R_3,\pi_1)$ and is a **SO**-chase derivation of depth 2; however, it is not breadth-first since now (R_2,π_1) is **SO**-applicable at rank 2 (thus, the breadth-first condition is not satisfied).

Example 27 (*Breadth-First Restricted Chase*)
Let $F = \{p(a,b), q(a,c)\}$ and $\mathcal{R} = \{R_1 = p(x,y) \to r(x,y),\ R_2 = r(x,y) \to \exists z.\ q(x,z),\ R_3 = r(x,y) \to t(x)\}$. Let $\mathcal{D}$ be the (terminating) breadth-first derivation of depth 2 from F whose sequence of associated triggers is $(R_1,\pi), (R_3,\pi)$ with $\pi = \{x \mapsto a, y \mapsto b\}$, which produces $r(a,b), t(a)$; note that the trigger (R_2,π) is **SO**-applicable but not **R**-applicable because of the presence of $q(a,c)$ in F. Let $F' = \{p(a,b)\}$. The restriction of $\mathcal{D}$ induced by F' is a restricted chase derivation of depth 2; however, it is not breadth-first since now (R_2,π) is **R**-applicable at rank 2 and thus has to be applied (for the derivation to be breadth-first).

This motivates the definition of preservation of ancestry. An X-chase preserves ancestry if, for any atom A in an X-derivation, there is an X-derivation that starts from the prime ancestors of A and is able to produce A at the same rank.

Definition 28 (*Preservation of Ancestry*)
The X-chase *preserves ancestry* if, for every X-derivation $\mathcal{D}$ from $(F,\mathcal{R})$ and atom $A \in F^{\mathcal{D}} \setminus F$, there is an X-derivation $\mathcal{D}'$ from $(Anc^0_{\mathcal{D}}(A),\mathcal{R})$ such that $A \in F^{\mathcal{D}'}$ and $rank_{\mathcal{D}}(A) = rank_{\mathcal{D}'}(A)$.

We will show that k-boundedness is decidable for any chase variant that satisfies this property. We achieve this by limiting the size of the factbases we need to consider. This implies that a finite number of factbases suffice in order to test k-boundedness of a given ruleset. To formally establish this implication, we introduce the following notion. Let F and F' be atomsets and $\tau : cnst(F) \to cnst(F')$, $\sigma : var(F) \to var(F')$ be mappings of constants and variables, respectively. If $h = \tau \cup \sigma$ is a bijection such that $h(F) = F'$, then h is a *quasi-isomorphism* from F to F'. Two KBs with the same ruleset and quasi-isomorphic factbases behave equivalently with respect to any chase variant.

Theorem 29
Determining if a set of rules is X-k-bounded is decidable if the X-chase preserves ancestry.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Proof
Let X be a chase variant that preserves ancestry. Let $\mathcal{R}$ be a ruleset. Suppose that $\mathcal{R}$ is not X-k-bounded. Therefore, there is a factbase F and a derivation $\mathcal{D}$ from $(F, \mathcal{R})$ with depth strictly greater than k. So there exists an atom $A \in F^{\mathcal{D}}$ with $rank(A) = k + 1$. Because the X-chase preserves ancestry, there exists an X-derivation $\mathcal{D}'$ from $(Anc^0_{\mathcal{D}}(A), \mathcal{R})$ which produces A with the same rank as $\mathcal{D}$. Therefore, $\mathcal{D}'$ is also of depth more than k. Let b be the maximum number of atoms in the bodies of the rules of $\mathcal{R}$. By Proposition 24, $Anc^0_{\mathcal{D}}$ has at most b^{k+1} atoms. We have shown that if a ruleset $\mathcal{R}$ is not X-k-bounded, then there exists a factbase F' of at most b^{k+1} (where b depends on $\mathcal{R}$) such that there is an X-derivation from $(F', \mathcal{R})$ of depth strictly greater than k. The converse of this statement is trivially true. In conclusion, if X is a chase variant that preserves ancestry, then a ruleset $\mathcal{R}$ (with b maximum body size) is X-k-bounded if and only if for every factbase F' of size at most b^{k+1}, every X-derivation from $(F', \mathcal{R})$ is of depth at most k. Up to quasi-isomorphism, there is a finite number of factbases of cardinality less or equal to b^{k+1} and for a given factbase F, there is a finite number of X-derivations from $(F, \mathcal{R})$. Hence, we can indeed compute all these derivations and verify whether $\mathcal{R}$ is X-k-bounded or not. □

We now show that the breadth-first variants of the semi-oblivious and restricted chases preserve ancestry. As illustrated by Examples 26 and 27, the restriction of a derivation to the prime ancestors of an atom A may not satisfy the conditions required by ancestry preservation, although it preserves the rank of A, because it may not be a breadth-first derivation. However, we will see that if we take only the subsequence of triggers that produce ancestors of a given atom A, then we can complete it by missing triggers to obtain a breadth-first derivation which preserves the rank of A. For other chase variants that preserve ancestry as well, we will prove that they satisfy a stronger property, namely heredity.

