

# Checking Chase Termination over Ontologies of Existential Rules with Equality

**David Carral**

Institute for Theoretical Computer Science  
Technische Universität Dresden, Germany  
david.carral@tu-dresden.de

**Jacopo Urbani**

Department of Computer Science  
Vrije Universiteit Amsterdam, The Netherlands  
jacopo@cs.vu.nl

## Abstract

The chase is a sound and complete algorithm for conjunctive query answering over ontologies of existential rules with equality. To enable its effective use, we can apply acyclicity notions; that is, sufficient conditions that guarantee chase termination. Unfortunately, most of these notions have only been defined for existential rule sets without equality. A proposed solution to circumvent this issue is to treat equality as an ordinary predicate with an explicit axiomatisation. We empirically show that this solution is not efficient in practice and propose an alternative approach. More precisely, we show that, if the chase terminates for any equality axiomatisation of an ontology, then it terminates for the original ontology (which may contain equality). Therefore, one can apply existing acyclicity notions to check chase termination over an axiomatisation of an ontology and then use the original ontology for reasoning. We show that, in practice, doing so results in a more efficient reasoning procedure. Furthermore, we present equality model-faithful acyclicity, a general acyclicity notion that can be directly applied to ontologies with equality.

## Introduction

Answering conjunctive queries (CQs) over ontologies of existential rules with equality is a relevant reasoning task, which is undecidable (Beeri and Vardi 1981). One approach to solve it in some cases is to use the *chase* (Maier, Mendelzon, and Sagiv 1979)—a forward-chaining algorithm, which is sound and complete but may not terminate. Despite the fact that checking chase termination is undecidable (Gogacz and Marcinkowski 2014; Grahne and Onet 2018), we can apply *acyclicity notions*—sufficient conditions that guarantee termination—to enable the effective use of the chase for a large subset of real-world ontologies (Cuenca Grau et al. 2013; Carral, Dragoste, and Krötzsch 2017).

Acyclicity notions have been widely researched and many such criteria have been developed (Fagin et al. 2005; Marnette 2009; Baget, Mugnier, and Thomazo 2011; Krötzsch and Rudolph 2011; Carral, Feier, and Hitzler 2016; Carral, Dragoste, and Krötzsch 2017; Karimi, Zhang, and You 2018). Alas, some of the most general notions, such as

model-faithful acyclicity (MFA) (Cuenca Grau et al. 2013), are only defined for existential rule sets without equality. This restriction limits their usefulness, since equality is a prevalent feature (for instance, equality is used in  $\sim 34\%$  of the logical theories in the Oxford Ontology Repository<sup>1</sup>).

A proposed solution to enable the use of existing acyclicity notions over ontologies with equality is to treat equality as an ordinary predicate with an explicit axiomatisation (see Sections 2.1 and 5 of (Cuenca Grau et al. 2013)). Intuitively, an axiomatisation of a rule set  $\mathcal{R}$  is another rule set that does not contain equality and can be exploited to solve CQ answering over ontologies with the rule set  $\mathcal{R}$ . More precisely, using axiomatisations, we can solve CQ answering over an ontology  $\mathcal{O} = \langle \mathcal{R}, \mathcal{F} \rangle$ , where  $\mathcal{R}$  is an existential rule set (possibly containing equality) and  $\mathcal{F}$  is a fact set, by implementing the following step-by-step approach:

1. Compute some equality axiomatisation  $\mathcal{R}'$  of  $\mathcal{R}$ .
2. Verify whether  $\mathcal{R}'$  is acyclic with respect to some acyclicity notion (e.g., MFA). If this is the case, then the chase of  $\mathcal{R}'$  terminates; that is, for any given fact set  $\mathcal{F}'$ , the chase terminates on input  $\langle \mathcal{R}', \mathcal{F}' \rangle$ .
3. Apply the chase on  $\langle \mathcal{R}', \mathcal{F} \rangle$ , and use the resulting output fact set to solve CQ answering over  $\mathcal{O}$ .

Note that, since  $\mathcal{R}'$  is an axiomatisation of  $\mathcal{R}$ ,  $\mathcal{R}'$  is equality-free and hence, we can check if this rule set is MFA in Step 2.

