

# Restricted Chase Termination: You Want More than Fairness

DAVID CARRAL, LIRMM, Inria, University of Montpellier, CNRS, France

LUKAS GERLACH, Knowledge-Based Systems Group, TU Dresden, Germany

LUCAS LARROQUE, Inria, DI ENS, ENS, CNRS, PSL University, France

MICHAËL THOMAZO, Inria, DI ENS, ENS, CNRS, PSL University, France

The chase is a fundamental algorithm with ubiquitous uses in database theory. Given a database and a set of existential rules (aka tuple-generating dependencies), it iteratively extends the database to ensure that the rules are satisfied in a most general way. This process may not terminate, and a major problem is to decide whether it does. This problem has been studied for a large number of chase variants, which differ by the conditions under which a rule is applied to extend the database. Surprisingly, the complexity of the universal termination of the restricted (aka standard) chase is not fully understood. We close this gap by placing universal restricted chase termination in the analytical hierarchy. This higher hardness is due to the fairness condition, and we propose an alternative condition to reduce the hardness of universal termination.

CCS Concepts: • Theory of computation → Constraint and logic programming; Logic and databases.

Additional Key Words and Phrases: Existential Rules, Tuple Generating Dependencies, Restricted Chase

## ACM Reference Format:

David Carral, Lukas Gerlach, Lucas Larroque, and Michaël Thomazo. 2025. Restricted Chase Termination: You Want More than Fairness. *Proc. ACM Manag. Data* 3, 2 (PODS), Article 109 (May 2025), 29 pages. <https://doi.org/10.1145/3725246>

## 1 Introduction

The chase is a fundamental algorithm in database theory that is applied to address a wide range of problems. For instance, it is used to check containment of queries under constraints, in data exchange settings, or to solve ontology-based query answering; see the introductions of [11, 13] for more information. Technically speaking, the chase is a bottom-up materialisation procedure that attempts to compute a universal model (a model that can be embedded into all other models via homomorphism) for a knowledge base (KB), consisting of an (existential) rule set<sup>1</sup> and a database.

**Example 1.** Consider the KB  $\mathcal{K}_1 = \langle \Sigma, D \rangle$  where  $D$  is the database  $\{\text{Bicycle}(b)\}$  and  $\Sigma$  contains:

$$\begin{aligned} \forall x. \text{Bicycle}(x) \rightarrow \exists y. \text{HasPart}(x, y) \wedge \text{Wheel}(y) & \quad \forall x, y. \text{HasPart}(x, y) \rightarrow \text{IsPartOf}(y, x) \\ \forall x. \text{Wheel}(x) \rightarrow \exists y. \text{IsPartOf}(x, y) \wedge \text{Bicycle}(y) & \quad \forall x, y. \text{IsPartOf}(x, y) \rightarrow \text{HasPart}(y, x) \end{aligned}$$

Then,  $\{\text{Bicycle}(b), \text{HasPart}(b, t), \text{IsPartOf}(t, b), \text{Wheel}(t)\}$  is a universal model of  $\mathcal{K}$ .

<sup>1</sup>Other researchers refer to these first-order formulas as “tuple generating dependencies” or simply as “TGDs”.

Authors’ Contact Information: David Carral, LIRMM, Inria, University of Montpellier, CNRS, Montpellier, France, [david.carral@inria.fr](mailto:david.carral@inria.fr); Lukas Gerlach, Knowledge-Based Systems Group, TU Dresden, Dresden, Germany, [lukas.gerlach@tu-dresden.de](mailto:lukas.gerlach@tu-dresden.de); Lucas Larroque, Inria, DI ENS, ENS, CNRS, PSL University, Paris, France, [lucas.larroque@inria.fr](mailto:lucas.larroque@inria.fr); Michaël Thomazo, Inria, DI ENS, ENS, CNRS, PSL University, Paris, France, [michael.thomazo@inria.fr](mailto:michael.thomazo@inria.fr).



This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2025 Copyright held by the owner/author(s).

ACM 2836-6573/2025/5-ART109

<https://doi.org/10.1145/3725246>

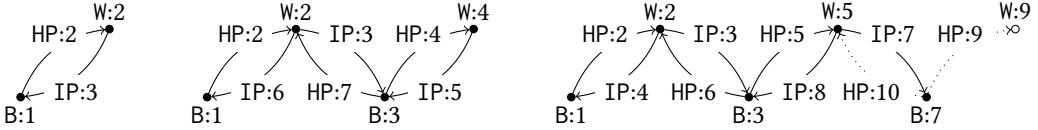


Fig. 1. Three Different Restricted Chase Sequences for the KB  $\mathcal{K}_1$  from Example 1

Although there are many variants of the chase, they all implement a similar strategy. Namely, they start with the database and then, in a step-by-step manner, extend this structure with new atoms to satisfy the rules in the input rule set in a most general way. Since none of these variants are guaranteed to terminate (some KBs do not even admit finite universal models), it is only natural to wonder about their respective halting problems [1, 5, 6, 10, 13, 18]. Despite intensive efforts, some results have remained open (until now!). Specifically, prior research has established tight bounds for all classes of chase terminating KBs and rule sets, except for the following:

- The class  $\text{CTK}_{\vee}^{\text{rest}}$  of all KBs that only admit finite restricted chase sequences.
- The class  $\text{CTR}_{\vee}^{\text{rest}}$  containing a rule set  $\Sigma$  if  $\langle \Sigma, D \rangle \in \text{CTK}_{\vee}^{\text{rest}}$  for every database  $D$ .

Our main contribution is to show that both classes are  $\Pi_1^1$ -complete, a surprising result given that these are significantly harder than the corresponding classes for other chase variants [13].

The restricted chase differs from other variants in that it introduces new terms to satisfy existential quantifiers in rules only if these are not already satisfied by existing terms. Because of this, the order of rule applications impacts the termination of a chase sequence. For instance, the KB  $\mathcal{K}_1$  from Example 1 admits finite and infinite restricted chase sequences; some of these are represented in Fig. 1, where atoms are numbered to denote the sequence step at which they were introduced.

$\text{CTK}_{\vee}^{\text{rest}}$  has been claimed to be recursively enumerable (RE) in [13], probably with the following procedure in mind: given an input KB, compute all of its restricted chase sequences in parallel, and halt and accept if all of them are finite. Alas, this strategy does not work as there are terminating input KBs that admit infinitely many finite sequences that are of ever-increasing length.

**Example 2.** Consider the KB  $\mathcal{K}_2 = \langle \Sigma, D \rangle$  where  $D$  is the database  $\{\text{Real}(a), \text{E}(a, c), \text{E}(c, b), \text{Real}(c), \text{E}(b, b), \text{Brake}(b)\}$  and  $\Sigma$  is the rule set that contains all of the following:

$$\begin{aligned} \forall x, y, z. \text{Real}(x) \wedge \text{E}(x, y) \wedge \text{Real}(y) \wedge \text{Brake}(z) &\rightarrow \exists v. \text{E}(y, v) \wedge \text{E}(v, z) \wedge \text{Real}(v) \\ \forall x. \text{Brake}(x) &\rightarrow \text{Real}(x) \end{aligned}$$

For any  $k \geq 1$ , there is a restricted chase sequence of  $\mathcal{K}_2$  that yields the (finite) universal model  $D \cup \{\text{E}(c, t_1)\} \cup \{\text{E}(t_i, t_{i+1}) \mid i < k\} \cup \{\text{E}(t_i, b), \text{Real}(t_i) \mid i \leq k\} \cup \{\text{Real}(b)\}$  of  $\mathcal{K}_2$ . Such a sequence is obtained by applying the first rule  $k$  consecutive times and then applying the second one once to derive  $\text{Real}(b)$ . After this application, the first rule is satisfied and the restricted chase halts.

The KB  $\mathcal{K}_2$  in the previous example is in  $\text{CTK}_{\vee}^{\text{rest}}$  because of *fairness*. This is a built-in condition in the definition of all chase variants that guarantees that the chase yields a model of the KB by requiring that, if a rule is applicable at some point during the computation of a sequence, then this rule must be eventually satisfied. Hence, the second rule in  $\mathcal{K}_2$  must sooner or later be applied in all restricted chase sequences and thus, all such sequences are finite.

The KB in Example 2 uses a technique called the *emergency brake*, initially proposed by Krötzsch et al. in [16]. The idea is to connect every term in the chase to a special term (the constant  $b$  in this example) that is not “Real” and acts as a “Brake”. Eventually, this term becomes “Real” because of fairness, all existential restrictions are satisfied, and the restricted chase halts. The emergency brake allows to grow the chase for an arbitrary number of steps whilst guaranteeing its termination.

By activating infinite sequences of emergency brakes, we emulate the eternal recurrence often displayed by  $\Pi_1^1$ -complete problems and thus define the reductions that lead to our main results.

After presenting the necessary preliminaries in Section 2, we discuss related work in Section 3. Then, we show that  $CTK_{\exists}^{rest}$  and  $CTR_{\forall}^{rest}$  are  $\Pi_1^1$ -complete in Sections 4 and 5, respectively. In Section 6, we propose an alternative to fairness for the restricted chase that simplifies its universal termination problem. We conclude with a brief discussion about future work in Section 7.

## 2 Preliminaries

*First-Order Logic.* Consider pairwise disjoint, countably infinite sets of *predicates*  $\text{Preds}$ , *variables*  $\text{Vars}$ , *constants*  $\text{Cons}$ , and *nulls*  $\text{Nulls}$ . Every predicate has an *arity* through  $\text{ar} : \text{Preds} \rightarrow \mathbb{N} \cup \{0\}$ . Elements in  $\text{Vars} \cup \text{Cons} \cup \text{Nulls}$  are called *terms*. An *atom* is an expression of the form  $P(\vec{t})$  where  $\vec{t}$  a list of terms and  $P$  is a  $|\vec{t}|$ -ary predicate. A *fact* is a variable-free atom. An (*existential*) *rule*  $R$  is a closed first-order formula of the form  $\forall \vec{x}, \vec{y}. B[\vec{x}, \vec{y}] \rightarrow \exists \vec{z}. H[\vec{y}, \vec{z}]$  where  $\vec{x}$ ,  $\vec{y}$ , and  $\vec{z}$  are pairwise disjoint lists of variables;  $B$  and  $H$  are null-free conjunctions of atoms featuring exactly the variables in  $\vec{x}$ ,  $\vec{y}$  and  $\vec{y}$ ,  $\vec{z}$ , respectively; and  $H$  is non-empty. We write  $\text{body}(R)$  and  $\text{head}(R)$  to denote  $B$  and  $H$ , respectively; and refer to the list  $\vec{y}$  of variables as the *frontier* of  $R$ . We omit universal quantifiers for brevity. A *database* is a finite fact set without nulls. A *knowledge base (KB)* is a pair  $\langle \Sigma, D \rangle$  consisting of a finite rule set  $\Sigma$  and a database  $D$ .

*The Chase.* A *substitution*  $\sigma$  is a partial mapping from variables to constants or nulls. For an (arbitrary) expression  $\varphi$ , let  $\sigma(\varphi)$  be the expression that results from  $\varphi$  by replacing all occurrences of every variable  $v$  in  $\varphi$  by  $\sigma(v)$  if the latter is defined. A *trigger* is a pair  $\langle R, \sigma \rangle$  consisting of a rule  $R$  and a substitution  $\sigma$  that is defined exactly on the universally quantified variables in  $R$ . The *support* of a trigger  $\langle R, \sigma \rangle$  is  $\text{support}(\langle R, \sigma \rangle) = \sigma(\text{body}(R))$ . A trigger  $\langle R, \sigma \rangle$  is *loaded* for a fact set  $F$  if this fact set includes its support; and *obsolete* for  $F$  if there exists a substitution  $\sigma'$  that extends  $\sigma$  to the existential variables in  $R$  such that  $\sigma'(\text{head}(R)) \subseteq F$ . The *output* of a trigger  $\langle R, \sigma \rangle$  that is not obsolete for  $F$  is  $\text{output}(\langle R, \sigma \rangle) = \sigma'(\text{head}(R))$ , where  $\sigma'$  is some substitution that extends  $\sigma$  by mapping every existential variable in  $R$  to a fresh null. A  $\Sigma$ -*trigger* is a trigger with a rule in  $\Sigma$ .

**Definition 3.** A (restricted) chase derivation for a KB  $\langle \Sigma, D \rangle$  is a possibly infinite sequence  $F_0, F_1, \dots$  of fact sets such that (1)  $F_0 = D$  and, (2) for each  $i \geq 0$ , there is some  $\Sigma$ -trigger  $\langle R, \sigma \rangle$  that is loaded and not obsolete for  $F_i$  such that  $F_{i+1} = F_i \cup \text{output}(\langle R, \sigma \rangle)$ . Such a chase derivation is a (restricted) chase sequence if, (3) for every  $\Sigma$ -trigger  $\lambda$  and every  $i \geq 0$  such that  $\lambda$  is loaded for  $F_i$ , there is some  $j \geq i$  such that  $\lambda$  is obsolete for  $F_j$ .

Condition (3) is known as *fairness*. Note that, if no appropriate trigger according to condition (2) exists for some  $i \geq 0$ , then the sequence necessarily ends at  $F_i$ . The *result* of a chase sequence  $\mathcal{F}$  is the union of all fact sets in  $\mathcal{F}$ . It is well-known that the result  $F$  of any chase sequence for a KB  $\mathcal{K} = \langle \Sigma, D \rangle$  is a *universal model* for  $\mathcal{K}$ . That is, every model of  $\mathcal{K}$  can be homomorphically embedded into  $F$ , which is also a model of this theory. Note that, if we consider infinite sequences, the result of the chase may not be a model of  $\mathcal{K}$  if we disregard fairness.

A chase sequence *terminates* if it is finite. A KB *existentially terminates* if it admits a terminating chase sequence; it *universally terminates* if all of its chase sequences terminate. A rule set  $\Sigma$  *existentially terminates* if every KB with  $\Sigma$  existentially terminates; it *universally terminates* if every KB with  $\Sigma$  universally terminates. The classes of knowledge bases that existentially and universally terminate are denoted by  $CTK_{\exists}^{rest}$  and  $CTK_{\forall}^{rest}$ , respectively. The classes of rule sets that existentially and universally terminate are denoted by  $CTR_{\exists}^{rest}$  and  $CTR_{\forall}^{rest}$ , respectively. We also consider similar classes for the oblivious and core chase variants, which we denoted in the obvious manner. For instance,  $CTR_{\exists}^{obl}$  is the set of all rule sets that existentially terminate for the oblivious chase.

*Turing Machines.* As per our definition, all machines reuse the same *initial* state. Moreover, machines do not write blanks and cannot access accepting or rejecting states; these are not relevant in our context because we only consider halting problems.

**Definition 4.** A (non-deterministic Turing) machine is a tuple  $\langle Q, \Gamma, \delta \rangle$  where  $Q$  is a set of states that contains the initial state  $q_0$ ,  $\Gamma$  is a tape alphabet with  $\Gamma \supseteq \{0, 1\}$  and  $B \notin \Gamma$ , and  $\delta$  is a transition function for  $Q$ . That is,  $\delta$  is a function that maps from  $Q \times \Gamma \cup \{B\}$  to  $\mathcal{P}(Q \times \Gamma \times \{\leftarrow, \rightarrow\})$ .

**Definition 5.** A configuration for a machine  $\langle Q, \Gamma, \delta \rangle$  is a tuple  $\langle n, t, p, q \rangle$  where  $n$  is a natural number;  $t : \{1, \dots, n\} \rightarrow \Gamma \cup \{B\}$  is a function such that  $t(n) = B$ , and  $t(i+1) = B$  if  $t(i) = B$  for some  $1 \leq i < n$ ;  $p$  is a number in  $\{1, \dots, n\}$ ; and  $q$  is a state in  $Q$ . The starting configuration on some word  $w_1, \dots, w_n \in \{0, 1\}^*$  is the tuple  $\langle n+1, t, 1, q_0 \rangle$  where  $t$  is the function that maps 1 to  $w_1$ , 2 to  $w_2, \dots, n$  to  $w_n$ , and  $n+1$  to  $B$ .

For a configuration  $\langle n, t, p, q \rangle$ , we use  $t$  to encode the contents of the tape at each position; moreover, we use  $p$  and  $q$  to encode the position of the head and the state of the machine, respectively. Note that elements of the tape alphabet  $\Gamma$  may not occur after a blank symbol in such a configuration.

**Definition 6.** Consider a machine  $M = \langle Q, \Gamma, \delta \rangle$  and a configuration  $\rho = \langle n, t, p, q \rangle$  with  $q \in Q$ . Then, let  $\text{Next}_M(\rho)$  be the smallest set that, for every  $\langle r, a, \leftrightarrow \rangle \in \delta(t(p), q)$  with  $\leftrightarrow = \rightarrow$  or  $p \geq 2$ , contains the configuration  $\langle n+1, t', p', r \rangle$  where:

- Let  $t'(p) = a$ , let  $t'(n+1) = B$ , and let  $t'(i) = t(i)$  for every  $1 \leq i \leq n$  with  $i \neq p$ .
- If  $\leftrightarrow = \leftarrow$ , then  $p' = p - 1$ ; otherwise,  $p' = p + 1$ .

As described above, any given machine defines a function that maps configurations to sets of configurations. An exhaustive traversal through a path in this possibly infinite tree of configurations that begins with a starting configuration yields a run:

**Definition 7.** A run of a machine  $M$  on a configuration  $\rho_1$  is a possibly infinite sequence  $S = \rho_1, \rho_2, \dots$  of configurations such that  $\rho_{i+1}$  is in  $\text{Next}_M(\rho_i)$  for every  $1 \leq i < |S|$ , and  $\text{Next}_M(\rho_{|S|}) = \emptyset$  if  $S$  is finite. A partial run of  $M$  on  $\rho_1$  is a sequence of configurations that can be extended into a run of  $M$  on  $\rho_1$ . A (partial) run of  $M$  on a word  $\vec{w}$  is a (partial) run on the starting configuration of  $\vec{w}$ .