First, notice that, when reducing the factbase, the application of the same triggers does not necessarily preserve their ranks and hence the ranks of the atoms they produce. Here is a simple example:

Example 30
Let $\mathcal{R}$ contain the rules $R_1 : p(x) \to q(x)$ and $R_2 : p(x) \wedge q(x) \to r(x)$. Let $F = \{p(a), q(a)\}$ and we denote $t_1 = (R_1, \{x \mapsto a\})$ and $t_2 = (R_2, \{x \mapsto a\})$. Here is a derivation $\mathcal{D}$ from $(F, \mathcal{R})$:

$$(\emptyset, F), (t_1, F), (t_2, F \cup \{r(a)\}).$$

Now, for $F' = \{p(a)\}$, there is a derivation $\mathcal{D}'$ from $(F', \mathcal{R})$ such that $triggers(\mathcal{D}') = triggers(\mathcal{D})$:

$$(\emptyset, F'), (t_1, F), (t_2, F \cup \{r(a)\}).$$

Notice that the trigger t_2 and correspondingly the atom $r(a)$ that it produces, have different ranks in the two derivations. In Figure 1, we have the corresponding chase graph(s).

However, it follows from Proposition 15 that by taking the restriction of a **bf-**X-derivation the ranks of atoms can only increase. Indeed, the restriction of a **bf-**X-

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

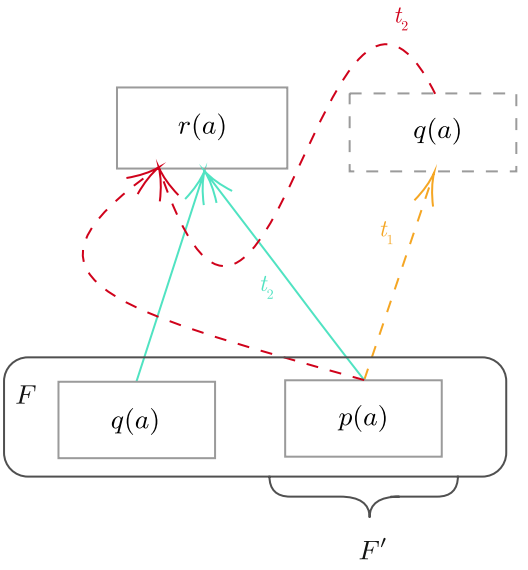


Fig. 1. Chase graph(s) associated with the derivations $\mathcal{D}$ and $\mathcal{D}'$ of Example 30 (dashed elements do not appear in the chase graph of $\mathcal{D}$).

derivation from $(F, \mathcal{R})$ is always an **O**-derivation from $(F, \mathcal{R})$. We conclude that although the ranks can be different, those obtained by the larger factbase provide a lower bound.

The following proposition applies to any breadth-first oblivious, semi-oblivious or restricted derivation: it shows that when we restrict the factbase to the prime ancestors of any atom A produced in a derivation $\mathcal{D}$, the **bf-O**, **bf-SO**, and **bf-R** ranks of all the ancestors of A in the smaller factbase are equal to their ranks in $\mathcal{D}$.

Proposition 31 (*Preservation of Ranks of Ancestors*)
Let $\mathcal{D}$ be a **bf-**X-derivation from $(F, \mathcal{R})$, where $X \in \{\mathbf{O}, \mathbf{SO}, \mathbf{R}\}$. Let $A \in F^{\mathcal{D}}$ be an atom and $\mathcal{D}'$ any **bf-O**-derivation from $(Anc^0_{\mathcal{D}}(A), \mathcal{R})$ of the same depth as $\mathcal{D}$. Then, $rank_{\mathcal{D}}(A') = rank_{\mathcal{D}'}(A')$ for every $A' \in Anc_{\mathcal{D}}(A)$.

Proof
Let $\mathcal{D}_A$ be the **O**-derivation from $(Anc^0_{\mathcal{D}}(A), \mathcal{R})$ given by the maximal subsequence of $triggers(\mathcal{D})$ made only by triggers that produce an ancestor of A. We first show that $rank_{\mathcal{D}}(A') = rank_{\mathcal{D}_A}(A')$, for every $A' \in Anc_{\mathcal{D}}(A)$. We use induction on the number of triggers in $\mathcal{D}_A$. If $\mathcal{D}_A$ has no trigger, then $A \in F$ as the atom has no ancestors, and the claim follows. Now, assume that $\mathcal{D}_A$ has i triggers and let t_i be its last trigger. All the atoms that are in $support(t_i)$ have been produced by the first $i-1$ triggers of $\mathcal{D}_A$. By the inductive hypothesis, since such atoms are ancestors of A, they have the same rank in $\mathcal{D}$ and $\mathcal{D}_A$. Hence, all ancestors of A in $output(t_i)$ have the same rank in $\mathcal{D}$ and $\mathcal{D}_A$.

Now let $\mathcal{D}'$ be any **bf-O**-derivation from $(Anc^0_{\mathcal{D}}(A), \mathcal{R})$. This can be seen as an **O**-derivation from $(F, \mathcal{R})$. So, by Proposition 15, the ranks of atoms in $\mathcal{D}'$ can only increase with respect to $\mathcal{D}$. But again by Proposition 15, the ranks of atoms in $\mathcal{D}_A$ can only increase with respect to $\mathcal{D}'$. Since the ranks of the ancestors of A are the same in $\mathcal{D}_A$ and $\mathcal{D}$, we conclude that so is the case for $\mathcal{D}'$. □

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

We are now ready to prove that the breadth-first semi-oblivious and restricted chase variants preserve ancestry.