The application of the above step-by-step approach to real-world ontologies is somewhat problematic. For instance, the use of the *standard axiomatisation* in Step 1 often causes the MFA check applied in Step 2 to fail (Cuenca Grau et al. 2013). As shown in this paper, the use of other axiomatisation techniques in Step 1, such as *singularisation* (Marnette 2009), fixes this issue to a large extent. Unfortunately, computing the chase of an ontology that features some singularisation of a rule set—as required in Step 3—is not efficient in practice. The fact that the use of axiomatisations leads to poor performance has previously been shown for the standard axiomatisation (Motik et al. 2015); we show that it is also the case when singularisation is applied.

To address these issues, we show that, if the chase of any equality axiomatisation of  $\mathcal{R}$  terminates, then so does the

<sup>1</sup><https://www.cs.ox.ac.uk/isg/ontologies/>

chase of the rule set  $\mathcal{R}$ . Hence, we can replace Step 3 in the above step-by-step procedure with the following alternative:

4. Compute the chase on input  $\mathcal{O}$  and use the resulting output fact set to solve CQ answering over  $\mathcal{O}$ .

Implementing Step 4 instead of Step 3 enables the use of *rewriting* to deal with equality when computing the chase; a technique that has already been proven more efficient than the use of axiomatisations in practice (Motik et al. 2015).

Still, there is yet another practical problem. Namely, checking if the singularisation of a rule set is MFA—as required in Step 2—is somewhat inefficient for many real-world rule sets. To solve this issue, we present *equality model-faithful acyclicity* (EMFA), a very general acyclicity notion based on MFA that can be directly applied to rule sets with equality. By applying this notion directly, we altogether remove the need for using equality axiomatisations.

Our contributions are as follows: we provably show that, if the chase of the standard axiomatisation or any singularisation of a rule set  $\mathcal{R}$  terminates, then so does the chase of  $\mathcal{R}$ ; we show that the converses of the previous implications do not hold; and we define EMFA, we study the complexity of checking EMFA membership and reasoning over EMFA ontologies, and we compare the expressivity of this notion with that of MFA. Moreover, we empirically show that

- I. computing the chase of an ontology featuring some singularisation of a rule set is not efficient in practice,
- II. the standard equality axiomatisation of a large subset of real-world rule sets is not MFA,
- III. EMFA is as general as “MFA plus singularisation”, and
- IV. checking if a rule set is EMFA is more efficient than checking if it is “MFA plus singularisation”.

Our treatment is fully self-contained, but a technical report with further information can be consulted if desired (Carral and Urbani 2019).

## Preliminaries

### Syntax and Semantics

We consider a signature based on mutually disjoint, finite sets of *constants*  $\mathbf{C}$ , *function symbols*  $\mathbf{F}$ , *variables*  $\mathbf{V}$ , and *predicates*  $\mathbf{P}$ . Every entity  $e \in \mathbf{F} \cup \mathbf{P}$  is associated with some arity  $ar(e) \geq 1$ . The set  $\mathbf{P}$  includes the special binary predicate  $\approx$ , which is referred to as the *equality predicate* or simply as *equality*. The set of *terms*  $\mathbf{T}$  is the minimal superset of  $\mathbf{C}$  and  $\mathbf{V}$  such that, for all  $f \in \mathbf{F}$  and all  $t_1, \dots, t_{ar(f)} \in \mathbf{T}$ , we have that  $f(t_1, \dots, t_{ar(f)}) \in \mathbf{T}$ . The set of *ground terms*  $\mathbf{GT}$  is the set of all terms without syntactic occurrences of a variable. For a term  $t$ , let  $dep(t) = 1$  if  $t \in \mathbf{C} \cup \mathbf{V}$ , and  $dep(t) = \max(dep(t_1), \dots, dep(t_n)) + 1$  if  $t$  is of the form  $f(t_1, \dots, t_n)$ . Given an entity set  $\mathbf{E}$  and a formula or set thereof  $\mathcal{U}$ , we write  $\mathbf{E}_{\mathcal{U}}$  to denote the set that contains all of the elements in  $\mathbf{E}$  that occur in  $\mathcal{U}$ . We abbreviate lists of terms  $t_1, \dots, t_n$  as  $\vec{t}$  and treat these as sets when order is irrelevant. An *atom* is a formula  $P(\vec{t})$  with  $P \in \mathbf{P}$ ,  $\vec{t} \in \mathbf{T}$ , and  $ar(P) = |\vec{t}|$ . As customary, we write  $t \approx u$  instead of  $\approx(t, u)$  to denote atoms defined over equality.