*Computability Theory.* The arithmetical hierarchy consists of classes of formal languages  $\Sigma_i^0$  with  $i \geq 1$  where  $\Sigma_1^0$  is the class of all semi-decidable languages and  $\Sigma_{i+1}^0$  is obtained from  $\Sigma_i^0$  with a Turing jump [19]. The co-classes are denoted by  $\Pi_i^0$ . Equivalently, these classes can be viewed as the sets of natural numbers definable by first-order logic formulas with bounded quantifier alternation. That is,  $\Sigma_i^0$  is the class of sets of natural numbers definable with a formula of the form  $\exists \vec{x}_1 \forall \vec{x}_2 \dots Q_i \vec{x}_i. \phi[x, \vec{x}_1, \dots, \vec{x}_i]$  where  $\phi$  is a quantifier-free formula and  $Q_i$  is  $\exists$  if  $i$  is odd or  $\forall$  otherwise. For  $\Pi_i^0$ , the alternation starts with  $\forall$ . We also the first level of the *analytical hierarchy*; that is,  $\Sigma_1^1$  and  $\Pi_1^1$  [19]. The analytical hierarchy can analogously be defined using second-order formulae with bounded second-order quantifier alternation. In the following, we introduce complete problems for these classes that we later use in our reductions. Consider a machine  $M$  and a state  $q_r$ .

- The machine  $M$  is *non-recurring through  $q_r$*  on some word  $\vec{w}$  if every run of  $M$  on  $\vec{w}$  features  $q_r$  finitely many times.
- It is *universally non-recurring through  $q_r$*  if it is non-recurring through  $q_r$  on all words.
- It is *robust non-recurring through  $q_r$*  if every run of  $M$  on any configuration features  $q_r$  finitely many times.

We obtain  $\Pi_1^1$ -completeness of the first problem by adjusting a proof from the literature [15] and for the latter two using simple reductions that we define in Appendix A.

	KB		Rule Set	
	Sometimes	Always	Sometimes	Always
<b>Oblivious</b>	RE-complete [6]		RE-complete [10, 18]	
<b>Restricted</b>	RE-complete [6]	$\Pi_1^1$ -complete	$\Pi_2^0$ -complete [13]	$\Pi_1^1$ -complete
<b>Core</b>	RE-complete [6]		$\Pi_2^0$ -complete [13]	

Table 1. Undecidability status of the main decision problems related to chase termination; the results presented without citations refer to the main contributions of this article

### 3 Related Work

*Novel Notation.* The notation introduced in Section 2 to refer to classes of terminating KBs and rule sets differs from previous literature [13]; for instance, we write  $\text{CTR}_{\forall}^{\text{rest}}$  instead of  $\text{CT}_{\forall\forall}^{\text{rest}}$ . Moreover, given some database  $D$ , we do not consider a class such as  $\text{CT}_{D\forall}^{\text{rest}}$  [13], which contains a rule set  $\Sigma$  if  $\langle \Sigma, D \rangle$  universally terminates for the restricted chase. For our purposes, it is clearer to consider a single class of terminating KBs (such as  $\text{CTK}_{\forall}^{\text{rest}}$ ) instead of one class of terminating rule sets for every possible database because of the following result.

**Proposition 8.** *For a database  $D'$ , a quantifier  $Q \in \{\forall, \exists\}$ , and a chase variant  $\text{var} \in \{\text{obl}, \text{rest}, \text{core}\}$ ; there is a many-one reduction from  $\text{CTK}_Q^{\text{var}}$  to  $\text{CT}_{D'Q}^{\text{var}}$  and vice-versa.*

**PROOF.** There is a many-one reduction  $\text{CT}_{D'Q}^{\text{var}}$  to  $\text{CTK}_Q^{\text{var}}$  since, for a rule set  $\Sigma$ , we have that  $\Sigma \in \text{CT}_{D'Q}^{\text{var}}$  if and only if  $\langle \Sigma, D' \rangle \in \text{CTK}_Q^{\text{var}}$ . To show that there is a many-one reduction in the other direction we describe a computable function that maps a KB  $\mathcal{K} = \langle \Sigma, D \rangle$  into the rule set  $\Sigma'$  such that  $\mathcal{K} \in \text{CTK}_Q^{\text{var}}$  if and only if  $\Sigma' \in \text{CT}_{D'Q}^{\text{var}}$ . Namely, let  $\Sigma'$  be the rule set that results from applying the following modifications to  $\Sigma$ : (i) replace all occurrences of every predicate  $P$  with a fresh predicate  $P'$ , (ii) add the conjunction  $\bigwedge_{P(\vec{c}) \in D} P'(\vec{c})$  to the body of every rule, and (iii) add the rule  $\rightarrow \bigwedge_{P(\vec{c}) \in D} P'(\vec{c})$ . The reduction is correct because one can easily establish a one-to-one correspondence between the sequences of  $\mathcal{K}$  and those of  $\langle \Sigma', D' \rangle$  once we ignore the single trigger with  $\rightarrow \bigwedge_{P(\vec{c}) \in D} P'(\vec{c})$  at the beginning of every sequence of the latter KB. Note that the sets of facts produced at subsequent steps of these corresponding sequences are identical modulo replacement of all occurrences of every predicate  $P$  by  $P'$ .  $\square$

*Chase Termination in the General Case.* All decision problems related to chase termination are undecidable. However, these are complete for different classes within the arithmetical and analytical hierarchies, as summarised in Table 1. In the following paragraphs, we discuss some simple proofs as well as the relevant references to understand all of the results in this table.

One can readily show via induction that, if a fact occurs in some oblivious chase sequence of some KB, then it also occurs in all oblivious chase sequences of this KB. Hence, all such chase sequences of a KB yield the same result, and thus we conclude that  $\text{CTK}_{\exists}^{\text{obl}} = \text{CTK}_{\forall}^{\text{obl}}$  and  $\text{CTR}_{\exists}^{\text{obl}} = \text{CTR}_{\forall}^{\text{obl}}$ .

Deutsch et al. proved that, if a KB admits a finite universal model, then all of its core chase sequences yield precisely this model and thus all of these sequences are finite; see Theorem 7 in [6]. Regardless of the variant, all terminating chase sequences yield a (not necessarily minimal) finite universal model; hence, if a KB does not admit a finite universal model, then it does not admit any finite chase sequence. Therefore, we have that either all core chase sequences of a KB are finite or all of them are infinite. Because of this, we conclude that  $\text{CTK}_{\exists}^{\text{core}} = \text{CTK}_{\forall}^{\text{core}}$  and  $\text{CTR}_{\exists}^{\text{core}} = \text{CTR}_{\forall}^{\text{core}}$ .

To understand why  $\text{CTK}_{\exists}^{\text{obl}}$  (resp.  $\text{CTK}_{\exists}^{\text{rest}}$  or  $\text{CTK}_{\exists}^{\text{core}}$ ) is recursively enumerable (RE), consider the following procedure: given some input KB, compute all of its oblivious (resp. restricted or core)

chase sequences in parallel and accept as soon as you find a finite one. Deutsch et al. proved that  $\text{CTK}_{\exists}^{\text{rest}}$  is RE-hard. More precisely, they defined a reduction that takes a machine  $M$  as input and produces a KB  $\mathcal{K}$  as output such that  $M$  halts on the empty word if and only  $\mathcal{K}$  is in  $\text{CTK}_{\exists}^{\text{rest}}$ ; see Theorem 1 in [6]. This reduction works because all restricted chase sequences of  $\mathcal{K}$  yield the same result, which encodes the computation of  $M$  on the empty word with a grid-like structure (as we ourselves do in later sections). One can use the same reduction to show that  $\text{CTK}_{\exists}^{\text{obl}}$  is also RE-hard.

Deutsch et al. also proved that  $\text{CTK}_{\exists}^{\text{core}}$  is RE-hard. More precisely, they showed that checking if a KB admits a universal model is undecidable; see Theorem 6 in [6]. Moreover, they proved that the core chase is a procedure that halts and yields a finite universal model for an input KB if this theory admits one; see Theorem 7 of the same paper. Therefore, the core chase can be applied as a semi-decision procedure for checking if a KB admits a finite universal model.

In Section 4, we argue that  $\text{CTK}_{\forall}^{\text{rest}}$  is  $\Pi_1^1$ -complete. This contradicts Theorem 5.1 in [13], which states that  $\text{CTK}_{\forall}^{\text{rest}}$  is RE-complete. Specifically, it is claimed that this theorem follows from results in [6], but the authors of that paper only demonstrate that  $\text{CTK}_{\forall}^{\text{rest}}$  is undecidable without proving that it is in RE. Before our completeness result, the tightest lower bound was proven by Carral et al., who proved that this class is  $\Pi_2^0$ -hard; see Proposition 42 in [5].

Marnette proved that  $\text{CTR}_{\exists}^{\text{obl}}$  is in RE. More precisely, he showed that a rule set  $\Sigma$  is in  $\text{CTR}_{\exists}^{\text{obl}}$  if and only if the KB  $\langle \Sigma, D_{\Sigma}^{\star} \rangle$  is in  $\text{CTK}_{\exists}^{\text{obl}}$  where  $D_{\Sigma}^{\star} = \{P(\star, \dots, \star) \mid P \in \text{Preds}(\Sigma)\}$  is the *critical instance* and  $\star$  is a special fresh constant; see Theorem 2 in [18]. This result follows because one can show that, for any database  $D$ , the (only) result of the oblivious chase of  $\langle \Sigma, D_{\Sigma}^{\star} \rangle$  includes the (only) result of the oblivious chase of  $\langle \Sigma, D \rangle$  if we replace all syntactic occurrences of constants in the latter with  $\star$ . Since  $\text{CTK}_{\exists}^{\text{obl}}$  is in RE, we conclude that  $\text{CTR}_{\exists}^{\text{obl}}$  is also in this class.

Gogacz and Marcinkowski proved that  $\text{CTR}_{\exists}^{\text{obl}}$  is RE-hard. More precisely, they presented a reduction that takes a 3-counter machine  $M$  as input and produces a rule set  $\Sigma$  such that  $M$  halts on  $\varepsilon$  if and only if  $\langle \Sigma, D_{\Sigma}^{\star} \rangle$  is in  $\text{CTK}_{\exists}^{\text{obl}}$ ; see Lemma 6 in [10].<sup>2</sup> Hence,  $M$  halts on the  $\varepsilon$  if and only if  $\Sigma$  is in  $\text{CTR}_{\exists}^{\text{obl}}$  by Theorem 2 in [18]. Furthermore, Bednarczyk et al. showed that this hardness result holds even when we consider single-head binary rule sets; see Theorem 1.1 in [1].

To understand why  $\text{CTR}_{\exists}^{\text{rest}}$  is in  $\Pi_2^0$ , consider the following semi-decision procedure that can access an oracle that decides the RE-complete class  $\text{CTK}_{\exists}^{\text{rest}}$ : given some input rule set  $\Sigma$ ; iterate through every database  $D$ , use the oracle to decide if  $\langle \Sigma, D \rangle$  is in  $\text{CTK}_{\exists}^{\text{rest}}$ , and accept if this is not the case. Consider an analogous procedure to understand why  $\text{CTR}_{\exists}^{\text{core}}$  is in  $\Pi_2^0$ .

Grahne and Onet proved that  $\text{CTR}_{\exists}^{\text{rest}}$  is  $\Pi_2^0$ -hard. To show this, they defined two reductions that take a word rewriting system  $R$  and a word  $\vec{w}$  as input, and produce a rule set  $\Sigma_R$  and a database  $D_{\vec{w}}$ , respectively. Then, they proved that  $R$  terminates on  $\vec{w}$  if and only if the KB  $\langle \Sigma_R, D_{\vec{w}} \rangle$  is in  $\text{CTK}_{\exists}^{\text{rest}}$ ; this claim holds because  $\langle \Sigma_R, D_{\vec{w}} \rangle$  only admits a single restricted chase result, which encodes all branches of computation of  $R$  on  $\vec{w}$  in an implicit tree-like structure. Therefore,  $R$  is uniformly terminating if  $\Sigma_R$  is in  $\text{CTR}_{\exists}^{\text{rest}}$ . To ensure that  $\Sigma_R$  is in  $\text{CTR}_{\exists}^{\text{rest}}$  if  $R$  is uniformly terminating, Grahne and Onet make use of “flooding”, a technique used in earlier work dealing with datalog boundedness [7]. For a comprehensive presentation of this technique and its applications, see Section 2 of [11]. Using the very same reduction, Grahne and Onet also proved that  $\text{CTR}_{\exists}^{\text{core}}$  is  $\Pi_2^0$ -hard.

In Section 5, we show that  $\text{CTR}_{\forall}^{\text{rest}}$  is  $\Pi_1^1$ -complete. This contradicts Theorem 5.16 in [13], where it is stated that this class is  $\Pi_2^0$ -complete. The error in the upper-bound of this theorem arose from the assumption that  $\text{CTK}_{\forall}^{\text{rest}}$  is in RE, which, as previously discussed, is not the case. Regarding the lower bound, they consider an extended version of this class of rule sets where they allow the inclusion of a single “denial constraint”; that is, an implication with an empty head that halts the

<sup>2</sup>We do not think that it is possible to intuitively explain this reduction in a couple of lines. Go read this paper, it’s worth it!

chase if the body is satisfied during the computation of a chase sequence. They prove that the always restricted halting problem for rule sets is  $\Pi_2^0$ -hard if one such constraint is allowed. Our results imply that we do not need to consider such an extension to obtain a higher lower bound.

*Chase Termination of Syntactic Fragments.* Undeterred by the undecidability results discussed above, researchers have proven we can decide chase termination if we consider syntactic fragments of existential rules for which query entailment is decidable [2, 3, 12, 17]. Another way of checking termination in practice is to develop acyclicity and cyclicity notions; that is, sufficient conditions for termination and non-termination of the chase. Indeed, experiments show that we can determine chase termination for a large proportion of real-world rule sets with these checks [4, 8, 9, 14].

#### 4 Knowledge base termination

**Theorem 9.** *The class  $\text{CTK}_{\forall}^{\text{rest}}$  is  $\Pi_1^1$ -complete.*

The theorem immediately follows from the upcoming Lemma 12 and Lemma 13.

For the membership part, we define a non-deterministic Turing machine that loops on  $q_r$  if and only if there is a non-terminating chase sequence for a given rule set.

**Definition 10.** *Consider a rule set  $\Sigma$ . For a fact set  $F$ , let  $\text{active}(F)$  be the set of all triggers with a rule in  $\Sigma$  that are loaded and not obsolete for  $F$ . Let  $\mathcal{M}_{\Sigma}$  be a non-deterministic Turing machine with start state  $q_0$  and a designated state  $q_r$  that executes the following procedure.*

- (1) *Check if the input tape contains a valid encoding of a database. If not, halt.*
- (2) *Initialize two counters  $i = j = 0$  and a set of facts  $F_0$  containing exactly the encoded database.*
- (3) *If  $\text{active}(F_i)$  is empty, halt.*
- (4) *Non-deterministically pick a trigger  $\langle R, \sigma \rangle$  from  $\text{active}(F_i)$  and let  $F_{i+1} = F_i \cup \sigma'(\text{head}(R))$  where  $\sigma'$  extends  $\sigma$  by mapping existential variables in  $R$  to fresh nulls (not occurring in  $F_i$ ).*
- (5) *If all triggers in  $\text{active}(F_j)$  are obsolete for  $F_i$ , then increment  $j$  and visit  $q_r$  once.*
- (6) *Increment  $i$  and go to 3.*

**Lemma 11.** *For every database  $D$  and rule set  $\Sigma$ , there is a run of  $\mathcal{M}_{\Sigma}$  on the encoding of  $D$  that visits  $q_r$  infinitely often if and only if there is a non-terminating chase sequence for  $\langle \Sigma, D \rangle$ .*

**PROOF.** Assume that there is a run of  $\mathcal{M}_{\Sigma}$  on the encoding of  $D$  that visits  $q_r$  infinitely many times. Then, the sequence  $F_0, F_1, \dots$  constructed by  $\mathcal{M}_{\Sigma}$  is an infinite restricted chase derivation for  $\langle \Sigma, D \rangle$  by construction. Since  $q_r$  is visited infinitely many times,  $j$  grows towards infinity. Therefore, every trigger that is loaded for some  $F_j$  with  $j \geq 0$  is obsolete for some  $i \geq j$ ; which is exactly *fairness*. Hence, the infinite derivation is a proper chase sequence.

Assume that there is an infinite chase sequence  $F_0, F_1, \dots$  for  $\langle \Sigma, D \rangle$ . By definition, for each  $i \geq 0$ , there is a trigger  $\lambda \in \text{active}(F_i)$  that yields  $F_{i+1}$ . Hence, there is a run of  $\mathcal{M}_{\Sigma}$  that non-deterministically picks these triggers. Because of fairness, for every trigger  $\lambda$  in  $\text{active}(F_j)$  with  $j \geq 0$ , there is  $i \geq j$  such that  $\lambda$  is obsolete for  $F_i$ . Hence, the run of  $\mathcal{M}_{\Sigma}$  visits  $q_r$  infinitely often.  $\square$

**Lemma 12.** *Deciding membership in  $\text{CTK}_{\forall}^{\text{rest}}$  is in  $\Pi_1^1$ .*

**PROOF.** We show a reduction to non-recurrence through  $q_r$  on the empty word. For a given rule set  $\Sigma$ , let  $\mathcal{M}_{\Sigma}^D$  be a non-deterministic Turing machine that results from  $\mathcal{M}_{\Sigma}$  by adding an initial step that replaces the initial tape content by an encoding of  $D$ . Then, by Lemma 11,  $\Sigma$  is in  $\text{CTK}_{\forall}^{\text{rest}}$  if and only if no run of  $\mathcal{M}_{\Sigma}^D$  on the empty input visits  $q_r$  infinitely many times.  $\square$

**Lemma 13.** *The class  $\text{CTK}_{\forall}^{\text{rest}}$  is  $\Pi_1^1$ -hard.*

To prove hardness, we reduce non-recurrence through  $q_r$  on the empty word to knowledge base termination. In other words, to a Turing machine  $M$ , we will associate a database  $D_\epsilon$  and a rule set  $\Sigma_M$  such that there exists a run of  $M$  on the empty word reaching  $q_r$  infinitely often if and only if the restricted chase of  $\Sigma_M$  on  $D_\epsilon$  does not halt.

A perhaps surprising feature of this reduction is that the restricted chase must halt for rule sets generated from Turing machines that do not halt on the empty word, as long as they reach  $q_r$  only finitely often. As we cannot get any computable bound on the number of steps required to reach  $q_r$ , we must simulate any finite run of the Turing machine in a terminating way. This calls for the use of *emergency brakes* as presented in the introduction. We “stack” such brakes, each one being responsible to prevent the non-termination for runs that do not go through  $q_r$ .

*Schema.* We will make use of the following predicates. Note that the last position usually holds an emergency brake. We introduce: For each letter  $a$  in the Turing machine alphabet or equal to the blank  $B$ , a binary predicate  $a$ . For each state  $q$  of the Turing machine, a binary predicate  $q$ . Two ternary predicates  $F$  and  $R$ , that encode the successor relation for time and for cells. Two binary predicates  $C_L$  and  $C_R$ , used to copy tapes content. A unary predicate  $Real$  and a binary predicate  $NextBr$ , used for the machinery of emergency brakes. Two unary predicates  $Brake$  and  $End$  to identify terms used as emergency brakes and the last element of a configuration, respectively.