Theorem 32
The **bf-**X-chase preserves ancestry when $X \in \{\mathbf{SO}, \mathbf{R}\}$.

Proof
We assume that $\mathcal{D}$ is an X-derivation from $(F, \mathcal{R})$ and t is a trigger that produces an atom A of rank $n > 0$ in $\mathcal{D}$.

We start from the derivation $\mathcal{D}_{|Anc^0_{\mathcal{D}}(A)}$ and remove all triggers that do not produce any ancestor of A in $\mathcal{D}$. Let $\mathcal{D}''$ be the obtained derivation. We build a derivation $\mathcal{D}'$ by applying the following completion procedure on $\mathcal{D}''$. For each rank $1 \leq k \leq n$, we first extend $\mathcal{D}'$ by applying all triggers of rank k in $\mathcal{D}''$, in the given order. Then, within the same rank, we further extend the resulting derivation by non-deterministically applying any sequence of X-applicable triggers of rank k until no trigger of rank k is X-applicable anymore. We repeat this process rank by rank.

We claim that $\mathcal{D}'$ is a **bf-**X-derivation. The first thing to show is that $\mathcal{D}'$ is rank-compatible. To do so, first notice that completion yields a derivation from $(Anc^0_{\mathcal{D}}(A), \mathcal{R})$ which, by Proposition 15, produces a subset of the atoms produced by a **bf-O**-derivation from $(Anc^0_{\mathcal{D}}(A), \mathcal{R})$ at the same rank. Then, by Proposition 31, the ranks of the ancestors of A are the same on $\mathcal{D}$ and the **bf-O** derivation, thus $rank_{\mathcal{D}}(t) = rank_{\mathcal{D}'}(t)$ for all $t \in triggers(\mathcal{D}'')$. We conclude that $\mathcal{D}'$ is rank-compatible, as the completion does not alter the ranks of the triggers. By construction, every rank in $\mathcal{D}'$ cannot be extended with other X-applicable triggers. So, if $\mathcal{D}'$ is an X-derivation, then it respects both conditions for being a breadth-first X-derivation. What remains to be shown is that all triggers in $\mathcal{D}'$ (and specifically those of $\mathcal{D}''$) are X-applicable on their respective prefix. We perform a case analysis on the type of chase.

(Case $X = \mathbf{SO}$)

We say that (R, π) and (R, π') are **SO**-equivalent if $\pi(x) = \pi'(x)$ for all $x \in fr(R)$.

We start from the breadth-first completion $\mathcal{D}'$ that has been described above. Let t be the first trigger of $\mathcal{D}'$ at position $i+1$ in $triggers(\mathcal{D}')$ that is not **SO**-applicable on $\mathcal{D}'_{|i}$. Then, there exists a trigger t' in $\mathcal{D}'_{|i}$ which is **SO**-equivalent with t. Notice that, by construction of $\mathcal{D}'$, all elements of $triggers(\mathcal{D}') \setminus triggers(\mathcal{D})$ are **SO**-applicable on their respective prefixes, so $t \in triggers(\mathcal{D})$. On the contrary, t' cannot occur in $\mathcal{D}$ because t occurs in $\mathcal{D}$ and no pair of distinct triggers in $\mathcal{D}$ can be **SO**-equivalent. Therefore, t' has been introduced by the completion procedure. According to this procedure, all triggers of $\mathcal{D}'$ that precede t within the same rank as t also occur in $\mathcal{D}$. Hence, the rank of t' in $\mathcal{D}'$ is strictly lower than that of t. Now, because $\mathcal{D}'_{|i}$ is a **bf-SO**-derivation, we can apply Propositions 31 and 17 to conclude that $rank_{\mathcal{D}}(t) = rank_{\mathcal{D}'}(t)$, so also $rank_{\mathcal{D}'}(t') < rank_{\mathcal{D}}(t)$.

At this point, we do not know whether t' was applicable on some prefix of $\mathcal{D}$. This would be the case if $support(t')$ was produced by $\mathcal{D}$. However, $support(t')$ may contain atoms that use some fresh terms introduced by a previous trigger added by the completion. Nevertheless, using Proposition 16, we can find a trigger that is **SO**-equivalent with t' (so also with t) and has to be applicable on some prefix of $\mathcal{D}$, leading to a contradiction. In particular, this proposition implies that there is a retraction h from $F^{\mathcal{D}'_{|i}} \cup F^{\mathcal{D}}$ to $F^{\mathcal{D}}$

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

that maps $support(t')$ to atoms of equal or lower rank in $\mathcal{D}$. Suppose that $t = (R, \pi)$ and $t' = (R, \pi')$. The SO-equivalence of t and t' guarantees that π and π' agree on the mapping of all frontier variables of R. Those variables are necessarily mapped by π (so also by π') to terms of $F^{\mathcal{D}}$. Let $t'' = (R, h \circ \pi')$. We have that $support(t'') = h(support(t'))$ so t'' is applicable on $F^{\mathcal{D}}$, with $rank_{\mathcal{D}}(t'') \leq rank_{\mathcal{D}'}(t')$ (again, the latter is a consequence of Proposition 16).