Each time a new term is created during the chase, we link it in a specific way to the relevant brake. To simplify the subsequent presentation, we denote by  $brSet(x, w)$  the set of atoms  $\{F(x, w, w), R(x, w, w), Real(x), Brake(w)\}$ . The remainder of this section is devoted to the reduction from the “non-recurrence through  $q_r$ ” problem to knowledge base restricted chase termination. We first present the reduction, and then focus on the main ideas required to show correctness.

*The Reduction.* Each configuration  $\rho$  of a Turing machine is encoded by a database as follows.

**Definition 14.** *The database  $D_\rho$  encoding a configuration  $\rho = \langle n, t, p, q \rangle$  is*

$$D_\rho = \{R(c_i, c_{i+1}, w_1), a_i(c_i, w_1) \mid 1 \leq i \leq n; a_i = t(i)\} \cup \{q(c_p, w_1), B(c_{n+1}, w_1), End(c_{n+1}, w_1)\} \cup \bigcup_{1 \leq i \leq n+1} brSet(c_i, w_1)$$

For a word  $w$ , we denote by  $D_w$  the database  $D_{\rho_w}$ , where  $\rho_w$  is the initial configuration of  $M$  on  $w$ .

Given a Turing machine  $M$  with states  $Q$  and tape alphabet  $\Gamma$ , we build  $\Sigma_M$  composed of the following rules. We first have a set of rules required for setting up emergency brakes.

$$\begin{aligned} Brake(w) \rightarrow \bigwedge_{a \in \Gamma \cup \{B\}} a(w, w), \bigwedge_{q \in Q} q(w, w), & F(w, w, w), R(w, w, w), \\ & C_L(w, w), C_R(w, w), Real(w), nextBr(w, w) \quad (R_{Brake}) \\ brSet(x, w), nextBr(w, w') \rightarrow brSet(x, w') & \quad (R_{nextBr}) \end{aligned}$$

The next four rules are responsible of simulating the moves of the head of the Turing machine. The first two rules deal with the case where the machine is not in  $q_r$ , and the head moves to the right (resp. to the left). The important feature of these rules is the presence in both the body and the head of the same brake  $w$ .

For all  $q \neq q_r, q' \in Q$  and  $a, b, c \in \Gamma \cup \{\mathsf{B}\}$  such that  $(q', b, \rightarrow) \in \delta(q, a)$ :

$$\begin{aligned} & \mathsf{q}(x, w), \mathsf{a}(x, w), \mathsf{R}(x, y, w), \mathsf{c}(y, w), \mathsf{brSet}(x, w), \mathsf{brSet}(y, w) \\ & \rightarrow \exists x', y' \mathsf{q}'(y', w), \mathsf{c}(y', w), \mathsf{b}(x', w), \mathsf{C_L}(x', w), \mathsf{C_R}(y', w), \\ & \quad \mathsf{R}(x', y', w), \mathsf{F}(x, x', w), \mathsf{F}(y, y', w), \mathsf{brSet}(x', w), \mathsf{brSet}(y', w) \end{aligned} \quad (R_{\neg q_r}^{\rightarrow})$$

For all  $q \neq q_r, q' \in Q$  and  $a, b, c \in \Gamma \cup \{\mathsf{B}\}$  such that  $(q', b, \leftarrow) \in \delta(q, a)$ :

$$\begin{aligned} & \mathsf{q}(x, w), \mathsf{a}(x, w), \mathsf{R}(y, x, w), \mathsf{c}(y, w), \mathsf{brSet}(x, w), \mathsf{brSet}(y, w) \\ & \rightarrow \exists x', y' \mathsf{q}'(y', w), \mathsf{c}(y', w), \mathsf{b}(x', w), \mathsf{C_L}(y', w), \mathsf{C_R}(x', w), \\ & \quad \mathsf{R}(y', x', w), \mathsf{F}(x, x', w), \mathsf{F}(y, y', w), \mathsf{brSet}(x', w), \mathsf{brSet}(y', w) \end{aligned} \quad (R_{\neg q_r}^{\leftarrow})$$

The following two rules treat the case where the transition is from  $q_r$ . The only difference with the two above rules is the introduction of a new brake  $w'$  in the head of the rules. This permits non-terminating restricted chase sequences in the presence of specific runs.

For all  $q' \in Q$  and  $a, b, c \in \Gamma \cup \{\mathsf{B}\}$  such that  $(q', b, \rightarrow) \in \delta(q_r, a)$ :

$$\begin{aligned} & \mathsf{q}_r(x, w), \mathsf{R}(x, y, w), \mathsf{a}(x, w), \mathsf{c}(y, w), \mathsf{brSet}(x, w), \mathsf{brSet}(y, w) \\ & \rightarrow \exists x', y', w', \mathsf{q}'(y', w'), \mathsf{c}(y', w'), \mathsf{b}(x', w'), \mathsf{R}(x', y', w'), \\ & \quad \mathsf{F}(x, x', w'), \mathsf{F}(y, y', w'), \mathsf{C_L}(x', w'), \mathsf{C_R}(y', w'), \\ & \quad \mathsf{brSet}(x', w'), \mathsf{brSet}(y', w'), \mathsf{nextBr}(w, w') \end{aligned} \quad (R_{q_r}^{\rightarrow})$$

For all  $q' \in Q$  and  $a, b, c \in \Gamma \cup \{\mathsf{B}\}$  such that  $(q', b, \leftarrow) \in \delta(q_r, a)$ :

$$\begin{aligned} & \mathsf{q}_r(x, w), \mathsf{R}(y, x, w), \mathsf{a}(x, w), \mathsf{c}(y, w), \mathsf{brSet}(x, w), \mathsf{brSet}(y, w) \\ & \rightarrow \exists x', y', w', \mathsf{q}'(y', w'), \mathsf{c}(y', w'), \mathsf{b}(x', w'), \mathsf{R}(y', x', w'), \\ & \quad \mathsf{F}(x, x', w'), \mathsf{F}(y, y', w'), \mathsf{C_L}(y', w'), \mathsf{C_R}(x', w'), \\ & \quad \mathsf{brSet}(x', w'), \mathsf{brSet}(y', w'), \mathsf{nextBr}(w, w') \end{aligned} \quad (R_{q_r}^{\leftarrow})$$

The following rules copy the content of unchanged cells to the right and the left of the head from one configuration to the next. We instantiate one of each rule for each  $a \in \Gamma \cup \{\mathsf{B}\}$ .

$$\begin{aligned} & \mathsf{C_R}(x', w'), \mathsf{F}(x, x', w'), \mathsf{R}(x, y, w), \mathsf{a}(y, w), \mathsf{brSet}(x, w), \mathsf{brSet}(x', w'), \mathsf{brSet}(y, w) \\ & \rightarrow \exists y' \mathsf{F}(y, y', w'), \mathsf{R}(x', y', w'), \mathsf{a}(y', w'), \mathsf{C_R}(y', w'), \mathsf{brSet}(y', w') \quad (R_{C_R}) \\ & \mathsf{C_L}(x', w'), \mathsf{F}(x, x', w'), \mathsf{R}(y, x, w), \mathsf{a}(y, w), \mathsf{brSet}(x, w), \mathsf{brSet}(x', w'), \mathsf{brSet}(y, w) \\ & \rightarrow \exists y' \mathsf{F}(y, y', w'), \mathsf{R}(y', x', w'), \mathsf{a}(y', w'), \mathsf{C_L}(y', w'), \mathsf{brSet}(y', w') \quad (R_{C_L}) \end{aligned}$$

Finally, we extend the represented part of the configuration by one cell at each step, as coherent with our definition of Turing machine runs:

$$\begin{aligned} & \mathsf{C_R}(x', w'), \mathsf{F}(x, x', w'), \mathsf{End}(x, w), \mathsf{brSet}(x, w), \mathsf{brSet}(x', w') \\ & \rightarrow \exists y', \mathsf{R}(x', y', w'), \mathsf{B}(y', w'), \mathsf{End}(y', w'), \mathsf{brSet}(y', w') \quad (R_{End}) \end{aligned}$$

**Example 15.** Consider a machine  $M = \langle \{q_0, q_r\}, \{0, 1\}, \delta \rangle$  where  $\delta$  is a transition function that maps  $\langle q_0, 0 \rangle$  to  $\{\langle q_r, 1, \rightarrow \rangle\}$ ,  $\langle q_r, \mathsf{B} \rangle$  to  $\{\langle q_0, 1, \leftarrow \rangle\}$ ,  $\langle q_0, 1 \rangle$  to  $\{\langle q_r, 1, \rightarrow \rangle\}$ , and  $\langle q_r, 1 \rangle$  to  $\{\langle q_0, 1, \leftarrow \rangle\}$ ; note how the (only) run of  $M$  on the word 0 contains infinitely many configurations with the state  $q_r$ . In this representation, every label on an edge or a term represents several facts in the chase. For the sake

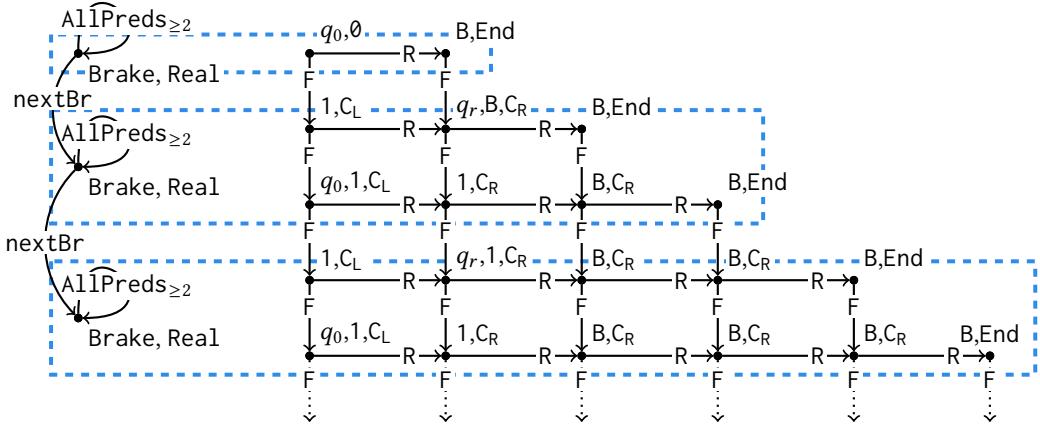


Fig. 2. An Infinite Restricted Chase Sequence of  $\langle \Sigma_M, D_0 \rangle$  where  $M$  is the machine from Example 15, and  $\text{AllPreds}_{\geq 2}$  above is a shortcuts for “ $F, R, q_0, 0, 1, B, C_R, C_L, \text{nextBr}$ ”.

of clarity, these labels can be extended with another argument, which should be some ‘Brake’ term in the same dashed or later dashed box.

*Correctness proof of the reduction.* The reduction is now fully described, and we claim that:

**Proposition 16.**  $\Sigma_M$  universally halts for the restricted chase on  $D_\rho$  if and only if there exists no run of  $M$  on  $\rho$  that goes infinitely often through  $q_r$ .

We first prove that if there exists a run of  $M$  going through  $q_r$  infinitely often, then there exists a non-terminating chase sequence. To that purpose, we identify interesting subsets of databases.

**Definition 17** (Wild Frontier of Configuration  $\rho$ ). A set of atoms  $F$  has a wild frontier of configuration  $\rho = \langle n, t, p, q \rangle$  overseen by  $w \in \text{terms}(F)$  if there exists  $x_1, \dots, x_{n+1} \in \text{terms}(F)$  such that:

- $\text{Real}(w) \notin F$ ;
- $\{R(x_i, x_{i+1}, w), a_i(x_i, w)\} \subseteq F$  for all  $i \in \{1, \dots, n\}$ ,  $a_i = t(i)$ ;
- $q(x_p, w), \text{End}(x_{n+1}, w), B(x_{n+1}, w) \in F$ ;
- $\text{brSet}(x_i, w) \in F$  for all  $i \in \{1, \dots, n+1\}$ ;
- any other atom of  $F$  having  $x_i$  as first argument has  $w$  as second.

A wild frontier has three important features (i) it contains the necessary atoms to simulate the run of a Turing machine on that configuration; (ii) it is correctly connected to a (not yet real) brake  $w$ ; (iii) it does not contain atoms preventing the above run to be simulated through a restricted derivation. By comparing Definition 14 and Definition 17, it is clear that  $D_\epsilon$  has a wild frontier of the configuration of  $M$  on the empty word, overseen by  $w_1$ . The construction of an infinite restricted derivation is made by inductively using the following key proposition.

**Proposition 18.** If  $F$  has a wild frontier of  $\rho$  overseen by  $w$ , and  $\rho'$  is reachable in one step by a transition of  $M$ , then there exists a restricted derivation  $\mathcal{D}_{\rho \rightarrow \rho'} = F, \dots, F'$  such that  $F'$  has a wild frontier of  $\rho'$  overseen by  $w'$ , where  $w' \neq w$  is a fresh existential if  $\rho$  is in  $q_r$ , and  $w' = w$  otherwise.

Concatenating the infinite sequence of derivations built in Proposition 18 does not however provide a fair sequence of derivations, because of Rules  $R_{\text{Brake}}$ ,  $R_{\text{nextBr}}$  and of the non-determinism

of  $M$ . Fairness is enforced by applying  $R_{\text{Brake}}$  and  $R_{\text{nextBr}}$  “late enough” to ensure that none of the triggers involved in the proof of Proposition 18 are made obsolete. This is possible because the run of  $M$  going infinitely many often through  $q_r$ , infinitely many brakes are created. Details are provided in the appendix.

**Lemma 19.** *Let  $(\rho_i)_{i \in \mathbb{N}}$  be a run of  $M$  on the empty word that visits  $q_r$  infinitely often. There exists an infinite restricted chase sequence for  $\langle \Sigma_M, D_\varepsilon \rangle$ .*

To show the converse, we fix an infinite restricted chase sequence  $\mathcal{D}$  as  $(F_i)_{i \in \mathbb{N}}$ , where  $F_0 = D_\varepsilon$ . We build from  $\mathcal{D}$  an infinite run that visits  $q_r$  infinitely often by identifying a substructure of the chase, consisting of *state atoms*. We then prove that a run can be built from these states atoms (and other elements of the chase), which fulfills the required conditions.

**Definition 20.** *A state atom of  $F$  is an atom of  $F$  of the form  $q(x, w)$  where  $q \in Q$  and  $x$  is not a brake. A state atom  $A$  precedes  $A'$  if there is a trigger  $t$  such that  $A \in \text{support}(t)$  and  $A' \in \text{output}(t)$ . In this case, we write  $A \prec A'$ .*

It is worth noticing that in the chase of  $\langle \Sigma_M, D_\varepsilon \rangle$ , state atoms are organised as a tree structure rooted in the unique state atom belonging to  $D_\varepsilon$ , and such that  $A$  is a parent of  $A'$  if and only if  $A \prec A'$ . Intuitively, we can assign with each of these state atoms a configuration such that the configuration associated with  $A'$  is reachable in one transition of  $M$  from the configuration associated with its parent  $A$ . The key property is that in an infinite restricted chase, there exists an infinite sequence  $(A_n)_{n \in \mathbb{N}}$  with good properties.

**Lemma 21.** *For all databases  $D$ , and all infinite chase sequences from  $\langle \Sigma_M, D \rangle$  with result  $F$ , there is an infinite sequence  $(A_n)_{n \in \mathbb{N}}$  of state atoms of  $F$  such that:*

- $A_0 \in D$ ;
- $A_n \prec A_{n+1}$  for all  $n \in \mathbb{N}$ ;
- for infinitely many  $i \in \mathbb{N}$ ,  $A_i$  is of the shape  $q_r(t_i, w_i)$ .

**PROOF SKETCH.** Since the rules that introduce state atoms (Rules  $R_{q_r}^\leftarrow$ ,  $R_{q_r}^\rightarrow$ ,  $R_{\neg q_r}^\leftarrow$  and  $R_{\neg q_r}^\rightarrow$ ) feature a state atom in their body,  $\prec$  defines a forest structure over state atoms, where the root of each tree is an atom of the database. There is thus a finite amount of trees. We can prove by induction that there is a finite amount of atoms that feature a given brake. Thus, there is an infinite amount of brakes in  $F$ . Then, since the rules that introduce new brakes (Rules  $R_{q_r}^\rightarrow$  and  $R_{q_r}^\leftarrow$ ) introduce a state atom too, there is an infinite number of state atoms. Thus, one of the trees must be infinite, and since branching can be proven to be finite, there must be an infinite branch by König’s lemma.  $\square$

**Lemma 22.** *For every configuration  $\rho$ , if the restricted chase does not terminate on  $\langle \Sigma_M, D_\rho \rangle$  then there exists a run of  $M$  on  $\rho$  which visits  $q_r$  infinitely many times.*

Lemmas 19 and 22 directly imply Proposition 16, and hence the correctness of the reduction.

## 5 Rule set termination

**Theorem 23.**  $\text{CTR}_\vee^{\text{rest}}$  is  $\Pi_1^1$ -complete.

The theorem immediately follows from the upcoming Lemma 24 and Lemma 25.

**Lemma 24.** *Deciding membership in  $\text{CTR}_\vee^{\text{rest}}$  is in  $\Pi_1^1$ .*

**PROOF.** We reduce to universal non-recurrence through  $q_r$ . More precisely, we show that a rule set  $\Sigma$  is in  $\text{CTR}_\vee^{\text{rest}}$  if and only if  $\mathcal{M}_\Sigma$  from Definition 10 is universally non-recurring through  $q_r$ .

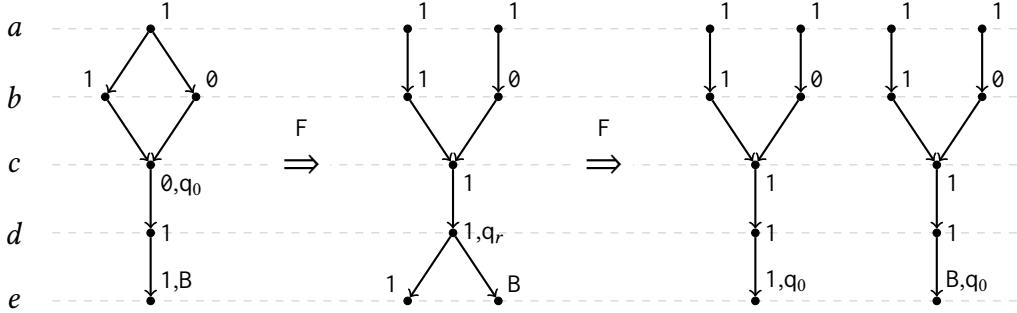


Fig. 3. First three steps of the restricted chase from  $\langle \mathcal{R}_M, D \rangle$  as defined in Example 26. The predicate  $F$  and the brakes are not represented for the sake of readability, but terms are connected through the future predicate to an element on the same line at the previous step. Unlabeled arrows represent  $R$ -atoms.

If  $\Sigma$  is in  $\text{CTR}_{\forall}^{\text{rest}}$ , then  $\langle \Sigma, D \rangle$  is in  $\text{CTK}_{\forall}^{\text{rest}}$  for each  $D$ . Hence, by Lemma 11,  $\mathcal{M}_\Sigma$  is non-recurring on every input that is the encoding of some database  $D$ . On inputs that are not encodings of databases,  $\mathcal{M}_\Sigma$  halts immediately by Definition 10. Therefore,  $\mathcal{M}_\Sigma$  is universally non-recurring.

If  $\mathcal{M}_\Sigma$  is universally non-recurring through  $q_r$ , then, in particular,  $\mathcal{M}_\Sigma$  is non-recurring through  $q_r$  on every input that is the encoding of a database. Hence, by Lemma 11, each restricted chase sequence for each knowledge base with  $\Sigma$  is finite. Therefore,  $\Sigma$  is in  $\text{CTR}_\forall^{\text{rest}}$ .  $\square$

**Lemma 25.**  $\text{CTR}_{\forall}^{\text{rest}}$  membership is  $\Pi_1^1$ -hard.

The rest of the section is dedicated to proving this lemma, by reducing robust non-recurrence through  $q_r$  to rule set termination. In fact, the reduction is very similar to the one we use for knowledge base termination: to a machine  $M$ , we associate the rule set  $\Sigma_M$ , which will belong to  $\text{CTR}_\vee^{\text{rest}}$  if and only if  $M$  is robust non-recurring through  $q_r$ . The direct implication follows from Lemma 22 by contrapositive: if a Turing machine  $M$  is *not* robust non-recurring through  $q_r$ , then there is a configuration  $\rho$  such that  $M$  visits  $q_r$  infinitely many times from  $\rho$ . Then, by Lemma 22, the restricted chase does not terminate on  $\langle \Sigma_M, D_\rho \rangle$ , and thus  $\Sigma_M \notin \text{CTR}_\vee^{\text{rest}}$ . The other direction requires more work. Consider a Turing machine  $M$ , and assume that there is some database  $D$  such that the restricted chase does not terminate on  $\langle \Sigma_M, D \rangle$ . We then show that  $M$  is not robust non-recurring through  $q_r$ .

Since the restricted chase does not terminate on  $\langle \Sigma_M, D \rangle$ , there is an infinite chase sequence from this knowledge base. We use  $F$  to denote its result. As in Section 4, by Lemma 21,  $F$  contains an infinite sequence of state atoms  $\mathcal{A} = (A_n)_{n \in \mathbb{N}}$  such that  $A_0 \in D$ ,  $A_n \prec A_{n+1}$  for all  $n \in \mathbb{N}$ , and there are infinitely many integers  $i$  such that  $A_i$  is a  $q_r$ -atom.

In the knowledge base case, we had control over the database as part of the knowledge base, which meant that we could start from a “well-formed” database (in the sense that it encodes a single start configuration). This allowed us to extract the *unique* configuration associated with a state atom. However, in the rule set case, the database  $D$  leading to non-termination is arbitrary and can contain any kind of structure, as highlighted by the following example.

**Example 26.** Consider a Turing machine  $M$  that moves to the right in every step, writing 1 regardless of the symbol it reads. It alternates between its start state  $q_0$  and the designated state  $q_r$ . Now, consider the database depicted on the left side of Fig. 3, which contains the atoms  $R(a, b_1, w)$ ,  $R(a, b_2, w)$ ,  $R(b_1, c, w)$ ,  $R(b_2, c, w)$ ,  $R(c, d, w)$ ,  $R(d, e, w)$ ,  $q_0(c, w)$ ,  $1(a, w)$ ,  $1(b_1, w)$ ,  $0(b_2, w)$ ,  $0(c, w)$ ,  $1(d, w)$ ,  $1(e, w)$ ,  $B(e, w)$ , and  $\text{brSet}(x, w)$  for all  $x \in \{a, b_1, b_2, c, d, e\}$ . This database represents four different configurations,

each with a tape of size 5, the start state  $q_0$ , and the head positioned at the third cell. These configurations correspond to tapes with contents 11011, 10011, 1101B, and 1001B.

As the simulation progresses, these configurations evolve simultaneously, creating new structures shown in the middle and right of Fig. 3. Notice how term  $e$  has two successors through the  $F$  predicate, one for each symbol atom it belongs to. Furthermore, when the head encounters a branching structure, it splits into two, as observed in the third step of the simulation. In a sense, if the machine simulation is able to perform steps on the database at all, then it will gradually “heal” the structure step by step towards proper encodings of machine configurations.

As highlighted by this example, the structure of the set of atoms connected to a state atom not present in the database is specific: it is the union of two trees rooted in the state atom. The first has arrows going towards the state atom, and the second one has arrows going away from the state atom. In fact, this structure represent the set of paths in the initial database (after the appropriate number of steps of simulation), which we coin a bow tie, due to its shape.

**Definition 27.** The inverse  $E^-$  of a binary relation  $E$  is the relation defined by  $(x, y) \in E^-$  if and only if  $(y, x) \in E$ . In a directed graph  $G = (V, E)$  we denote with  $V_x^{-y}$  the connected component<sup>3</sup> of  $x$  in the subgraph induced by  $V \setminus \{y\}$  on  $G$ , for any two vertices  $x$  and  $y$ . A bow tie is a graph  $(V, E)$  with two distinguished vertices  $x$  and  $y$  that has the following properties:

- (1)  $(x, y) \in E$ ;
- (2) The subgraph induced by  $V_x^{-y}$  on  $(V, E^-)$  is a directed tree rooted in  $x$ ;
- (3) The subgraph induced by  $V_y^{-x}$  on  $(V, E)$  is a directed tree rooted in  $y$ ;
- (4) The sets  $V_x^{-y}$  and  $V_y^{-x}$  form a partition of  $V$ ; that is, they are disjoint and  $V = V_x^{-y} \cup V_y^{-x}$ .

The edge  $(x, y)$  is called the center of the bow tie, and the sets  $V_x^{-y}$  and  $V_y^{-x}$  are called the left and right parts of the bow tie, respectively.

In the following, we denote with  $\text{semterms}(F)$  (for semantically meaningful terms) the set of all the terms in  $F$ , except the brakes (which appear in the last position of atoms). We also define  $E_R$  as the relation over  $\text{semterms}(F)$  such that  $(x, y) \in E_R$  if and only if there is  $w$  such that  $R(x, y, w) \in F$ . For all state atoms  $A = q(x, w)$  generated during the chase, we denote the connected component of  $x$  in the graph  $(\text{semterms}(F), E_R)$  with  $\text{bowtie}(A)$ . The following lemma explains how this bow tie structure is generated at each step.

**Lemma 28.** For all database  $D$ , and every  $F$  result of a chase sequence for  $\langle \Sigma_M, D \rangle$ , the graph  $\text{bowtie}(A)$  is a finite bow tie for all state atoms  $A \in F \setminus D$ . In addition:

- The center of the bow tie is the atom generated along with  $A$ , by rule  $R_{\neg q_r}^\leftarrow$ ,  $R_{\neg q_r}^\rightarrow$ ,  $R_{q_r}^\leftarrow$  or  $R_{q_r}^\rightarrow$ ;
- all the atoms in the left part of the bow tie are generated by rule  $R_{C_L}$ ;
- all the atoms in the right part of the bow tie are generated by rule  $R_{C_R}$ , except possibly the end of a maximal path, which may have been generated by rule  $R_{\text{End}}$ .

**PROOF SKETCH.** This proof relies on an analysis of how R-atoms are generated during the chase. All the rules that generate R-atoms (over non-brake terms) generate R-atoms containing at least one existentially quantified variable. Three cases occur:

- Rules  $R_{\neg q_r}^\leftarrow$ ,  $R_{\neg q_r}^\rightarrow$ ,  $R_{q_r}^\leftarrow$  and  $R_{q_r}^\rightarrow$  generate an R-atom  $R(u, v, w)$  where  $u$  and  $v$  are both existentially quantified.
- Rule  $R_{C_L}$  generates an R-atom  $R(u, v, w)$  where  $u$  is existentially quantified and  $v$  is a frontier variable.

<sup>3</sup>We consider here weakly connected components; a weakly connected component in a directed graph is a maximal subgraph such that there is an undirected path between any two vertices.

- Rules  $R_{C_R}$  and  $R_{End}$  generate an R-atom  $R(u, v, w)$  where  $u$  is a frontier variable and  $v$  is existentially quantified.

Thus, all connected components are generated by a rule of the first kind, and then extended to the left by a rule of the second kind, and to the right by a rule of the third kind. Since no rule can generate an atom  $R(u, v, w)$  where  $u$  and  $v$  are both frontier variable (assuming  $u$  and  $v$  are not brakes), this yields the wanted structure. Finiteness is guaranteed by the emergency brakes.  $\square$

We now have a bit of structure to work with. Let us give a bit of intuition before concluding the proof. We have considered an infinite sequence  $\mathcal{A} = (A_n)_{n \in \mathbb{N}}$  of state atoms, with  $A_0 \in D$  and  $A_n \prec A_{n+1}$  for all  $n \in \mathbb{N}$ , and we have just shown that to each state atom (not in  $D$ ) is attached a bow tie structure. As mentioned before, the bow tie  $\text{bowtie}(A_n)$  consists in a set of (non-disjoint) paths that represent configurations that can be obtained from a configuration containing  $A_0$  in the database, after  $n$  steps of simulation. In addition, Lemma 28 shows how each of these paths is constructed using a path from  $\text{bowtie}(A_{n-1})$ . We also have seen in Example 26 that a bow tie can get split. From these two facts we get that the number of configurations represented by  $\text{bowtie}(A_n)$  decreases as  $n$  grows. Since this number is an integer, and each bow tie represents at least one configuration, this sequence will be stationnary at some point  $N$ . At this point, we know that each of the configurations represented by  $\text{bowtie}(A_N)$  visits  $q_r$  infinitely many time. Thus, we pick such a configuration  $\rho$ , and we show that the restricted chase does not terminate on  $\langle \Sigma_M, D_\rho \rangle$ , which is enough to conclude the proof by Lemma 22. We then formalize this argument.

**Definition 29.** The set of configurations  $\text{configs}(A_n)$  associated to a state atom  $A_n = q(x, w) \in \mathcal{A}$ , with  $n > 0$ , is the set whose elements are the sets

$$\{A_n\} \cup \bigcup_{i \leq m} \text{brSet}(x_i, w) \cup \{P(y_1, \dots, y_k, w) \in F \mid P \in \{R, \theta, 1, B, \text{End}\} \text{ and } \forall i, y_i \in \{x_1, \dots, x_m\}\}$$

for all maximal paths  $(x_1, \dots, x_m)$  in  $\text{bowtie}(A_n)$ .

**Lemma 30.** For all  $n > 0$ ,  $\text{configs}(A_n)$  is finite, non-empty, and each of its elements homomorphically embeds into  $D_\rho$  for some configuration  $\rho$ . Also, there is an injective function  $\text{pred}_n$  from  $\text{configs}(A_{n+1})$  to  $\text{configs}(A_n)$  such that  $S \in \text{configs}(A_{n+1})$  can be generated using only atoms in  $\text{pred}_n(S)$ .

**PROOF SKETCH.** To each set  $S \in \text{configs}(A_{n+1})$  we can associate a configuration  $\rho$  and a path  $p$  in  $\text{bowtie}(A_{n+1})$ . We then define  $\text{pred}_n(S)$  as the set of atoms that was used to generate it, which is not hard: its associated configuration is an extension of a configuration that yields  $\rho$ , and its associated path is connected through the  $F$ -predicate to  $p$ . To show injectivity of  $\text{pred}_n$ , we then rely on a lemma stating that if  $F(x, z, w)$  and  $F(y, z, w)$  are both in  $F$ , then  $x = y$ .  $\square$

Since for all  $n$ , there is an injective function from  $\text{configs}(A_{n+1})$  to  $\text{configs}(A_n)$ , the sequence  $(|\text{configs}(A_n)|)_{n \in \mathbb{N}_{>0}}$  is a decreasing sequence of natural numbers, as mentioned before. Thus, there must be some  $N \in \mathbb{N}$  such that for all  $n \geq N$ ,  $|\text{configs}(A_n)| = |\text{configs}(A_N)| > 0$ . We pick  $S_0$  in  $\text{configs}(A_N)$ , and let  $\rho$  be a configuration such that  $S_0$  homomorphically embeds into  $D_\rho$ .

**Lemma 31.** The restricted chase does not terminate on  $\langle \Sigma_M, D_\rho \rangle$ .

**PROOF SKETCH.** Since for all  $n \geq N$ ,  $|\text{configs}(A_n)| = |\text{configs}(A_N)|$ ,  $\text{pred}_n$  is actually a bijection. We thus define  $S_{n+1}$  as  $\text{pred}_{N+n}^{-1}(S_n)$ . Intuitively, the sequence  $(S_n)_{n \in \mathbb{N}}$  encodes the run of  $M$  that visits  $q_r$  infinitely many times from  $\rho$ . We then construct an infinite chase sequence from  $\langle \Sigma_M, D_\rho \rangle$  such that  $S_n$  homomorphically embed in it for all  $n$ .  $\square$

By Lemma 22, this means that there is a run of  $M$  which visits  $q_r$  infinitely many times, and thus that  $M$  is not robust non-recurring through  $q_r$ , concluding the reduction.

## 6 An Alternative to Fairness to Simplify Restricted Chase Termination

All chase variants can be applied to semi-decide Boolean conjunctive query (BCQ) entailment. This is the case because, if a KB  $\mathcal{K}$  entails a BCQ  $Q$  under standard first-order semantics, then every chase sequence of  $\mathcal{K}$  features a fact set that entails  $Q$ . Consequently, it suffices to compute an (arbitrarily large) finite prefix of any (arbitrarily chosen) chase sequence of  $\mathcal{K}$  to semi-decide whether  $\mathcal{K}$  entails  $Q$ . Note that semi-decidability of BCQ entailment breaks down if we remove the fairness condition from the definition of a chase sequence. Unfortunately, this condition complicates the problem of universal termination for the restricted chase (see Theorems 9 and 23). To address this situation, we propose an alternative to fairness in the following definition that retains semi-decidability while simplifying the termination problem of the chase (see Theorem 34).

**Definition 32.** A breadth-first chase sequence for a KB  $\langle \Sigma, D \rangle$  is a chase derivation  $F_0, F_1, \dots$  such that,  $(\dagger)$  if some  $\Sigma$ -trigger  $\lambda$  is loaded for some  $F_i$ , then there is some  $j \in \{i, \dots, i+n\}$  such that  $\lambda$  is obsolete for  $F_j$  and  $n$  is the (finite) number of  $\Sigma$ -triggers that are loaded and not obsolete for  $F_i$ .

Note that, since  $(\dagger)$  implies fairness as introduced in Definition 3, every breadth-first chase sequence is also a chase sequence and we preserve semi-decidability of BCQ entailment.

**Definition 33.** Let  $\text{CTK}_{\vee}^{b1r}$  be the class of all KBs that only admit finite breadth-first chase sequences. Let  $\text{CTR}_{\vee}^{b1r}$  be the class containing a rule set if  $\text{CTK}_{\vee}^{b1r}$  contains all KBs with this rule set.

**Theorem 34.** The class  $\text{CTK}_{\vee}^{b1r}$  is in RE, and the class  $\text{CTR}_{\vee}^{b1r}$  is in  $\Pi_2^0$ .

PROOF. To show that  $\text{CTK}_{\vee}^{b1r}$  is in RE, we define a semi-decision procedure, which executes the following instructions on a given input KB  $\mathcal{K} = \langle \Sigma, D \rangle$ :

- (1) Initialise the set  $\mathcal{P}_1$  of lists of facts that contains the (unary) list  $D$ , and a counter  $i := 2$ .
- (2) Compute the set  $C_i$  of all chase derivations of length  $i$  of  $\mathcal{K}$  that can be obtained by extending a chase derivation in  $\mathcal{P}_{i-1}$  with one fact set. Intuitively,  $C_i$  includes all lists of length  $i$  that can be extended into breadth-first chase sequences for  $\mathcal{K}$ .
- (3) Compute the maximal subset  $\mathcal{P}_i$  of  $C_i$  that does not contain a chase derivation  $F_1, \dots, F_i \in C_i$  if there is some  $1 \leq k \leq i$  and some  $\Sigma$ -trigger  $\lambda$  such that  $\lambda$  is loaded for  $F_k$ , the trigger  $\lambda$  is not obsolete for  $F_i$ , and  $i - k$  is larger than the number of  $\Sigma$ -triggers that are loaded and not obsolete for  $F_k$ . Intuitively,  $\mathcal{P}_i$  filters out prefixes in  $C_i$  that already violate  $(\dagger)$ .
- (4) If  $\mathcal{P}_i$  is empty, *accept*. Otherwise, increment  $i := i + 1$  and go to 2.

If the procedure accepts, then  $\mathcal{P}_i$  is empty for some  $i$  and all breadth-first chase sequences of  $\mathcal{K}$  are of length at most  $i - 1$ . If the procedure loops, then there is an infinite chase derivation  $F_0, F_1, \dots$  of  $\mathcal{K}$  such that  $F_0, \dots, F_{i-1} \in \mathcal{P}_i$  for every  $i \geq 1$ , which is a breadth-first derivation for  $\mathcal{K}$ .

The class  $\text{CTR}_{\vee}^{b1r}$  is in  $\Pi_2^0$  because we can semi-decide if a rule set  $\Sigma$  is not in  $\text{CTR}_{\vee}^{b1r}$  using an oracle that solves  $\text{CTK}_{\vee}^{b1r}$ . We simply enumerate every database  $D$ , use the oracle to check if the KB  $\langle \Sigma, D \rangle$  is in  $\text{CTK}_{\vee}^{b1r}$ , and *accept* if this is not the case.  $\square$

The previous result holds because the condition  $(\dagger)$  is *finitely verifiable*; that is, every infinite chase derivation that does not satisfy this condition has a finite prefix that witnesses this violation. Note that fairness does not have this property since any finite prefix of any chase derivation can be extended into a (fair) chase sequence. In fact, we can readily show a version of Theorem 34 for any other alternative condition if it is finitely verifiable. For an example of one such trigger application strategy, consider the one from [20], which is a bit more complex to define than  $(\dagger)$  but nevertheless results in a very efficient implementation of the restricted chase.