Since h does not affect any variable of $F^{\mathcal{D}}$, it does not affect the mapping of the frontier variables of R, that is, for every $x \in fr(R)$, $h \circ \pi'(x) = \pi'(x)$. This implies that t'' is **SO**-equivalent with t', so also with t. Let $\mathcal{D}''$ be the prefix of $\mathcal{D}$ with all elements of rank strictly less than $rank_{\mathcal{D}}(t)$. $\mathcal{D}''$ does not include t nor any trigger of the same **SO**-equivalence class. But $support(t'') = h(support(t')) \subseteq F^{\mathcal{D}}$ and, from $rank_{\mathcal{D}}(t'') < rank_{\mathcal{D}}(t)$, we know in particular that $support(t'') \subseteq F^{\mathcal{D}''}$. So t'' is **SO**-applicable on $\mathcal{D}''$. But we know that $t'' \notin triggers(\mathcal{D})$ because it is **SO**-equivalent with t. This is a contradiction because $\mathcal{D}$ is breadth-first, so t'' must have been applied at its respective rank. We conclude that all the triggers of $\mathcal{D}'$ are **SO**-applicable on their respective prefix and $\mathcal{D}'$ is a **bf-SO**-derivation.

(Case $X = \mathbf{R}$)

Let t_{i+1} be the first trigger of $\mathcal{D}'$ at position $i+1$ in $triggers(\mathcal{D}')$ that is not **R**-applicable on $\mathcal{D}'_{|i}$. Assume that t_{i+1} is of rank $m+1$ in $\mathcal{D}'$ and let $\mathcal{D}'_{|depth(m)}$ be the maximal prefix of $\mathcal{D}'$ of depth m. Hence, $F^{\mathcal{D}'_{|i}} = F^{\mathcal{D}'_{|depth(m)}} \cup output(t_{n+1}) \cup \cdots \cup output(t_i)$ where, by construction of $\mathcal{D}'$, the (possibly empty) sequence of triggers $t_{n+1}, \ldots, t_i$ of $\mathcal{D}'$ have the same rank as t_{i+1}.

Now, let $\mathcal{D}_{|depth(m)}$ be the maximal prefix of $\mathcal{D}$ of depth m. Of course $\mathcal{D}_{|depth(m)}$ is a breadth-first **R**-derivation with the same depth as $\mathcal{D}'_{|depth(m)}$. Let $\mathcal{D}''$ be the derivation from $(F, \mathcal{R})$ with $triggers(\mathcal{D}'') = triggers(\mathcal{D}')$. Then $\mathcal{D}''$ produces the same atoms as $\mathcal{D}'$ and does not have greater depth. So, by Proposition 16, there is a retraction h from $F^{\mathcal{D}''_{|depth(m)}} \cup F^{\mathcal{D}_{|depth(m)}}$ to $F^{\mathcal{D}_{|depth(m)}}$, so also from $F^{\mathcal{D}'_{|depth(m)}} \cup F^{\mathcal{D}_{|depth(m)}}$ to $F^{\mathcal{D}_{|depth(m)}}$.

Recall that $t_{n+1}, \ldots, t_{i+1}$ are the first triggers at the beginning of rank $m+1$ in $\mathcal{D}'$. These triggers are necessarily producing an ancestor of A since it is the case for t_{i+1}. And since $\mathcal{D}'$ is rank-compatible, they also have rank $m+1$ in $\mathcal{D}$. By Proposition 31, the ranks of the ancestors of A are the same in $\mathcal{D}$ and $\mathcal{D}'$. So the support of each trigger is in $F^{\mathcal{D}_{|depth(m)}}$ and since (i) h preserves the terms of $F^{\mathcal{D}_{|depth(m)}}$ and (ii) fresh nulls are named after the trigger that generated them, we deduce that h behaves as the identity on $output(t_{n+1}) \cup \cdots \cup output(t_{i+1})$. We get $h(F^{\mathcal{D}'_{|i}}) = h(F^{\mathcal{D}'_{|depth(m)}} \cup output(t_{n+1}) \cup \cdots \cup output(t_i)) \subseteq F^{\mathcal{D}_{|depth(m)}} \cup output(t_{n+1}) \cup \cdots \cup output(t_i) \subseteq F^{\mathcal{D}_{|j}}$ where $j+1$ is the position of t_{i+1} in $triggers(\mathcal{D})$.

Now, as t_{i+1} is not **R**-applicable on $\mathcal{D}'_i$, there is a retraction σ from $F^{\mathcal{D}'_{|i+1}}$ to $F^{\mathcal{D}'_{|i}}$. But this means that t_{i+1} was not applicable on the prefix $\mathcal{D}_{|j}$ of $\mathcal{D}$ because of the retraction $h \circ \sigma$ from $F^{\mathcal{D}_{|j+1}}$ to $F^{\mathcal{D}_{|j}}$; indeed, $h \circ \sigma(output(t_{i+1})) \subseteq h(F^{\mathcal{D}'_{|i}}) \subseteq F^{\mathcal{D}_{|j}}$. This is a contradiction. □

4.2 Heredity

Heredity is a second property that leads to the decidability of k-boundedness. A chase variant X is hereditary if by restricting an X-derivation to a subset of a factbase, we still obtain an X-derivation.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Definition 33 (*Heredity*)
The X-chase is *hereditary* if, for any X-derivation $\mathcal{D}$ from $(F, \mathcal{R})$ and subset $G \subseteq F$, the restriction $\mathcal{D}_{|G}$ is also an X-derivation.