## 7 Open Problems

After settling the general case regarding restricted chase termination and proposing an alternative fairness condition, there are still open challenges. Namely, what is the undecidability status of the classes  $CTK_{\forall}^{rest}$  and  $CTR_{\forall}^{rest}$  if we only consider single-head rules or only guarded (multi-head) rules? Note that, with guarded rules, it is not obvious how to simulate a Turing machine. For single-head rules, we cannot implement the emergency brake and thus our proofs do not apply. Moreover, if we only consider single-head rule sets, we can ignore fairness when determining restricted chase termination because of the “fairness theorem” [12]: a single-head KB admits an infinite (possibly unfair) chase derivation if and only if admits an infinite (fair) chase sequence. We think that answers to these problems will help to develop a better understanding for the (restricted) chase overall.

## Acknowledgments

On TU Dresden side, this work is partly supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in project number 389792660 (TRR 248, Center for Perspicuous Systems), by the Bundesministerium für Bildung und Forschung (BMBF, Federal Ministry of Education and Research) in the Center for Scalable Data Analytics and Artificial Intelligence (project SCADS25B, ScaDS.AI), and by Bundesministerium für Bildung und Forschung (BMBF, Federal Ministry of Education and Research) and Deutscher Akademischer Austauschdienst (DAAD, German Academic Exchange Service) in project 57616814 (SECAI, School of Embedded and Composite AI).

## References

- [1] Bartosz Bednarczyk, Robert Ferens, and Piotr Ostropolski-Nalewaja. 2020. All-Instances Oblivious Chase Termination is Undecidable for Single-Head Binary TGDs. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, Christian Bessiere (Ed.), ijcai.org, 1719–1725. <https://doi.org/10.24963/IJCAI.2020/238>
- [2] Marco Calautti, Georg Gottlob, and Andreas Pieris. 2015. Chase Termination for Guarded Existential Rules. In *Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, Tova Milo and Diego Calvanese (Eds.). ACM, 91–103. <https://doi.org/10.1145/2745754.2745773>
- [3] Marco Calautti and Andreas Pieris. 2021. Semi-Oblivious Chase Termination: The Sticky Case. *Theory Comput. Syst.* 65, 1 (2021), 84–121. <https://doi.org/10.1007/S00224-020-09994-5>
- [4] David Carral, Irina Dragoste, and Markus Krötzsch. 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*, Carles Sierra (Ed.), ijcai.org, 922–928. <https://doi.org/10.24963/IJCAI.2017/128>
- [5] David Carral, Lucas Larroque, Marie-Laure Mugnier, and Michaël Thomazo. 2022. Normalisations of Existential Rules: Not so Innocuous!. In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning, KR 2022, Haifa, Israel, July 31 - August 5, 2022*, Gabriele Kern-Isbner, Gerhard Lakemeyer, and Thomas Meyer (Eds.). <https://proceedings.kr.org/2022/11/>
- [6] Alin Deutsch, Alan Nash, and Jeffrey B. Remmel. 2008. The chase revisited. In *Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2008, June 9-11, 2008, Vancouver, BC, Canada*, Maurizio Lenzerini and Domenico Lembo (Eds.). ACM, 149–158. <https://doi.org/10.1145/1376916.1376938>
- [7] Haim Gaifman, Harry G. Mairson, Yehoshua Sagiv, and Moshe Y. Vardi. 1993. Undecidable Optimization Problems for Database Logic Programs. *J. ACM* 40, 3 (1993), 683–713. <https://doi.org/10.1145/174130.174142>
- [8] Lukas Gerlach and David Carral. 2023. Do Repeat Yourself: Understanding Sufficient Conditions for Restricted Chase Non-Termination. In *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning, KR 2023, Rhodes, Greece, September 2-8, 2023*, Pierre Marquis, Tran Cao Son, and Gabriele Kern-Isbner (Eds.), 301–310. <https://doi.org/10.24963/KR.2023/30>
- [9] Lukas Gerlach and David Carral. 2023. General Acyclicity and Cyclicity Notions for the Disjunctive Skolem Chase. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, Brian Williams, Yiling Chen, and Jennifer Neville (Eds.). AAAI Press, 6372–6379. <https://doi.org/10.1609/AAAI.V37I5.25784>
- [10] Tomasz Gogacz and Jerzy Marcinkowski. 2014. All-Instances Termination of Chase is Undecidable. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014*,

*Proceedings, Part II (Lecture Notes in Computer Science, Vol. 8573)*, Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias (Eds.). Springer, 293–304. [https://doi.org/10.1007/978-3-662-43951-7\\_25](https://doi.org/10.1007/978-3-662-43951-7_25)

[11] Tomasz Gogacz and Jerzy Marcinkowski. 2014. Termination of oblivious chase is undecidable. *CoRR* abs/1401.4840 (2014). arXiv:1401.4840 <http://arxiv.org/abs/1401.4840>

[12] Tomasz Gogacz, Jerzy Marcinkowski, and Andreas Pieris. 2023. Uniform Restricted Chase Termination. *SIAM J. Comput.* 52, 3 (2023), 641–683. <https://doi.org/10.1137/20M1377035>

[13] Gösta Grahne and Adrian Onet. 2018. Anatomy of the Chase. *Fundam. Informaticae* 157, 3 (2018), 221–270. <https://doi.org/10.3233/FI-2018-1627>

[14] Bernardo Cuenca Grau, Ian Horrocks, Markus Krötzsch, Clemens Kupke, Despoina Magka, Boris Motik, and Zhe Wang. 2013. Acyclicity Notions for Existential Rules and Their Application to Query Answering in Ontologies. *J. Artif. Intell. Res.* 47 (2013), 741–808. <https://doi.org/10.1613/JAIR.3949>

[15] David Harel. 1986. Effective transformations on infinite trees, with applications to high undecidability, dominoes, and fairness. *J. ACM* 33, 1 (jan 1986), 224–248. <https://doi.org/10.1145/4904.4993>

[16] Markus Krötzsch, Maximilian Marx, and Sebastian Rudolph. 2019. The Power of the Terminating Chase (Invited Talk). In *22nd International Conference on Database Theory, ICDT 2019, March 26–28, 2019, Lisbon, Portugal (LIPIcs, Vol. 127)*, Pablo Barceló and Marco Calautti (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 3:1–3:17. <https://doi.org/10.4230/LIPICS.ICDT.2019.3>

[17] Michel Leclère, Marie-Laure Mugnier, Michaël Thomazo, and Federico Ulliana. 2019. A Single Approach to Decide Chase Termination on Linear Existential Rules. In *22nd International Conference on Database Theory, ICDT 2019, March 26–28, 2019, Lisbon, Portugal (LIPIcs, Vol. 127)*, Pablo Barceló and Marco Calautti (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 18:1–18:19. <https://doi.org/10.4230/LIPICS.ICDT.2019.18>

[18] Bruno Marnette. 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 2009, June 19 – July 1, 2009, Providence, Rhode Island, USA*, Jan Paredaens and Jianwen Su (Eds.). ACM, 13–22. <https://doi.org/10.1145/1559795.1559799>

[19] Hartley Rogers, Jr. 1987. *Theory of recursive functions and effective computability (Reprint from 1967)*. MIT Press. <http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=3182>

[20] Jacopo Urbani, Markus Krötzsch, Cerial J. H. Jacobs, Irina Dragoste, and David Carral. 2018. Efficient Model Construction for Horn Logic with VLog - System Description. In *Automated Reasoning - 9th International Joint Conference, IJCAR 2018, Held as Part of the Federated Logic Conference, FloC 2018, Oxford, UK, July 14–17, 2018, Proceedings (Lecture Notes in Computer Science, Vol. 10900)*, Didier Galmiche, Stephan Schulz, and Roberto Sebastiani (Eds.). Springer, 680–688. [https://doi.org/10.1007/978-3-319-94205-6\\_44](https://doi.org/10.1007/978-3-319-94205-6_44)

Received December 2024; revised February 2025; accepted March 2025

## A $\Pi_1^1$ -complete Turing Machine Problems for Reductions

Our definition of non-recurring machines differs slightly from descriptions found in previous literature. Indeed, Harel showed that the following problem is  $\Pi_1^1$ -complete: decide if a (non-deterministic Turing) machine admits a run on the empty word that features the initial state  $q_0$  infinitely many times (see Corollary 6.2 in [15]). Our definition is slightly different since we choose a different state  $q_r$  to keep track of this infinite recurrence; note that this state may be different from the initial state. Fortunately, the choice of the initial state in the proof of Corollary 6.2 of Harel [15] is arbitrary, making it straightforward to adapt his proof to any given state. We first prove this in Section A.1, and then use this result to get  $\Pi_1^1$ -completeness for the other Turing machine problems we consider in Section A.2.

### A.1 Non-Recurrence on the Empty Word

To show that checking if a machine is non-recurring on the empty word is  $\Pi_1^1$ -complete, we adapt the proof of Corollary 6.2 in [15]. To do so, we first need to introduce some preliminary notions. A *list* is a finite sequence. The *concatenation* of two lists  $u = u_1, \dots, u_n$  and  $v = v_1, \dots, v_m$  is the list  $u \cdot v = u_1, \dots, u_n, v_1, \dots, v_m$ . A list  $u_1, \dots, u_n$  with  $n \geq 2$  is the *child* of  $u_1, \dots, u_{n-1}$ . A list  $u$  is an *ancestor* of another list  $v$ , written  $u \prec v$ , if  $u$  is a prefix of  $v$ ; that is, if  $u \cdot w = v$  for some list  $w$ .

**Definition 35.** An  $\omega$ -tree  $T$  is a set of lists of natural numbers closed under  $\prec$ . A node is an element in  $T$ ; a leaf is a node without children in  $T$ . Such a tree is computable if so is the following function:

$$\chi_T(u) = \begin{cases} 0 & \text{if } u \notin T \\ 1 & \text{if } u \in T \text{ and } u \text{ is a leaf} \\ 2 & \text{if } u \in T \text{ and } u \text{ is not a leaf} \end{cases}$$

A possibly infinite sequence of natural numbers is a branch of  $T$  if the latter contains every finite prefix of the former. Such a tree is well founded if all of its branches are finite.

In the following, we identify a computable  $\omega$ -tree  $T$  with the machine that computes the function  $\chi_T$ . Note that this is a machine that implements a function mapping lists of natural numbers to elements of  $\{0, 1, 2\}$  as indicated in Definition 35. Checking if such a machine does correspond to a well-founded tree is a  $\Pi_1^1$ -complete problem.

**Lemma 36** ([19], Theorem 16). *Checking if a computable  $\omega$ -tree is well founded is  $\Pi_1^1$ -complete.*

**Definition 37.** For a natural number  $k \geq 0$ , a  $k$ -tree  $T$  is an  $\omega$ -tree that does not contain sequences with numbers larger than  $k$ . A b-tree (b for bounded) is a  $k$ -tree for some  $k \geq 0$ . A marked b-tree is a pair  $(T, \mu)$  consisting of a b-tree  $T$  and a marking function  $\mu$ ; that is, a function from  $T$  to  $\{0, 1\}$ . A marked b-tree is computable if the following function is computable:

$$\chi_T^\mu(u) = \begin{cases} 0 & \text{if } u \notin T \\ 1 & \text{if } u \in T \text{ and } u \text{ is marked (that is, } \mu(u) = 1) \\ 2 & \text{if } u \in T \text{ and } u \text{ is not marked} \end{cases}$$

A marked b-tree is recurring if it has a branch with infinitely many marked prefixes.

As we do for computable  $\omega$ -trees, we identify a computable marked b-tree  $(T, \mu)$  with the decider that implements the function  $\chi_T^\mu$ .

**Lemma 38** ([15], Corollary 5.3). *Checking if a computable b-tree is non-recurring is  $\Pi_1^1$ -complete.*

We are ready now to show the main result in this subsection.

**Proposition 39.** *The problem of checking if a machine is non-recurring through some state  $q_r$  on the empty word  $\varepsilon$  is  $\Pi_1^1$ -complete.*

**PROOF.** To show membership, we present a reduction that maps a machine  $M = (Q, \Gamma, \delta)$  to a computable marked b-tree  $(T, \mu)$  such that  $M$  is non-recurring through a given state  $q_r \in Q$  on the empty word  $\varepsilon$  if and only if  $(T, \mu)$  is non-recurring. To define  $(T, \mu)$ , we consider an (arbitrarily chosen) enumeration  $q_1, \dots, q_n$  of the states in  $Q$ .

- Let  $T$  be the set containing a list of natural numbers  $i_1, \dots, i_n$  if there is a partial run  $\rho_1, \dots, \rho_n$  of  $M$  on  $\varepsilon$  such that  $\rho_j$  features the state  $q_{i_j}$  for every  $1 \leq j \leq n$ .
- Let  $\mu$  be the function that maps a list  $u \in T$  to 1 if and only if  $q_i = q_r$  where  $i$  is the last element in  $u$ . That is, if the last element of  $u$  is the index that corresponds to  $q_r$  in the enumeration  $q_1, \dots, q_n$ .

For every infinite branch of  $T$ , there is an infinite run of  $M$  and vice-versa. Furthermore, by the definition of  $\mu$ , a branch of  $(T, \mu)$  containing infinitely many marked nodes corresponds to a run of  $M$  visiting  $q_r$  infinitely many times. Therefore,  $M$  is non-recurring through  $q_r$  if and only if  $(T, \mu)$  is non-recurring.

For hardness, we present a reduction that maps a computable  $\omega$ -tree  $T$  to a non-deterministic machine  $M = (Q, \Gamma, \delta)$  such that  $T$  is well-founded if and only if  $M$  is non-recurring through a state

$q_r \in Q$  on the empty word  $\epsilon$ . Intuitively, the machine  $M$  proceeds by doing a traversal of the full  $\omega$ -tree; formally, it implements the following instructions on input  $\epsilon$ :

- (1) Initialise the variable  $u = 0$ , which stores a list of natural numbers.
- (2) If  $u \notin T$ , replace the last element  $i$  in  $u$  with  $i + 1$ .
- (3) If  $u \in T$ , make a non-deterministic choice between the following options:
  - (a) Replace the last element  $i$  in  $u$  with  $i + 1$ .
  - (b) Append 0 to the list stored in  $u$  and visit the state  $q_r$ .
- (4) Go to (2).

We can effectively check if a list  $u$  is a node in  $T$  above because  $T$  is a computable  $\omega$ -tree and hence, so is function  $\chi_T$ . Intuitively, each run of  $M$  on the empty word corresponds to a traversal of a branch in  $T$ ; note how we use non-determinism in (3) to alternatively visit the sibling (Instruction 3.a) or the child (Instruction 3.b) of a node in the tree. Furthermore, note that  $M$  only visits  $q_r$  when it moves deeper on a given branch; that is, when it executes instruction (3.b). Therefore, there is a run of  $M$  visiting  $q_r$  infinitely often if and only if there is an infinite branch in  $T$ .  $\square$

## A.2 Reductions between Turing Machine Problems

**Proposition 40.** *The problem of checking if a machine is universally non-recurring through a given state  $q_r$  is  $\Pi_1^1$ -complete.*

**PROOF.** To show membership, we present a reduction that maps a machine  $M$  to another machine  $M'$  such that  $M$  is universally non-recurring through a state  $q_r$  if and only if  $M'$  is non-recurring through a state  $q'_r$  on  $\epsilon$ . On input  $\epsilon$ , the machine  $M'$  first guesses some input word and then simulates  $M$  on this input. Formally, it executes the following instructions:

- (1) Make a non-deterministic choice to decide whether to go to (2) or to (3).
- (2) Replace the first occurrence of the blank symbol  $B$  in the input tape with some non-deterministically chosen symbol in the input alphabet of  $M$ . Then, go to (1).
- (3) Simulate  $M$  on the (finite) word written down in the input tape. During this simulation, visit  $q'_r$  whenever  $M$  would have visited  $q_r$ .

Note that there are infinite runs of  $M'$  on  $\epsilon$  where the machine never executes Instruction 3. This does not invalidate our reduction since  $M'$  never visits  $q'_r$  in these branches.

To show hardness, we present a reduction that maps a machine  $M$  to another machine  $M'$  such that  $M$  is non-recurring through a state  $q_r$  on  $\epsilon$  if and only if  $M'$  is universally non-recurring through a state  $q'_r$ . The machine  $M'$  first discards its input by replacing it with a special symbol that is treated like the blank symbol  $B$ .<sup>4</sup> Then,  $M'$  simulates  $M$  on  $\epsilon$ ; during this simulation,  $M'$  visits  $q'_r$  whenever  $M$  would have visited  $q_r$ .  $\square$

**Proposition 41.** *Checking if a machine is robust non-recurring through  $q_r$  is  $\Pi_1^1$ -complete.*

**PROOF.** To show membership, we present a reduction from a machine  $M$  to a machine  $M'$  such that  $M$  is robust non-recurring through a state  $q_r$  if and only if  $M'$  is universally non-recurring through a state  $q'_r$ . The machine  $M'$  scans its input and halts if it does not encode a configuration of  $M$ . Otherwise,  $M'$  simulates  $M$  starting on this input configuration; during this simulation,  $M'$  visits  $q'_r$  whenever  $M$  would have visited  $q_r$ .

To show hardness, we present a reduction from a machine  $M$  to another machine  $M'$  such that  $M$  is non-recurring through a state  $q_r$  on the empty word  $\epsilon$  if and only if  $M'$  is robust non-recurring through a state  $q'_r$ . The machine  $M'$  executes the following instructions:

- (1) Halt if the input does not contain some configuration  $\rho$  of  $M$ .

<sup>4</sup>We consider a special symbol here because, as per our definition, machines may not print the blank symbol  $B$ .

- (2) If the configuration in the tape  $\rho$  features the special state  $q_r$ , then visit  $q'_r$ .
- (3) After the encoding of  $\rho$  in the tape, (non-deterministically) simulate a run of  $M$  on  $\varepsilon$  until it terminates or you reach the configuration  $\rho$ . If the run terminates, without finding  $\rho$ , halt. Otherwise, continue in (4).
- (4) If  $\text{Next}_M(\rho)$  is empty, halt. Otherwise, replace the configuration  $\rho$  in the tape with a non-deterministically chosen configuration in  $\text{Next}_M(\rho)$ , and go to (1).

Intuitively speaking, Instruction 3 implements a reachability check for the configuration  $\rho$  in the tape. That is, this procedure ensures that this configuration is reachable from the starting configuration of  $M$  on the empty word  $\varepsilon$  by some run. Note that the reachability check makes non-deterministic choices itself. So it can happen that  $M'$  terminates early or even runs forever because it picks the wrong run in Instruction 3. This does not invalidate our reduction though since on those runs,  $M'$  only visits  $q'_r$  finitely many times.

If  $M'$  is robust non-recurring through  $q'_r$ , then it is also non-recurring on the starting configuration with the encoding of the starting configuration of  $M$  on  $\varepsilon$ . Since  $M'$  uses non-determinism to simulate all possible runs  $M$  on  $\varepsilon$  and visits  $q'_r$  whenever  $M$  would have visited  $q_r$ , we conclude that  $M$  is non-recurring through  $q_r$  on  $\varepsilon$ .

Suppose that there is a configuration  $\rho'$  of  $M'$  that may lead to a run that visits  $q'_r$  infinitely many times. In turn, this implies that there is a configuration  $\rho$  of  $M$  that leads to a run of  $M$  that visits  $q_r$  infinitely many times. Moreover, all the configurations of  $M$  in this infinite run are reachable from the start configuration of  $M$  on  $\varepsilon$  because of the check implemented in Instruction 3. Therefore,  $M$  is recurring through  $q_r$  on the empty word.  $\square$

## B Proofs for Section 4 (Knowledge base termination)

**Proposition 18.** *If  $F$  has a wild frontier of  $\rho$  overseen by  $w$ , and  $\rho'$  is reachable in one step by a transition of  $M$ , then there exists a restricted derivation  $\mathcal{D}_{\rho \rightarrow \rho'} = F, \dots, F'$  such that  $F'$  has a wild frontier of  $\rho'$  overseen by  $w'$ , where  $w' \neq w$  is a fresh existential if  $\rho$  is in  $q_r$ , and  $w' = w$  otherwise.*

**PROOF.** We consider the case where  $\rho = \langle n, t, p, q \rangle$ , with  $q \neq q_r$ , and where  $\rho'$  is obtained from  $\rho$  because  $(b, q', \rightarrow) \in \delta(t(p), q)$ . We consider  $x_1, \dots, x_{n+1}, w$  as provided by the definition of a wild frontier of configuration  $\rho$ .

- we start by applying Rule  $R_{\neg q_r}^{\rightarrow}$ , mapping  $x$  to  $x_p$ ,  $y$  to  $x_{p+1}$  and  $w$  to  $w$ . This produces the atoms  $q'(x'_{p+1}, w), c(x'_{p+1}, w), b(x'_p, w), C_L(x'_p, w), C_R(x'_{p+1}, w), R(x'_p, x'_{p+1}, w), F(x_p, x'_p, w), F(x_{p+1}, x'_{p+1}, w), \text{brSet}(x'_p, w), \text{brSet}(x'_{p+1}, w)$ ;
- we apply Rule  $R_{C_L}$   $p - 1$  times. The  $i^{\text{th}}$  (for  $i$  from 1 to  $p - 1$ ) application maps  $x$  to  $x_{p-i+1}$ ,  $x'$  to  $x'_{p-i+1}$ ,  $y$  to  $x_{p-i}$ ,  $w'$  and  $w$  to  $w$ . It creates atoms  $F(x_{p-i}, x'_{p-i}, w), R(x'_{p-i}, x'_{p-i+1}, w), t(p-i)(x'_{p-i}, w), C_L(x'_{p-i}, w), \text{brSet}(x'_{p-i}, w')$ ;
- we apply Rule  $R_{C_R}$   $n - p$  times. The  $i^{\text{th}}$  (for  $i$  from 1 to  $n - p$ ) application maps  $x$  to  $x_{p+i}$ ,  $x'$  to  $x'_{p+i}$ ,  $y$  to  $x_{p+i+1}$ ,  $w'$  and  $w$  to  $w$ . It creates atoms  $F(x_{p+i+1}, x'_{p+i+1}, w), R(x'_{p+i}, x'_{p+i+1}, w), t(p+i+1)(x'_{p+i+1}, w), C_R(x'_{p+i+1}, w), \text{brSet}(x'_{p+i+1}, w)$ ;
- we apply Rule  $R_{\text{End}}$ , mapping  $x'$  to  $x'_{n+1}$ ,  $x$  to  $x_{n+1}$ ,  $w$  and  $w'$  to  $w$ . It creates the atoms  $R(x'_{n+1}, x'_{n+2}, w), B(x'_{n+2}, w), \text{End}(x'_{n+2}, w), \text{brSet}(x'_{n+2}, w)$ .

The result of that derivation has a wild frontier of configuration  $\rho'$  overseen by  $w$ , as witnessed by terms  $x'_1, \dots, x'_{n+2}$ .

If  $\rho'$  is obtained from  $\rho$  because  $(b, q', \leftarrow) \in \delta(t(p), q)$ , with  $q \neq q_r$ , we consider  $x_1, \dots, x_{n+1}, w$  as provided by the definition of a wild frontier of configuration  $\rho$ .

- we start by applying Rule  $R_{q_r}^{\leftarrow}$ , mapping  $x$  to  $x_p$ ,  $y$  to  $x_{p-1}$ ,  $w$  to  $w$ . This produces the atoms  $q'(x'_{p-1}, w)$ ,  $c(x'_{p-1}, w)$ ,  $b(x'_p, w)$ ,  $C_L(x'_{p-1}, w)$ ,  $C_R(x'_p, w)$ ,  $R(x'_{p-1}, x'_p, w)$ ,  $F(x_p, x'_p, w)$ ,  $F(x_{p-1}, x'_{p-1}, w)$ ,  $brSet(x'_p, w)$ ,  $brSet(x'_{p-1}, w)$ ;
- we apply Rule  $R_{C_L}$   $p - 2$  times. The  $i^{\text{th}}$  (for  $i$  from 1 to  $p - 2$ ) application maps  $x$  to  $x_{p-i}$ ,  $x'$  to  $x'_{p-1}$ ,  $y$  to  $x_{p-i-1}$ , and  $w, w'$  to  $w$ . It creates atoms  $F(x_{p-i-1}, x'_{p-i-1}, w)$ ,  $R(x'_{p-i-1}, x'_{p-i}, w)$ ,  $t(p - i - 1)(x'_{p-i-1}, w)$ ,  $C_L(x_{p-i-1}, w)$ ,  $brSet(x_{p-i-1}, w)$ ;
- we apply Rule  $R_{C_R}$   $n - p + 1$  times. The  $i^{\text{th}}$  (for  $i$  from 1 to  $n - p + 1$ ) application maps  $x$  to  $x_{p-1+i}$ ,  $x'$  to  $x'_{p-1+i}$ ,  $y$  to  $x_{p+i}$ , and  $w, w'$  to  $w$ . It creates atoms  $F(x_{p+i}, x'_{p+i}, w)$ ,  $R(x'_{p+i-1}, x'_{p+i}, w)$ ,  $t(p - i - 1)(x'_{p+i}, w)$ ,  $C_R(x'_{p+i}, w)$ ,  $brSet(x'_{p+i}, w)$ ;
- we apply Rule  $R_{\text{End}}$ , mapping  $x'$  to  $x'_{n+1}$ ,  $x$  to  $x_{n+1}$ ,  $w$  and  $w'$  to  $w$ . It creates the atoms  $R(x'_{n+1}, x'_{n+2}, w)$ ,  $B(x'_{n+2}, w)$ ,  $\text{End}(x'_{n+2}, w)$ ,  $brSet(x'_{n+2}, w)$ .

The result of that derivation has a wild frontier of configuration  $\rho'$  overseen by  $w$ , as witnessed by terms  $x'_1, \dots, x'_{n+2}$ .

If  $\rho'$  is obtained from  $\rho$  because  $(b, q', \rightarrow) \in \delta(t(p), q_r)$ , we consider  $x_1, \dots, w_{n+1}$  as provided by the definition of a wild frontier of configuration  $\rho$ .

- we start by applying Rule  $R_{q_r}^{\rightarrow}$ , mapping  $x$  to  $x_p$ ,  $y$  to  $x_{p+1}$  and  $w$  to  $w$ . This rule application produces the atoms  $q'(x'_{p+1}, w')$ ,  $c(x'_{p+1}, w')$ ,  $b(x'_p, w')$ ,  $C_L(x'_p, w')$ ,  $C_R(x'_{p+1}, w')$ ,  $R(x'_p, x'_{p+1}, w')$ ,  $F(x_p, x'_p, w')$ ,  $F(x_{p+1}, x'_{p+1}, w')$ ,  $brSet(x'_p, w')$ ,  $brSet(x'_{p+1}, w')$ ;
- we apply Rule  $R_{C_L}$   $p - 1$  times. The  $i^{\text{th}}$  (for  $i$  from 1 to  $p - 1$ ) application maps  $x$  to  $x_{p-i+1}$ ,  $x'$  to  $x'_{p-i+1}$ ,  $y$  to  $x_{p-i}$ ,  $w$  to  $w$  and  $w'$  to  $w'$ . It creates atoms  $F(x_{p-i}, x'_{p-i}, w')$ ,  $R(x'_{p-i}, x'_{p-i+1}, w')$ ,  $t(p - i - 1)(x'_{p-i}, w')$ ,  $C_L(x'_{p-i}, w')$ ,  $brSet(x'_{p-i}, w')$ ;
- we apply Rule  $R_{C_R}$   $n - p$  times. The  $i^{\text{th}}$  (for  $i$  from 1 to  $n - p$ ) application maps  $x$  to  $x_{p+i}$ ,  $x'$  to  $x'_{p+i}$ ,  $y$  to  $x_{p+i+1}$ ,  $w$  to  $w$  and  $w'$  to  $w'$ . It creates atoms  $F(x_{p+i+1}, x'_{p+i+1}, w')$ ,  $R(x'_{p+i}, x'_{p+i+1}, w')$ ,  $t(p + i + 1)(x'_{p+i+1}, w')$ ,  $C_R(x'_{p+i+1}, w')$ ,  $brSet(x'_{p+i+1}, w')$ ;
- we apply Rule  $R_{\text{End}}$ , mapping  $x'$  to  $x'_{n+1}$ ,  $x$  to  $x_{n+1}$ ,  $w$  to  $w$  and  $w'$  to  $w'$ . It creates atoms  $R(x'_{n+1}, x'_{n+2}, w')$ ,  $B(x'_{n+2}, w')$ ,  $\text{End}(x'_{n+2}, w')$ ,  $brSet(x'_{n+2}, w')$ .

The result of that derivation has a wild frontier of configuration  $\rho'$  overseen by  $w'$ , as witnessed by terms  $x'_1, \dots, x'_{n+2}$ .

If  $\rho'$  is obtained from  $\rho$  because  $(b, q', \leftarrow) \in \delta(t(p), q_r)$ , we consider  $x_1, \dots, x_{n+1}, w$  as provided by the definition of a wild frontier of configuration  $\rho$ .

- we start by applying Rule  $R_{q_r}^{\leftarrow}$ , mapping  $x$  to  $x_p$ ,  $y$  to  $x_{p-1}$ ,  $w$  to  $w$ . This rule application produces the atoms  $q'(x'_{p-1}, w')$ ,  $c(x'_{p-1}, w')$ ,  $b(x'_p, w')$ ,  $C_L(x'_{p-1}, w')$ ,  $C_R(x'_p, w')$ ,  $R(x'_{p-1}, x'_p, w')$ ,  $F(x_p, x'_p, w')$ ,  $F(x_{p-1}, x'_{p-1}, w')$ ,  $brSet(x'_p, w')$ ,  $brSet(x'_{p-1}, w')$ ;
- we apply Rule  $R_{C_L}$   $p - 2$  times. The  $i^{\text{th}}$  (for  $i$  from 1 to  $p - 2$ ) application maps  $x$  to  $x_{p-i}$ ,  $x'$  to  $x'_{p-1}$ ,  $y$  to  $x_{p-i-1}$ , and  $w$  to  $w$  and  $w'$  to  $w'$ . It creates atoms  $F(x_{p-i-1}, x'_{p-i-1}, w')$ ,  $R(x'_{p-i-1}, x'_{p-i}, w')$ ,  $t(p - i - 1)(x'_{p-i-1}, w')$ ,  $C_L(x_{p-i-1}, w')$ ,  $brSet(x_{p-i-1}, w')$ ;
- we apply Rule  $R_{C_R}$   $n - p + 1$  times. The  $i^{\text{th}}$  (for  $i$  from 1 to  $n - p + 1$ ) application maps  $x$  to  $x_{p-1+i}$ ,  $x'$  to  $x'_{p-1+i}$ ,  $y$  to  $x_{p+i}$ , and  $w$  to  $w$  and  $w'$  to  $w'$ . It creates atoms  $F(x_{p+i}, x'_{p+i}, w')$ ,  $R(x'_{p+i-1}, x'_{p+i}, w')$ ,  $t(p - i - 1)(x'_{p+i}, w')$ ,  $C_R(x'_{p+i}, w')$ ,  $brSet(x'_{p+i}, w')$ ;
- we apply Rule  $R_{\text{End}}$ , mapping  $x'$  to  $x'_{n+1}$ ,  $x$  to  $x_{n+1}$ ,  $w$  to  $w$  and  $w'$  to  $w'$ . It creates atoms  $R(x'_{n+1}, x'_{n+2}, w')$ ,  $B(x'_{n+2}, w')$ ,  $\text{End}(x'_{n+2}, w')$ ,  $brSet(x'_{n+2}, w')$ .

The result of that derivation has a wild frontier of configuration  $\rho'$  overseen by  $w'$ , as witnessed by terms  $x'_1, \dots, x'_{n+2}$ .  $\square$

A rule is *datalog* if its head does not contain any existentially quantified variable.

**Proposition 42.** *Let  $F_0, \dots, F_k$  be restricted derivation. Let  $w^* \in \text{terms}(F_k) \setminus \text{terms}(F_0)$  such that  $\text{Real}(w^*) \in F_k$ . Then for any  $j > k$ , the only rules generating  $w^*$  as a last argument having non-obsolete triggers on  $F_j$  are datalog rules.*

**PROOF.** We consider a non-datalog rule  $R$  and  $\sigma$  a homomorphism of  $\text{body}(R)$  into  $F_j$ . We prove that  $\sigma$  can be extended in a homomorphism of  $\text{head}(R)$  into  $F_j$ , showing that  $\langle R, \sigma \rangle$  is obsolete. Note that as  $\text{Real}(w^*) \in F_k$ , it holds that Rule  $R_{\text{Brake}}$  has been applied by mapping  $w$  to  $w^*$ .

- Rules  $R_{\neg q_r}^\leftarrow$  and  $R_{\neg q_r}^\rightarrow$ : extend  $\sigma$  by mapping  $x'$  and  $y'$  to  $\sigma(w) = w^*$ .
- Rules  $R_{q_r}^\leftarrow$  and  $R_{q_r}^\rightarrow$ : extend  $\sigma$  by mapping  $x', y', w'$  to  $\sigma(w) = w^*$ .
- Rules  $R_{C_L}$ ,  $R_{C_R}$ , and  $R_{\text{End}}$ : extend  $\sigma$  by mapping  $y'$  to  $\sigma(w') = w^*$ .

$\square$

**Lemma 19.** *Let  $(\rho_i)_{i \in \mathbb{N}}$  be a run of  $M$  on the empty word that visits  $q_r$  infinitely often. There exists an infinite restricted chase sequence for  $\langle \Sigma_M, D_\varepsilon \rangle$ .*

**PROOF.** Let  $(i_j)_{j \in \mathbb{N}}$  be the infinite strictly increasing sequence of integers such that  $i_1 = 1$  and  $\rho_k$  is in  $q_r$  if and only if  $k = i_j$  for some  $j$ . We denote by  $\mathcal{D}_{\rho_{i_j} \rightarrow \rho_{i_{j+1}}}$  the concatenation of the restricted derivations provided by Proposition 18. Let us consider the derivation build by induction:

- $\mathcal{D}_1 = \mathcal{D}_{\rho_{i_1} \rightarrow \rho_{i_2}}$
- $\mathcal{D}'_1$  extends  $\mathcal{D}_1$  by the application of Rule  $R_{\text{Brake}}$  mapping  $w$  to the brake overseeing the wild frontier of the last element of  $\mathcal{D}_1$ , as well as by applying any datalog rule mapping  $w$  to that brake.
- $\mathcal{D}_j$  extends  $\mathcal{D}'_{j-1}$  by the derivation  $\mathcal{D}_{\rho_{i_j} \rightarrow \rho_{i_{j+1}}}$ ;
- $\mathcal{D}'_j$  extends  $\mathcal{D}_j$  by the application of Rule  $R_{\text{Brake}}$  mapping  $w$  to the brake overseeing the wild frontier of the last element of  $\mathcal{D}_j$ , and by applying Rule  $R_{\text{nextBr}}$  in any possible way that maps  $w$  to the brake overseeing the wild frontier of the last element of  $\mathcal{D}_j$ , as well by applying any datalog rule mapping  $w$  to that brake.

This derivation is fair:

- any created atom has a brake as argument;
- brakes are created exactly once in each derivation  $\mathcal{D}_{\rho_{i_j} \rightarrow \rho_{i_{j+1}}}$  (by definition of  $(i_j)_{j \in \mathbb{N}}$ ); let us call  $w_1$  the brake appearing in  $D_\varepsilon$ , and  $w_{j+1}$  the brake created in  $\mathcal{D}_{\rho_{i_j} \rightarrow \rho_{i_{j+1}}}$ ;
- by Proposition 42, the application of Rule  $R_{\text{Brake}}$  mapping  $w$  to  $w_j$  deactivates any trigger of a non-datalog rule mapping creating an atom with  $w_j$  as a last argument;
- by definition of  $\mathcal{D}'_j$ , all datalog rules creating an atom with  $w_j$  as last argument are applied.

$\square$

**Lemma 43.** *Let  $F$  be the result of an infinite restricted chase sequence  $F_0, F_1, \dots$  from  $\langle \Sigma_M, D \rangle$  for some  $D$ . For any  $w$  such that  $\text{Brake}(w) \in F$ , there are finitely many atoms having  $w$  as last argument in  $F$ . There is thus an infinite amount of brakes in  $F$ .*

**PROOF.** Consider a term  $w$  such that  $\text{Brake}(w) \in F$ , which we call a brake. By fairness and Rule  $R_{\text{Brake}}$ , there must be some integer  $i$  such that  $\text{Real}(w) \in F_i$ . At this step, there is a finite number of atoms with  $w$  as last argument, and by Proposition 42, the only rules that can generate

such atoms after step  $i$  are datalog. Rule  $R_{\text{Brake}}$  only generates atoms over  $w$ , so it is applicable at most one, and will yield at most 6 new atoms. Thus, the only rule left is Rule  $R_{\text{nextBr}}$ .