This property is satisfied by the oblivious, the semi-oblivious, and the restricted chase variants.

Theorem 34
The X-chase is hereditary for $X \in \{\mathbf{O}, \mathbf{bf\text{-}O}, \mathbf{SO}, \mathbf{R}\}$.

Proof
Let $\mathcal{D}$ be an X-derivation from $(F, \mathcal{R})$ and $G \subseteq F$. We do a case analysis on X.
(Case $X = \mathbf{O}$) Clearly $\mathcal{D}_{|G}$ is an $\mathbf{O}$-chase derivation; therefore, the $\mathbf{O}$-chase is hereditary.
(Case $X = \mathbf{bf\text{-}O}$) Since $\mathcal{D}$ is rank-compatible and since the ordering of triggers is preserved in $\mathcal{D}_{|G}$, we get that $\mathcal{D}_{|G}$ is rank-compatible. Moreover, because $\mathcal{D}$ is a $\mathbf{bf\text{-}O}$-derivation, all triggers which are applicable on atoms whose prime ancestors are in G are also applicable on $\mathcal{D}_{|G}$. Therefore $\mathcal{D}_{|G}$ is also breadth-first, since at every rank, all possible rule applications are performed.
(Case $X = \mathbf{SO}$) The condition for $\mathbf{SO}$-applicability is that we do not have two triggers from the same rule mapping frontier variables in the same way. We know that $\mathit{triggers}(\mathcal{D})$ satisfies this condition, hence so does its subseqence $\mathit{triggers}(\mathcal{D}_{|G})$.
(Case $X = \mathbf{R}$) Let $\mathcal{D}'$ be a prefix of $\mathcal{D}$. The condition for $\mathbf{R}$-applicability of a trigger t on $\mathcal{D}'$ is that there is no retraction from the immediate derivation from $F^{\mathcal{D}'}$ through t back to $F^{\mathcal{D}'}$. Since $\mathcal{D}_{|G}$ generates factbases that are included in the factbases generated by $\mathcal{D}$ at the moment of the application of the same trigger t, we conclude that $\mathbf{R}$-applicability is preserved. □

Preservation of ancestry is a generalization of heredity, as we state in Theorem 35. Moreover, this generalization is strict, since the breadth-first semi-oblivious and breadth-first restricted variants preserve ancestry but are not hereditary, as shown in Examples 26 and 27.

Theorem 35
Every hereditary chase variant preserves ancestry.

Proof
Let X be a hereditary chase variant. Let $\mathcal{D}$ be an X-derivation from $(F, \mathcal{R})$. Let $G = \mathit{Anc}^0_{\mathcal{D}}(A)$. Since X is hereditary, $\mathcal{D}_{|G}$ is an X-derivation. By Proposition 25, $\mathcal{D}_{|G}$ produces at least $\mathit{Anc}^0_{\mathcal{D}} \cup \{A\}$. To conclude, we must show that $\mathit{rank}_{\mathcal{D}}(A) = \mathit{rank}_{\mathcal{D}_{|G}}(A)$.

Given (t_i, F_i) an element of a derivation $\bar{\mathcal{D}}$ such that $i \geq 1$, we denote $\mathit{new}_{\bar{\mathcal{D}}}(t_i) = \mathit{output}(t_i) \setminus F_{i-1}$, the set of atoms produced by t_i in $\bar{\mathcal{D}}$, that is, all atoms in the specialization of the rule head used by the trigger that did not already appear in the last factbase. We first show that $\mathit{new}_{\mathcal{D}}(t_i) \subseteq \mathit{new}_{\mathcal{D}_{|G}}(t_i)$, for all $t_i \in \mathit{triggers}(\mathcal{D}_{|G})$. That is, the triggers in the restriction of a derivation potentially produce more facts. Let $\ell \geq i$ be the position of t_i in $\mathit{triggers}(\mathcal{D})$. Then, $G_{i-1} = G \cup \bigcup_{j<i} \mathit{output}(t_j)$ and $F_{\ell-1} \subseteq F \cup \bigcup_{j<i} \mathit{output}(t_j)$. It follows that $\mathit{new}_{\mathcal{D}}(t_i) = \mathit{output}(t_i) \setminus F_{\ell-1} \subseteq \mathit{output}(t_i) \setminus G_{i-1} = \mathit{new}_{\mathcal{D}_{|G}}(t_i)$.

We now show that, for all $A_1, A_2 \in \mathit{Anc}_{\mathcal{D}}(A) \cup \{A\}$, if A_1 is a direct ancestor of A_2 in $\mathcal{D}$, then A_1 is also a direct ancestor of A_2 in $\mathcal{D}_{|G}$. Let t be the trigger that produces A_2

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

in $\mathcal{D}$. We know that $t \in \mathit{triggers}(\mathcal{D}_{|G})$. Since $\mathit{new}_{\mathcal{D}}(t) \subseteq \mathit{new}_{\mathcal{D}_{|G}}(t)$ and $A_2 \in \mathit{new}_{\mathcal{D}}(t)$, the trigger t produces A_2 in $\mathcal{D}_{|G}$, that is, A_1 is a direct ancestor of A_2 in $\mathcal{D}_{|G}$.

Hence, the subgraphs induced by $\mathit{Anc}_{\mathcal{D}}(A) \cup \{A\}$ and $\mathit{Anc}_{\mathcal{D}_{|G}}(A) \cup \{A\}$ in their respective chase graphs coincide, and the rank of A in both derivations is the same. We have shown that there is a derivation from $(\mathit{Anc}^0_{\mathcal{D}}(A), \mathcal{R})$ that produces A in the same rank as $\mathcal{D}$, so the X-chase preserves ancestry. □

Gathering the previous results, we can now state that $\forall$-X-k-boundedness is decidable for all identified chase variants that preserve ancestry.

Corollary 36
$\forall$-X-k-boundedness is decidable when $X \in \{\mathbf{O}, \mathbf{bf\text{-}O}, \mathbf{SO}, \mathbf{bf\text{-}SO}, \mathbf{R}, \mathbf{bf\text{-}R}\}$.

Proof
By Theorem 32, **bf-SO** and **bf-R** preserve ancestry. By Theorem 34, **O**, **bf-O**, **SO** and **R** are hereditary; hence, they preserve ancestry (by Theorem 35). Finally, by Theorem 29, preservation of ancestry is a sufficient condition for the decidability of $\forall$-X-k-boundedness. □

Naturally, this implies the decidability of $\exists$-X-k-boundedness when both problems are equivalent, which is in particular the case for the chase variants identified in Propositions 22.

Corollary 37
$\exists$-X-k-boundedness is decidable when $X \in \{\mathbf{O}, \mathbf{bf\text{-}O}, \mathbf{SO}, \mathbf{bf\text{-}SO}\}$.

The arguments used to show the decidability of $\forall$-X-k-boundedness also allows one to upper-bound the complexity of the problem.

Proposition 38
The $\forall$-X-k-boundedness problem (where k is unary-encoded) is in 2-EXPTIME for $X \in \{\mathbf{bf\text{-}O}, \mathbf{bf\text{-}SO}\}$ and in 3-EXPTIME for $X \in \{\mathbf{O}, \mathbf{SO}, \mathbf{R}, \mathbf{bf\text{-}R}\}$.

Proof
Clearly, **bf-O** and **bf-SO** always produce isomorphic factbases at the same rank. Therefore, to decide whether all **bf-O**-derivations from a given factbase F have depth at most k, it is sufficient to build one **bf-O**-derivation of depth at most k and, if the depth k is reached, check whether one trigger is still **O**-applicable, in which case we obtain a counter-example to the $\forall$-**bf-O**-k-boundedness. The same holds for the **bf-SO**-chase. On the contrary, the depth of the derivations for the other variants depends on the order in which triggers are applied. Hence, for a given factbase, we have to check all possible orders on triggers, instead of building a single derivation.

Given a ruleset $\mathcal{R}$ with at most b atoms in the bodies of its rules, the number of factbases of size at most b^k is in the worst-case double exponential with respect to k. The length of a derivation of depth at most k from a given KB is also double exponential in k. Finally, the number of all the different derivations of depth at most k is triple exponential in k. Hence, the complexity of the $\forall$-X-k-boundedness problem (where k is unary-encoded) is in 2-EXPTIME for $X \in \{\mathbf{bf\text{-}O}, \mathbf{bf\text{-}SO}\}$ and in 3-EXPTIME for $X \in \{\mathbf{O}, \mathbf{SO}, \mathbf{R}, \mathbf{bf\text{-}R}\}$. □

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

4.3 The **E***-chase and the* **bf-E***-chase do not preserve ancestry*

Finally, the following example shows that the **E** and **bf-E**-chase do not preserve ancestry. Hence, the decidability of X-k-boundedness for those variants remains an open question.

Example 39
Let $(F, \mathcal{R})$ be a KB, where $F = \{p(x_1, x_1), p(x_2, x_2), r(x_1, x_2)\}$ and $\mathcal{R} = \{R\}$, where $R = p(x, x) \wedge p(y, y) \rightarrow p(x, y) \wedge p(y, x)$. Let $t_1 = (R, \{x \mapsto x_1, y \mapsto x_2\})$. Then $\mathcal{D} = (\emptyset, F), (t_1, F \cup \{p(x_1, x_2), p(x_2, x_1)\})$ is a (terminating) **bf-E**-derivation from $(F, \mathcal{R})$. The set of ancestors of the atom $p(x_1, x_2)$ (of rank 1 in $\mathcal{D}$) is $F' = \{p(x_1, x_1), p(x_2, x_2)\}$. However, there is no **E**-derivation of depth 1 from $(F', \mathcal{R})$ because the obtained factbase would be equivalent to F'. Therefore, neither the **E**-chase nor the **bf-E**-chase preserve ancestry.