Only two rules create new Brake-atoms that do not already appear in their bodies, which are Rules  $R_{q_r}^{\rightarrow}$  and  $R_{q_r}^{\leftarrow}$ . Both these rules also generate an atom of the form  $\text{nextBr}(w, w')$ , where  $\text{Brake}(w)$  is the brake in their body, and  $\text{Brake}(w')$  is the newly created brake. As this is the only way to generate  $\text{nextBr}$ -atoms, the predicate  $\text{nextBr}$  defines a forest relationship over the brakes, where the root of each tree is a term  $w_0$  such that  $\text{Brake}(w_0) \in D$ . There is thus a finite number of trees. We then show that Rule  $R_{\text{nextBr}}$  can only create a finite number of atoms by induction on this forest structure.

- If  $\text{Brake}(w) \in D$ , then all the atoms of the form  $\text{nextBr}(w', w)$  are in  $D$ , so  $w'$  is in  $D$  too. Thus, Rule  $R_{\text{nextBr}}$  can only create sets of atoms of the form  $\text{brSet}(x, w')$ , where  $x$  is a database term. As there is a finite amount of database terms, this yields a finite number of atoms.
- If  $\text{Brake}(w') \in F \setminus D$ ,  $\text{nextBr}(w, w') \in F$  and there is a finite number of atoms having  $w$  as last argument, then first notice that  $w$  is the only term such that  $\text{nextBr}(w, w') \in F$ , since Rules  $R_{q_r}^{\rightarrow}$  and  $R_{q_r}^{\leftarrow}$  both generate  $\text{nextBr}$ -atoms featuring an existential variable in second position. Then, as there is a finite amount of atoms featuring  $w$  as their last argument, there is a finite amount of terms  $x$  such that  $\text{brSet}(x, w) \subseteq F$ . Thus, Rule  $R_{\text{nextBr}}$  generates at most  $\text{brSet}(x, w')$  for all these terms, which represents a finite number of atoms.

Thus, there is a finite number of atoms that feature a given brake as their last argument. As  $F_0, F_1, \dots$  is infinite,  $F$  must have an infinite amount of atoms, that were generated during the chase. Since  $\text{Brake}(w)$  is required in the body of all the rules where  $w$  appears as the last argument of an atom, there is thus an infinite amount of brakes in  $F$ .  $\square$

**Lemma 21.** *For all databases  $D$ , and all infinite chase sequences from  $\langle \Sigma_M, D \rangle$  with result  $F$ , there is an infinite sequence  $(A_n)_{n \in \mathbb{N}}$  of state atoms of  $F$  such that:*

- $A_0 \in D$ ;
- $A_n \prec A_{n+1}$  for all  $n \in \mathbb{N}$ ;
- for infinitely many  $i \in \mathbb{N}$ ,  $A_i$  is of the shape  $q_r(t_i, w_i)$ .

**PROOF.** Since the rules that introduce state atoms (Rules  $R_{q_r}^{\leftarrow}$ ,  $R_{q_r}^{\rightarrow}$ ,  $R_{\neg q_r}^{\leftarrow}$ , and  $R_{\neg q_r}^{\rightarrow}$ ) feature a state atom in their body,  $\prec$  defines a forest structure over state atoms, where the root of each tree is an atom of the database. There is thus a finite amount of trees (as there is a finite amount of atoms in the database). By Lemma 43, there is an infinite amount of brakes in  $F$ . Then, since the rules that introduce new brakes (Rules  $R_{q_r}^{\rightarrow}$  and  $R_{q_r}^{\leftarrow}$ ) introduce a state atom too, there is an infinite number of state atoms. Thus, one of the trees must be infinite. In addition, since there is a finite amount of atoms that feature a given brake as last argument, and each state atom features a brake as last argument, each state atom only has a finite number of successors for  $\prec$ . Indeed, infinitely many successors would require infinitely many rule applications, and thus infinitely many atoms featuring the same last argument as the state atom. We thus have an infinite tree with finite branching. It thus features an infinite branch, which must contain infinitely many  $q_r$ -atoms (as there are infinitely many  $q_r$ -atoms), by König's lemma.  $\square$

To each state atom, we associate both a set of atoms and a configuration.

**Definition 44** (Atoms associated with a state atom). *Let  $F_k$  be a fact set occurring in a chase derivation from  $D_\rho$ . The atoms associated with a state atom  $q(x, w)$  in  $F_k$  is the largest subset of  $F_k$  whose terms are included in  $\{x, w\} \cup X_i$  and such that:*

- $X_i$  is the set of terms reachable or co-reachable through an  $R$ -path from  $x$  not going through a brake;
- $w$  can only appear in the last position of atoms.

**Definition 45** (Configuration of a state atom). Let  $F_k$  appearing in a restricted chase sequence for  $\langle \Sigma, D_\rho \rangle$ . The configuration associated with a state atom  $q(x, w)$ , written  $\text{conf}(q(x, w))$ , is defined by induction:

- if  $q(x, w) \in D_\rho$ ,  $\text{conf}(q(x, w)) = \rho$
- otherwise, let  $A$  be the unique state atom such that  $A \prec q(x, w)$ . Let  $(q'b, d)$  be the element of  $\delta(q, a)$  due to which the rule whose application generated  $q(w, x)$  belongs to  $\Sigma_M$ . We define  $\text{conf}(q(x, w))$  as the configuration obtained from  $\text{conf}(A)$  where the content of the head cell of  $\text{conf}(A)$  is replaced by  $b$ , the head moves by  $d$  and switches to state  $q'$ .

Note that the above definition implies that if  $A_p$  is the parent of  $A$ , then  $\text{conf}(A)$  is reachable in one transition of  $M$  from  $\text{conf}(A_p)$ . Intuitively, the configuration of a state atom is encoded by the atoms associated with it. However, the restricted derivation may not have derived all such atoms, hence we consider a weaker notion, that we coin consistency.

**Definition 46** (Consistency). A set of atoms  $A$  associated with a state atom is consistent with the configuration  $\langle n, t, p, q \rangle$  if:

- there exists  $x$  and  $w$  such that  $q(x, w)$  is the only state atom in  $A$ , and  $\text{conf}(x)$  is in state  $q$ ;
- $a(x, w)$  is the only letter predicate having  $x$  as argument in  $A$ , and  $t(p) = a$ ;
- if there is an  $R$  path of length  $i$  from  $x$  to  $x'$ , and there is an atom  $a(x', x'')$ , then  $x'' = w$ ,  $p + i \leq n + 1$  and  $t(p + i) = a$ ;
- there exists at most one atom  $\text{End}(x', x'')$  in  $A$ , and if it exists then  $x'' = w$  and there is an  $R$ -path from  $x$  to  $x'$  of length  $i$ , such that  $p + i = n + 1$ ;
- if there is an  $R$  path of length  $i$  from  $x'$  to  $x$ , and there is an atom  $a(x', x'')$ , then  $x'' = w$ ,  $p - i \geq 1$  and  $t(p - i) = a$ .

As expected, the set of atoms associated with a state atom is consistent with its configuration, and this allows us to prove Lemma 22.

**Proposition 47.** Let  $F_k$  appearing in a restricted chase sequence for  $\langle \Sigma, D_\rho \rangle$ . For any state atom  $A$  of  $F_k$ , the set of atoms associated with  $A$  is consistent with  $\text{conf}(A)$ .

**PROOF.** We prove the result by induction. If  $A$  is a state atom that does not have any parent, then  $A \in F_0 = D_\rho$ . The set of atoms associated with  $A$  is  $D_\rho$ , which is consistent with the initial configuration of  $M$  on  $\rho$  by definition, which is configuration of  $A$ ;

Otherwise, let  $A = q'(y', w)$  be a state atom of  $F_k$ . We prove the result assuming that  $A$  has been created by the application of Rule  $R_{-q_r}^{\leftarrow}$ , mapping  $x$  to  $x_p$ ,  $w$  to  $w$  and  $y$  to  $y_p$ .  $A_p$ , the parent of  $A$ , is thus of the shape  $q(x_p, w)$  (other possible case would be  $A$  being created by Rules  $R_{-q_r}^{\rightarrow}$ ,  $R_{q_r}^{\leftarrow}$  or  $R_{q_r}^{\rightarrow}$ , which are treated similarly). It is easy to check that any term reachable from  $y'$  by an  $R$ -path not going through a brake is either created by the same rule application as  $y'$ , or has been created by an application of Rule  $R_{C_R}$  mapping  $x'$  to a term reachable by an  $R$ -path from  $y'$  (and similarly for terms co-reachable and Rule  $R_{C_L}$ ). Then:

- if there exists  $y'_{-1}$  such that  $R(y'_{-1}, y', w) \in F_k$ , then  $y'_{-1}$  has been created by the application of Rule  $R_{C_L}$ , in which case  $a(y'_{-1}, w)$  is generated if the cell two positions on the left of the head of  $\text{conf}(A_p)$  contains an  $a$ , that is, if the cell one position on the left of the head of  $\text{conf}(A)$  contains an  $a$ ; predecessors of  $y'$  further away from  $y'$  are treated by induction in a similar way.

- there exists a  $y'_{+1}$  such that  $R(y', y'_{+1}, w) \in F_k$ , as such an element is created by the application of Rule  $R_{\neg q_r}^{\leftarrow}$ . The same application create the atom  $b(y'_{+1}, w)$ , which is consistent with the fact that  $\text{conf}(A_p)$  contains a  $b$  in the first cell at the right of the head; cells further to the right are treated similarly to cells to the left, which are necessarily created by Rule  $R_{C_R}$ .
- the only way to derive an atom of the shape  $\text{End}(x', x'')$  is to apply Rule  $R_{\text{End}}$ , which can only be done after  $n - p$  rule applications of Rule  $R_{C_R}$ , yielding a path of length  $n - p + 2$  from position  $p - 1$  of the current configuration, which fulfills the condition of Definition 46 (remember that the length of  $\text{conf}(A)$  is incremented by 1 with respect to the length of  $\text{conf}(A_p)$ ).

□

**Lemma 22.** *For every configuration  $\rho$ , if the restricted chase does not terminate on  $\langle \Sigma_M, D_\rho \rangle$  then there exists a run of  $M$  on  $\rho$  which visits  $q_r$  infinitely many times.*

PROOF. We consider the sequence of states atoms  $(A_n)_{n \in \mathbb{N}}$  provided by Lemma 21, and the sequence  $(\text{conf}(A_n))_{n \in \mathbb{N}}$ .

- $\text{conf}(A_0)$  is the starting configuration of  $M$  on  $\varepsilon$ , and thus a run of  $M$  on that configuration;
- if  $(A_n)_{n \in \mathbb{N}}$  is not a run, there exists a smallest  $j \in \mathbb{N}$  such that  $(\text{conf}(A_n))_{1 \leq n \leq j}$  is not a run.  $\text{conf}(A_{j-1})$  is consistent with the set of atoms associated with  $A_{j-1}$  by Proposition 47. Hence  $\text{conf}(A_j)$  is obtained by applying the transition correponding to the rule creating  $A_j$ , and thus  $(\text{conf}(A_n))_{1 \leq n \leq j}$  is a run, which leads to a contradiction.

□

## C Proofs for Section 5 (Rule set termination)

The following lemmas are used in later proofs of the section.

**Lemma 48.** *For all databases  $D$ , and every  $F$  result of a chase sequence for  $\langle \Sigma_M, D \rangle$ , if the atoms  $F(x, z, w)$  and  $F(y, z, w)$  are both in  $F$  and  $z$  is a null, then  $x = y$ .*

PROOF. This result follows from the fact that whenever an  $F$ -atom appears in the head of a rule, it contains an existentially quantified variable in second position, and no two  $F$ -atoms contain this variable in second position. Thus, if  $z$  is a null and  $x$  and  $y$  are different, the atoms  $F(x, z, w)$  and  $F(y, z, w)$  must have been generated by two rule applications, which both introduce  $z$ , which is impossible. □

**Lemma 49.** *For all databases  $D$ , and every  $F$  result of a chase sequence for  $\langle \Sigma_M, D \rangle$ , for each null  $y$  in semiterms  $(F \setminus D)$ , there are a unique  $w$  and a unique  $a \in \Gamma \cup \{B\}$  such that  $a(y, w) \in F$ .*

PROOF. Whenever there is an existentially quantified variable  $x$  in the head of a rule in  $\Sigma_M \setminus \{R_{\text{Brake}}\}$ , it appears in a unique atom of the form  $a(x, w)$  in the same head. In addition, all the atoms of the same form in heads of rules feature an existentially quantified variable in first position (except for  $R_{\text{Brake}}$ , which feature a brake). Thus, when the null  $y$  is introduced in the chase, there is a unique atom  $a(y, w)$  introduced along with it (hence implying existence), and no other rule can introduce an atom of the same form (hence implying uniqueness). □

**Lemma 28.** *For all database  $D$ , and every  $F$  result of a chase sequence for  $\langle \Sigma_M, D \rangle$ , the graph  $\text{bowtie}(A)$  is a finite bow tie for all state atoms  $A \in F \setminus D$ . In addition:*

- The center of the bow tie is the atom generated along with  $A$ , by rule  $R_{\neg q_r}^{\leftarrow}$ ,  $R_{\neg q_r}^{\rightarrow}$ ,  $R_{q_r}^{\leftarrow}$  or  $R_{q_r}^{\rightarrow}$ ;
- all the atoms in the left part of the bow tie are generated by rule  $R_{C_L}$ ;
- all the atoms in the right part of the bow tie are generated by rule  $R_{C_R}$ , except possibly the end of a maximal path, which may have been generated by rule  $R_{\text{End}}$ .

PROOF. First, notice that all the rules that generate R-atoms (over non-brake terms) generate R-atoms containing at least one existentially quantified variable. Three cases occur:

- Rules  $R_{\neg q_r}^{\leftarrow}, R_{\neg q_r}^{\rightarrow}, R_{q_r}^{\leftarrow}$  and  $R_{q_r}^{\rightarrow}$  generate an R-atom  $R(u, v, w)$  where  $u$  and  $v$  are both existentially quantified.
- Rule  $R_{C_L}$  generates an R-atom  $R(u, v, w)$  where  $u$  is existentially quantified and  $v$  is a frontier variable.
- Rules  $R_{C_R}$  and  $R_{End}$  generate an R-atom  $R(u, v, w)$  where  $u$  is a frontier variable and  $v$  is existentially quantified.

Thus, no connected component can contain two R-atoms that are generated using a rule among rules  $R_{\neg q_r}^{\leftarrow}, R_{\neg q_r}^{\rightarrow}, R_{q_r}^{\leftarrow}$  and  $R_{q_r}^{\rightarrow}$ . Indeed, these rules create a new connected component, and to connect two connected components, we need a rule generating an R-atom  $R(u, v, w)$  where  $u$  and  $v$  are both frontier variables, which is not the case with this rule set. This also implies that  $(\text{bowtie}(A), E_R)$  is acyclic, even when seen as an undirected graph, for the same reason.

Thus, since  $A = q(x, w)$  is generated by a rule among  $R_{\neg q_r}^{\leftarrow}, R_{\neg q_r}^{\rightarrow}, R_{q_r}^{\leftarrow}$  and  $R_{q_r}^{\rightarrow}$  along with an R-atom, all the other atoms in the connected component of  $x$  must have been generated by  $R_{C_L}, R_{C_R}$  or  $R_{End}$ . We assume here that  $A$  was generated by rule  $R_{\neg q_r}^{\leftarrow}$  or  $R_{q_r}^{\leftarrow}$ , as the other cases are symmetric. Then,  $A$  is generated along the atom  $R(x, y, w)$ , which will be the center of our bow tie, and atoms  $C_L(x, w)$  and  $C_R(y, w)$ .

We then consider the sets  $\text{bowtie}(A)_x^{-y}$  and  $\text{bowtie}(A)_y^{-x}$  as defined in Definition 27. First, as mentioned before, the undirected graph induced by  $(\text{bowtie}(A), E_R)$  is acyclic and connected, so these sets form a partition of  $\text{bowtie}(A)$ . Thus, it only remains to show that the subgraphs induced by  $\text{bowtie}(A)_x^{-y}$  on  $(\text{bowtie}(A), E_R^-)$  and by  $\text{bowtie}(A)_y^{-x}$  on  $(\text{bowtie}(A), E_R)$  are trees. Again, since both proofs are similar, we only prove it for the second graph.

A directed tree is an acyclic and connected graph such that each vertex has in-degree at most one. Since  $(\text{bowtie}(A), E_R)$  is acyclic, the subgraph induced by  $\text{bowtie}(A)_y^{-x}$  is acyclic too, and as it is a connected component, it is connected. Thus, it only remains to show that each term in  $\text{bowtie}(A)_y^{-x}$  has an in-degree of at most one. Our previous analysis of the rules entails that only a term  $t$  such that  $C_L(t, w) \in F$  can have an in-degree greater than one. Indeed, the only rule that can increase the in-degree of an existing element is rule  $R_{C_L}$ , which requires this atom in its body. We thus show that there is no  $t$  in  $\text{bowtie}(A)_y^{-x}$  such that  $C_L(t, w) \in F$ .

Only two kinds of rules can generate  $C_L$ -atoms (over non-brakes), which are transition rules ( $R_{\neg q_r}^{\leftarrow}, R_{\neg q_r}^{\rightarrow}, R_{q_r}^{\leftarrow}$  and  $R_{q_r}^{\rightarrow}$ ), and the rule  $R_{C_L}$ . All these rules generate atoms of the form  $C_L(u, w)$  where  $u$  is existentially quantified. As stated before, in  $\text{bowtie}(A)$ , only the atom  $R(x, y, w)$  has been generated using a transition rule, and every other R-atom has been generated using  $R_{C_L}$  or  $R_{C_R}$ . Now, for a contradiction, assume that  $t$  is the first term of  $\text{bowtie}(A)_y^{-x}$  introduced during the chase such that  $C_L(t, w) \in F$ . Since the trigger generating  $R(x, y, w)$  only generates  $C_L(x, w)$ , and  $x \notin \text{bowtie}(A)_y^{-x}$ , the term  $t$  has been generated by rule  $R_{C_L}$ . This means that there is a term  $u \in \text{bowtie}(A)_y^{-x}$  such that  $C_L(u, w) \in F$  before  $C_L(t, w)$  is introduced, which contradicts our hypothesis. Note that  $u$  does have to be in  $\text{bowtie}(A)_y^{-x}$ , since otherwise,  $t \notin \text{bowtie}(A)_y^{-x}$ , as no rule can connect two disjoint connected components.

Thus, there is no  $C_L$ -atom over a term in  $\text{bowtie}(A)_y^{-x}$ , meaning that  $(\text{bowtie}(A)_y^{-x}, E_R)$  is a tree. As mentioned before, an analog line of reasoning can be used to show that  $(\text{bowtie}(A)_x^{-y}, E_R^-)$  is also a tree, so  $\text{bowtie}(A)$  is indeed a bow tie. Note also that since no  $C_L$ -atom over a term in  $\text{bowtie}(A)_y^{-x}$ , all the R-atoms of the right part of the bow tie must have been generated by rule  $R_{C_R}$  or  $R_{End}$ . However, rule  $R_{End}$  generates a new null  $y$  such that  $C_R(y, w) \notin F$  (by the same description as previously), and both rules  $R_{C_R}$  and  $R_{End}$  require an atom of this form to extend a path. Thus, if an R-atom is generated using rule  $R_{End}$ , it is necessarily the end of a maximal path.

It remains to show that  $\text{bowtie}(A)$  is finite: if  $F$  is finite, then so is  $A$  and therefore  $\text{bowtie}(A)$ . Otherwise, note that all atoms in  $\text{bowtie}(A)$  are associated with the same brake  $w$ . Then, by Lemma 43,  $\text{bowtie}(A)$  must be finite.  $\square$

Recall that  $\mathcal{A} = (A_n)$  is the sequence of state atoms provided by Lemma 21.

**Lemma 30.** *For all  $n > 0$ ,  $\text{configs}(A_n)$  is finite, non-empty, and each of its elements homomorphically embeds into  $D_\rho$  for some configuration  $\rho$ . Also, there is an injective function  $\text{pred}_n$  from  $\text{configs}(A_{n+1})$  to  $\text{configs}(A_n)$  such that  $S \in \text{configs}(A_{n+1})$  can be generated using only atoms in  $\text{pred}_n(S)$ .*

**PROOF. Non-emptiness and finiteness.** Non-emptiness and finiteness of  $\text{configs}(A_n)$  follow from Lemma 28, since a finite bow tie has a finite non-zero amount of maximal paths.

**The elements of  $\text{configs}(A_n)$  embed into some  $D_\rho$ .** We then consider an element  $S$  of  $\text{configs}(A_n)$ , and  $(x_1, \dots, x_n)$  the path associated with it. Also, let  $A_n = q(x, w)$ . First, since  $(x_1, \dots, x_n)$  is a path in  $(\text{bowtie}(A_n), E_R)$ , for all  $i$ , the atom  $R(x_i, x_{i+1}, w)$  is in  $S$  for all  $i$ , and these are all the R-atoms in  $S$  by Lemma 28. Then, by Lemma 49, there is a unique atom  $a_i(x_i, w)$  for all  $i \leq n$ . In addition, since all the maximal paths in a bow tie go through its center, there is some  $p$  such that  $x_p = x$ . We thus define the configuration  $\rho = \langle n, (a_i)_{i \leq n}, p, q \rangle$ .

By mapping  $x_i$  to  $c_i$  for all  $i$ , and  $w$  to  $w_1$ , we get that  $S$  and  $D_\rho$  are isomorphic, except for the End-atoms. However, as per the last item of Lemma 28, the only position that can have  $\text{End}(x_i, w)$  is the end of a maximal path (since this kind of atoms can only be generated by rule  $R_{\text{End}}$  over non-brakes). Thus, the only possible End-atom in  $S$  is  $\text{End}(x_n, w)$ , which has a counterpart in  $D_\rho$ . Thus,  $S$  homomorphically embeds into  $D_\rho$ .

**Construction of  $\text{pred}_n$ .** We then construct the function  $\text{pred}_n$ . Let  $A_n = q(x, w)$  and  $A_{n+1} = q'(y, w')$ . First notice that the rule that generates  $A_{n+1}$  in the chase is among  $R_{\neg q_r}^\leftarrow, R_{\neg q_r}^\rightarrow, R_{q_r}^\leftarrow$  and  $R_{q_r}^\rightarrow$ . Then, there are some atoms  $F(x, z, w')$ , and  $R(z, y, w')$  or  $R(y, z, w')$  depending on the direction of the transition. We then assume that the transition is to the right, as the left case is analogous.

Consider a set  $S \in \text{configs}(A_{n+1})$ ,  $(y_1, \dots, y_k)$  the associated path, and  $\rho' = \langle k, (b_1, \dots, b_k), p', q' \rangle$  the associated configuration, as defined earlier in the proof. Then, let  $\text{pred}_n(S)$  be one of the sets in  $\text{configs}(A_n)$  with associated path  $(x_1, \dots, x_m)$  and configuration  $\rho = \langle m, (a_1, \dots, a_m), p, q \rangle$  such that there is an integer  $l$  such that

- for all  $i < k$ , we have  $F(x_{i+l}, y_i, w') \in F$ ;
- if  $\text{End}(x_k, w') \in S$ , then  $\text{End}(y_{k+l-1}) \in F$ , and otherwise  $F(x_{k+l}, y_k, w') \in F$ ;
- for all  $i \neq p' - 1$ ,  $b_i = a_{i+l}$ ;
- $p' + l = p + 1$ .

**The function  $\text{pred}_n$  is well-defined.** By definition of  $S$  and its associated path and configuration, there must be some atoms  $R(y_i, y_{i+1}, w')$  and  $b_i(y_i, w')$  for all  $i$ , with  $y = y_{p'}$ . By Lemma 28,  $R(y_{p'-1}, y_{p'})$  has been generated by rule  $R_{\neg q_r}^\rightarrow$  or  $R_{q_r}^\rightarrow$  along with  $A_{n+1}$ , and  $R(x_{k-1}, x_k)$  may have been generated by rule  $R_{\text{End}}$  or  $R_{C_R}$ . Other than that, all the R-atoms in the path  $y_1, \dots, y_k$  have been generated by rules  $R_{C_L}$  and  $R_{C_R}$ . We then show that there is a path  $x'_1, \dots, x'_{k'}$  such that for all  $i < k$ ,  $F(x'_i, y_i, w) \in F$ , for all  $i \neq p' - 1$ ,  $b_i(x'_i, w) \in F$ , and either  $\text{End}(x'_{k'-1}, w) \in F$  (and  $k' = k - 1$ ) or  $F(x'_k, y_k, w) \in F$  (and  $k' = k$ ), depending on whether  $\text{End}(y_k, w) \in S$  or not.

First, since the atom  $A_{n+1}$  has been generated by rule  $R_{\neg q_r}^\rightarrow$  or  $R_{q_r}^\rightarrow$ , there must be a term  $z$  and some atoms  $R(x, z, w)$ ,  $R(y_{p'-1}, y, w)$ ,  $F(x, y_{p'-1}, w)$  and  $F(z, y, w)$  in  $F$ . Thus, let  $x'_{p'-1} = x$  and  $x'_{p'} = z$ . We will then extend this path in both directions to construct  $x'_1, \dots, x'_{k'}$ .

If the path has been extended up to  $x'_{p'+i}$  for some  $i < k - 1 - p'$ , we then extend it to  $x'_{p'+i+1}$ . As mentioned before, the atom  $R(y_{p'+i}, y_{p'+i+1}, w)$  has been generated by rule  $R_{C_R}$  (since  $p' + i < k$ ). Thus, there must be some terms  $z, t$  and atoms  $R(z, t, w)$ ,  $F(z, y_{p'+i}, w)$ ,  $F(t, y_{p'+i+1}, w)$  and  $b_i(t, w)$  in  $F$ . By Lemma 48, we then have  $z = x'_{p'+i}$ , since both  $F(z, y_{p'+i}, w)$  and  $F(x'_{p'+i}, y_{p'+i}, w)$  are present

in  $F$ . We thus set  $t = x'_{p'+i+1}$ . The same reasoning lets us extend the path to  $x'_{p'-i-1}$  provided we have extended it to  $x'_{p'-i}$ , using the left copy rule instead of the right copy.

We now treat the case where  $i = k - 1 - p$ . If  $\text{End}(y_k, w) \notin S$ , then  $\text{R}(y_{k-1}, y_k, w)$  has been generated by rule  $R_{C_R}$ , so the same reasoning as before applies, and  $k' = k$ . Otherwise,  $\text{R}(y_{k-1}, y_k, w)$  has been introduced by rule  $R_{\text{End}}$ , meaning that there are some term  $z$  and atoms  $\text{End}(z, w)$  and  $\text{F}(z, y_{k-1})$  in  $F$ . Thus, since both  $\text{F}(z, y_{k-1})$  and  $\text{F}(x'_{k-1}, y_{k-1})$  are in  $F$ , by Lemma 48,  $z = y_{k-1}$ , and we have the atom  $\text{End}(y_{k-1}, w)$  in  $F$  as promised, and  $k' = k - 1$ .

We thus have a path  $x'_1, \dots, x'_k$  in  $\text{bowtie}(A_n)$  as described before. However, this path does not define an element of  $\text{configs}(A_n)$ , since it is not maximal. Thus, consider any maximal path  $x_1, \dots, x_m$  in  $\text{bowtie}(A_n)$  that extends  $x'_1, \dots, x'_k$ ,  $\text{pred}_n(S)$  the corresponding set in  $\text{configs}(A_n)$ ,  $\langle m, (a_1, \dots, a_m), p, q \rangle$  the corresponding configuration, and let  $l$  be the integer such that  $x_{l+1} = x'_1$ . Then, by definition of  $(x'_1, \dots, x'_{k'})$ , the first two points of the definition of  $\text{pred}_n(S)$  hold. Then, since  $A_n = q(x'_{p'-1}, w)$ , and  $x'_{p'-1} = x_{p'-1+l}$ , we have  $p = p' - 1 + l$ , so  $p + 1 = p' + l$ . In addition, since for all  $i \neq p' - 1$ ,  $b_i(x_{i+l}, w) \in F$ , we have  $a_{i+l} = b_i$ . Thus, there is indeed a set in  $\text{configs}(A_n)$  that fits the definition of  $\text{pred}_n(S)$ . Note however that this path is not necessarily unique, but we only need an injective function, so this is fine.

**The set  $\text{pred}_n(S)$  is enough to generate  $S$ .** First note that all the rule applications described earlier suffice to generate  $S$ . It is then enough to notice that all the atoms in the support of the mentioned triggers are present in  $\text{pred}_n(S)$ , or generated during the application of the previous triggers.

### Injectivity of $\text{pred}_n$ .

Consider two sets  $S_1$  with associated path  $(y_1, \dots, y_{k_1})$  and configuration  $\langle k_1, (b_1, \dots, b_{k_1}), p_1, q' \rangle$ , and  $S_2$  with associated path  $(y'_1, \dots, y'_{k_2})$  and configuration  $\langle k_2, (b'_1, \dots, b'_{k_2}), p_2, q' \rangle$ , such that  $\text{pred}_n(S_1) = \text{pred}_n(S_2) = S'$ , and  $S'$  has path  $(x_1, \dots, x_m)$  and configuration  $\langle m, (a_1, \dots, a_m), p, q \rangle$ . Thus, there must be some  $l_1$  and  $l_2$  such that:

- for all  $i$ ,  $\text{F}(x_{i+l_1}, y_i, w') \in F$  and  $\text{F}(x_{i+l_2}, y'_i, w') \in F$ ;
- for all  $i \neq p_1 - 1$ ,  $b_i = a_{i+l_1}$  and for all  $i \neq p_2 - 1$ ,  $b'_i = a_{i+l_2}$ ;
- $p_1 + l_1 = p + 1 = p_2 + l_2$ .

Assume w.l.o.g. that  $l_1 \geq l_2$ , and let  $d = l_1 - l_2$ . We then get that  $p_2 = p_1 + d$ , and  $b_i = a_{i+l_1} = a_{i+d+l_2} = b'_{i+d}$ , for all  $i \neq p_1$ . We then show that for all  $i$  such that  $1 \leq i \leq k_1$  and  $1 \leq i + d \leq k_2$ , we have  $y_i = y'_{i+d}$ . First, this is true for  $i = p_1$ , since  $y_{p_1} = y = y'_{p_2}$  (where  $A_{n+1} = q'(y, w')$ ) and  $p_2 = p_1 + d$ . This is also true for  $i = p_1 - 1$ , since by definition of a bow tie and Lemma 28, there is only one term  $t$  such that  $\text{R}(t, y_{p_1}, w') \in F$ . We then extend this to all  $i$  by induction.

Assume that  $1 \leq i + 1 \leq k_1$  and  $1 \leq i + 1 + d \leq k_2$ , and that  $y_i = y'_{i+d}$  for some  $i \geq p_1$  (the case where  $i \leq p_1 - 1$  is similar, using  $R_{C_L}$  instead of  $R_{C_R}$ ). We then show that  $y_{i+1} = y'_{i+1+d}$ . Both the atoms  $\text{R}(y_i, y_{i+1}, w')$  and  $\text{R}(y_i, y'_{i+1+d}, w')$  have been generated using rule  $R_{C_R}$ . We then show that the triggers generating these atoms are equal, so these atoms must be equal. The body of rule  $R_{C_R}$  is  $\{C_R(x', w'), \text{F}(x, x', w'), \text{R}(x, y, w), b_i(y, w), \text{Real}(x), \text{Real}(x'), \text{Real}(y)\}$ . To generate  $\text{R}(y_i, y_{i+1}, w')$ ,  $x'$  must be mapped to  $y_i$  (and  $w'$  to himself). Then, by Lemma 48, each term  $v$  can only have one term  $u$  such that  $\text{F}(u, v, w) \in F$ , so  $x$  is mapped to  $x_{i+l_1}$  and  $y$  to  $x_{i+1+l_1}$  (and  $w$  to himself), since  $\text{F}(x_{i+l_1}, y_i, w')$  and  $\text{F}(x_{i+1+l_1}, y_{i+1}, w')$ . However, we also have  $\text{F}(x_{i+l_1}, y'_{i+d}, w')$  and  $\text{F}(x_{i+1+l_1}, y'_{i+1+d}, w')$ , so the triggers generating  $\text{R}(y_i, y_{i+1}, w')$  and  $\text{R}(y_i(\rho, \vec{x}), y'_{i+1+d}, w')$  are equal, and  $y_{i+1} = y'_{i+1+d}$ .

Thus,  $l_1 = l_2$  and  $k_1 = k_2$ . Indeed, if  $l_1 > l_2$ , then we can extend  $y_1, \dots, y_{k_1}$  into a bigger path  $y'_1, \dots, y'_{l_1}, y_1, \dots, y_{k_1}$ , which contradicts its maximality. If  $k_1 \neq k_2$ , then we can extend the shortest path into the longest, also contradicting its maximality. Thus, both paths are equal, and  $S_1 = S_2$ . From this we deduce that  $\text{pred}_n$  is injective.  $\square$

**Lemma 31.** *The restricted chase does not terminate on  $\langle \Sigma_M, D_\rho \rangle$ .*

PROOF. First note that since for all  $n \geq N$ ,  $|\text{configs}(A_n)| = |\text{configs}(A_N)|$ ,  $\text{pred}_n$  is actually a bijection, since it is injective between sets of equal sizes. It thus has an inverse  $\text{pred}_n^{-1}$ . Thus, for all  $n \in \mathbb{N}$ , we define  $S_{n+1}$  as  $\text{pred}_{N+n}^{-1}(S_n)$ . Note that we picked  $S_0$  from  $\text{configs}(A_N)$  and that  $S_0$  homomorphically embeds into  $D_\rho$ .

We then inductively construct a sequence of derivations  $(\mathcal{D}_n)_{n \in \mathbb{N}}$  such that for all  $n \in \mathbb{N}$ , if  $S_n$  is over terms  $x_1, \dots, x_k, w$ , then

- $\mathcal{D}_{n+1}$  extends  $\mathcal{D}_n$ ;
- there is a homomorphism  $\pi_n$  from  $S_n$  to the result  $R_n$  of  $\mathcal{D}_n$ ;
- if  $F(\pi_n(x_i), y, \pi_n(w)) \in \mathcal{D}_n$  for some  $i$ , then  $y = \pi_n(w)$ ;
- $\text{Real}(\pi_n(w)) \notin \mathcal{D}_n$

First, as stated in Lemma 30,  $S_0$  embeds in  $D_\rho$ , so we let  $\mathcal{D}_0 = D_\rho$ , which does fulfill all the conditions above. Then, assume that we have constructed derivation  $\mathcal{D}_n$  as described. By Lemma 30 again, all the atoms in  $S_{n+1}$  can be generated using only atoms in  $\text{pred}_n(S_{n+1}) = S_n$ . We thus extend the derivation  $\mathcal{D}_n$  into a derivation  $\mathcal{D}'_{n+1}$  with the triggers needed to generate the atoms in  $S_{n+1}$ , composed with  $\pi_n$ . All these triggers are applicable since they all create atoms of the form  $F(\pi_n(x_i), y, \pi_n(w))$  and  $\text{Real}(y)$ , which are not in the database by the third and forth item. The homomorphism  $\pi_{n+1}$  is then defined naturally (the triggers that generate  $S_{n+1}$  from  $S_n$  were used here to generate new nulls, to which we can map nulls of  $S_{n+1}$ ). Then, if  $S_n$  contains an atom of the form  $q_r(x, w')$ , we add the trigger  $(R_{\text{Brake}}, \{w \rightarrow \pi_n(w)\})$  at the end of this new derivation, to construct  $\mathcal{D}_{n+1}$ . The first and second point then follow by design. The third point follows from the fact that the triggers that were used to generate  $S_{n+1}$  from  $S_n$  do not generate other  $F$ -atoms, and the last point from the fact that if  $q_r(x, w') \in S_n$ , then  $S_n$  and  $S_{n+1}$  use different brakes.

We now show that the derivation  $\mathcal{D} = \bigcup_n \mathcal{D}_n$  is fair. First, by Lemma 21, there are infinitely many  $q_r$ -atoms in  $(A_n)_{n \in \mathbb{N}}$ , and thus infinitely many  $n \in \mathbb{N}$  such that  $S_n$  contains a  $q_r$ -atom. Then, notice that whenever we encounter a  $q_r$ -atom in  $\mathcal{D}$ , we make the previous brake real, blocking any rule application involving the atoms containing it. Thus, for any trigger that is applicable at some step  $n$ , there is a step  $m$  at which the brake that appears in this trigger's support gets real, making this trigger obsolete. Thus,  $\mathcal{D}$  is fair, and the restricted chase does not terminate on  $\langle \Sigma_M, D_\rho \rangle$ .  $\square$