5 Concluding remarks

Let us situate this work in the broader context of ontology-mediated query answering, where the ontology is a set of existential rules. The main techniques investigated to address this issue are either based on the chase or on query rewriting, which may both not terminate, since even ground atom entailment from an existential rule base is undecidable (from, e.g., Beeri and Vardi 1981). However, a wide range of syntactic conditions on rulesets have been defined, which ensure either termination of some chase variant or first-order rewritability for conjunctive queries. A third family of syntactic conditions rely on another decidability paradigm, namely guardedness and its extensions (Calì *et al*. 2008; Baget *et al*. 2010; Krötzsch and Rudolph 2011; Baget *et al*. 2011; Thomazo *et al*. 2012).

Actually, most sufficient conditions for chase termination apply to the (semi-)oblivious chase: from the simplest ones, namely rich-acyclicity (Hernich and Schweikardt 2007), weak-acyclicity (Fagin *et al*. 2005), and refinements like joint-acyclicity (Krötzsch and Rudolph 2011) or super-weak-acyclicity (Marnette 2009), acyclic-Graph of rule dependency (GRD) (Deutsch *et al*. 2008; Baget *et al*. 2011), to combinations of acyclicity criteria (Baget *et al*. 2014) and model-faithful acyclicity (MFA), which strictly generalizes all the previous acyclicity conditions (Cuenca Grau *et al*. 2013). Recently, MFA was extended to restricted-MFA, which ensures the termination of a specific restricted chase algorithm, even on KBs for which the semi-oblivious chase may not terminate (Carral *et al*. 2017). While boundedness was deeply investigated for Datalog, its study for existential rules is only beginning. Among known classes of rules that ensure chase termination, only acyclic-GRD ensures boundedness (of any chase variant), and boundedness cannot be decided for the other classes as they all generalize Datalog. Very recently, work reported in Bourhis *et al*. (2019) gave a characterization of boundedness in terms of chase termination and first-order rewritability. Precisely, a ruleset is X-bounded, for X being the breadth-first oblivious or semi-oblivious chase, if and only if it ensures both X-chase termination and first-order rewritability of conjunctive queries. This characterization allows one to obtain the decidability of (semi-)oblivious-boundedness for classes having decidable chase termination and decidable first-order rewritability, such as the important classes of sticky and guarded existential rules.

In this article, which extends the work presented in Delivorias *et al*. (2018), we have followed another path: instead of specific classes of existential rules, we have considered

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

the weaker problem of k-boundedness. We have shown that k-boundedness is decidable for the main chase variants: oblivious, semi-oblivious, and restricted chase, as well as their breadth-first versions. These results rely on establishing a common property that ensures the decidability of k-boundedness, namely "preservation of ancestry." We note that results concerning the semi-oblivious chase also apply to the logic programs associated with existential rules, since the semi-oblivious chase behaves as the Skolem chase.

We leave for further work the study of the precise complexity of deciding k-boundedness according to each kind of chase. Also, the decidability of $\exists$-**R**-k-boundedness as well as that of $\exists$-**bf-R**-k-boundedness remain open issues. Finally, we leave open the question of the decidability of the k-boundedness for chase variants that detect more redundancies, such as the equivalent chase and the core chase.

References

Abiteboul, S. 1989. Boundedness is undecidable for datalog programs with a single recursive rule. *Information Processing Letters 32,* 6, 281–287.

Abiteboul, S., Hull, R. and Vianu, V. 1995. *Foundations of Databases: The Logical Level*, 1st ed. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Baader, F., Brandt, S. and Lutz, C. 2005. Pushing the EL envelope. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJCAI'05*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 364–369.

Baget, J., Garreau, F., Mugnier, M. and Rocher, S. 2014. Extending acyclicity notions for existential rules. In *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic – Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, 39–44.

Baget, J., Leclère, M. and Mugnier, M. 2010. Walking the decidability line for rules with existential variables. In *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning, KR'10*. AAAI Press, 466–476.

Baget, J., Leclère, M., Mugnier, M. and Salvat, E. 2009. Extending decidable cases for rules with existential variables. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence, IJCAI'09*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 677–682.

Baget, J., Mugnier, M., Rudolph, S. and Thomazo, M. 2011. Walking the complexity lines for generalized guarded existential rules. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16–22, 2011*, 712–717.

Baget, J.-F., Leclère, M., Mugnier, M.-L. and Salvat, E. 2011. On rules with existential variables: Walking the decidability line. *Artificial Intelligence 175,* 9–10, 1620–1654.

Beeri, C. and Vardi, M. Y. 1981. The implication problem for data dependencies. In *International Colloquium on Automata, Languages, and Programming*. Springer, 73–85.

Beeri, C. and Vardi, M. Y. 1984. A proof procedure for data dependencies. *Journal of the Association for Computing Machinery 31,* 4 (September), 718-741.

Bourhis, P., Leclère, M., Mugnier, M., Tison, S., Ulliana, F. and Gallois, L. 2019. Oblivious and semi-oblivious boundedness for existential rules. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10–16, 2019*, 1581–1587.

Calì, A., Gottlob, G. and Kifer, M. 2008. Taming the infinite chase: Query answering under expressive relational constraints. In *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning, KR'08*. AAAI Press, 70-80.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Calì, A., Gottlob, G. and Lukasiewicz, T. 2009. A general datalog-based framework for tractable query answering over ontologies. In *Proceedings of the Twenty-Eighth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS'09*. Association for Computing Machinery, New York, NY, USA, 77-86.

Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M. and Rosati, R. 2005. DL-Lite: Tractable description logics for ontologies. In *Proceedings of the 20th National Conference on Artificial Intelligence – Volume 2, AAAI'05*. AAAI Press, 602–607.

Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M. and Rosati, R. 2007. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated Reasoning (JAR) 39,* 3, 385–429.

Carral, D., Dragoste, I. and Krötzsch, M. 2017. Restricted 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*, 922–928.

Chandra, A. K. and Merlin, P. M. 1977. Optimal implementation of conjunctive queries in relational data bases. In *Proceedings of the 9th Annual ACM Symposium on Theory of Computing, May 4–6, 1977, Boulder, Colorado, USA*, 77–90.

Cosmadakis, S. S., Gaifman, H., Kanellakis, P. C. and Vardi, M. Y. 1988. Decidable optimization problems for database logic programs (preliminary report). In *ACM Symposium on Theory of Computing*, 477–490.

Cuenca Grau, B., Horrocks, I., Krötzsch, M., Kupke, C., Magka, D., Motik, B. and Wang, Z. 2013. Acyclicity notions for existential rules and their application to query answering in ontologies. *Journal of Artificial Intelligence Research 47,* 1 (May), 741-808.

Delivorias, S., Leclère, M., Mugnier, M. and Ulliana, F. 2018. On the k-boundedness for existential rules. In *Computing Research Repository* (initially published in *Rules and Reasoning – Second International Joint Conference, RuleML+RR 2018, Luxembourg, September 18–21, 2018, Proceedings*) *abs/1810.09304*.

Deutsch, A., Nash, A. and Remmel, J. B. 2008. The chase revisited. In *Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS '08*. Association for Computing Machinery, New York, NY, USA, 149-158.

Fagin, R., Kolaitis, P. G., Miller, R. J. and Popa, L. 2005. Data exchange: Semantics and query answering. *Theoretical Computer Science 336,* 1 (May), 89-124.

Gaifman, H., Mairson, H. G., Sagiv, Y. and Vardi, M. Y. 1993. Undecidable optimization problems for database logic programs. *Journal of the Association for Computing Machinery 40,* 3 (July), 683–713.

Gogacz, T. and Marcinkowski, J. 2014. All-instances termination of chase is undecidable. In *International Colloquium on Automata, Languages, and Programming 2014 Proceedings, Part II*, 293–304.

Gottlob, G., Orsi, G., Pieris, A. and Simkus, M. 2012. Datalog and its extensions for semantic web databases. In *Reasoning Web*, T. Eiter and T. Krennwallner, Eds. Springer Berlin Heidelberg, Berlin, Heidelberg, 54–77.

Grahne, G. and Onet, A. 2018. Anatomy of the chase. *Fundamenta Informaticae 157,* 3, 221–270.

Guessarian, I. and Peixoto, M. V. 1994. About boundedness for some datalog and DATALOG$^{\text{neg}}$ programs. *Journal of Logic and Computation 4,* 4, 375–403.

Hernich, A. and Schweikardt, N. 2007. 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*, 113–122.

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

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press

Krötzsch, M. and Rudolph, S. 2011. Extending decidable existential rules by joining acyclicity and guardedness. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence – Volume Volume Two*. IJCAI'11. AAAI Press, 963–968.

Krötzsch, M., Rudolph, S. and Hitzler, P. 2007. Complexity boundaries for Horn description logics. In *Proceedings of the 22nd National Conference on Artificial Intelligence – Volume 1, AAAI'07*. AAAI Press, 452–457.

Leclère, M., Mugnier, M. and Ulliana, F. 2016. On bounded positive existential rules. In *Proceedings of the 29th International Workshop on Description Logics*.

Leone, N., Manna, M., Terracina, G. and Veltri, P. 2012. Efficiently computable Datalog∃ programs. In *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning, KR'12*. AAAI Press, 13–23.

Marcinkowski, J. 1999. Achilles, turtle, and undecidable boundedness problems for small datalog programs. *Society for Industrial and Applied Mathematics Journal on Computing 29,* 1, 231–257.

Marnette, B. 2009. Generalized schema-mappings: From termination to tractability. In *Proceedings of the Twenty-Eighth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS'09*. Association for Computing Machinery, New York, NY, USA, 13–22.

Mugnier, M. and Thomazo, M. 2014. An introduction to ontology-based query answering with existential rules. In *Reasoning Web 2014*, M. Koubarakis, G. Stamou, G. Stoilos, I. Horrocks, P. Kolaitis, G. Lausen and G. Weikum, Eds. Springer International Publishing, Cham, 245–278.

Rocher, S. 2016. Querying existential rule knowledge bases: Decidability and complexity *(Interrogation de Bases de Connaissances avec Règles Existentielles : Décidabilité et Complexité)*. Ph.D. thesis, University of Montpellier, France.

Thomazo, M., Baget, J., Mugnier, M. and Rudolph, S. 2012. A generic querying algorithm for greedy sets of existential rules. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10–14, 2012*.

https://doi.org/10.1017/S1471068420000083 Published online by Cambridge University Press