A *fact* is an atom  $P(\vec{t})$  with  $\vec{t} \in \mathbf{C}$ . For a formula  $v$  and a list of variables  $\vec{x}$ , we write  $v[\vec{x}]$  to indicate that  $\vec{x}$  is the set of all free variables occurring in  $v$  (i.e., the set of all variables that are not quantified in  $v$ ). An (*existential*) *rule* is a function- and constant-free first-order logic (FOL) formula of either of the following forms.

$$\forall \vec{x}, \vec{y}. (\beta[\vec{x}, \vec{y}] \rightarrow \exists \vec{w}. \eta[\vec{x}, \vec{w}]) \quad (1)$$

$$\forall \vec{x}. (\beta[\vec{x}] \rightarrow x \approx y) \quad (2)$$

In the above,  $\vec{x}$ ,  $\vec{y}$ , and  $\vec{w}$  are pairwise disjoint lists of variables;  $\vec{x}$  is non-empty;  $\beta$  and  $\eta$  are non-empty conjunctions of atoms without equality; and  $x, y \in \vec{x}$ . The *body* (resp. *head*) of a rule is the conjunction of atoms to the left (resp. right) of its implication symbol. We omit universal quantifiers when writing rules and treat conjunctions of atoms, such as  $\beta$  and  $\eta$  above, as atom sets. We refer to rules of the form (1) and (2) as *tuple generating dependencies* (TGDs) and *equality generating dependencies* (EGDs), respectively.

A *boolean conjunctive query* (BCQ) is a function-free FOL formula  $\gamma = \exists \vec{x}. \beta[\vec{x}]$  with  $\beta$  a non-empty conjunction of atoms that, without loss of generality, does not contain constants or equality. We refer to  $\beta$  as the *body* of  $\gamma$ . Since CQ answering can be reduced to BCQ entailment, we confine our attention to the latter without loss of generality.

We consider finite rule sets  $\mathcal{R}$ , where we assume without loss of generality that existentially quantified variables do not reoccur across different rules ( $\dagger$ ). An *ontology*  $\mathcal{O}$  is a tuple  $\langle \mathcal{R}, \mathcal{F} \rangle$  with  $\mathcal{R}$  a rule set and  $\mathcal{F}$  a fact set. Without loss of generality, we assume that, for an ontology  $\langle \mathcal{R}, \mathcal{F} \rangle$ , the set  $\mathcal{F}$  is equality-free and  $\mathbf{P}_{\mathcal{F}} \subseteq \mathbf{P}_{\mathcal{R}}$ .

For an ontology  $\mathcal{O}$  and a BCQ  $\gamma$ , we write  $\mathcal{O} \models \gamma$  to indicate that  $\mathcal{O}$  entails  $\gamma$  under FOL semantics. That is, to indicate that  $\bigwedge_{\rho \in \mathcal{R}} \rho \wedge \bigwedge_{\varphi \in \mathcal{F}} \varphi$  entails  $\gamma$ .

### The Non-Oblivious Chase Algorithm

We present the (*non-oblivious*) *chase* (Fagin et al. 2005)—a chase variant that expands existential quantifiers only if necessary, and merges terms to comply with the semantics of equality. Unlike Fagin et al. (2005), we do not contemplate the *unique name assumption* and introduce Skolem functional terms instead of “unlabelled” nulls to satisfy existential restrictions. The use of “labelled” Skolem terms simplifies some the formal arguments presented in the following sections (e.g., see the proofs of Theorems 1, 3, and 5).

**Definition 1** (Skolemisation). *The skolemisation  $sk(\rho)$  of a TGD  $\rho$  of the form (1) is the formula  $\beta \rightarrow sk(\eta)$  where  $sk(\eta)$  is the conjunction of atoms obtained from  $\eta$  by replacing each variable  $w \in \vec{w}$  by the Skolem term  $f_w(\vec{x})$ , with  $f_w$  a fresh Skolem function symbol of arity  $|\vec{x}|$  unique for  $w$ .*

By ( $\dagger$ ), an existentially quantified variable  $w$  in some rule set  $\mathcal{R}$  may occur in at most one rule  $\rho \in \mathcal{R}$  and hence, the function symbol  $f_w$  is unique for both  $w$  and  $\rho$  in  $sk(\mathcal{R})$ .

We write  $[t_1/u_1, \dots, t_n/u_n]$  to denote the function over  $\mathbf{T}$  mapping  $t_i$  to  $u_i$  for all  $i \in \{1, \dots, n\}$ . Given such a function  $\sigma$  and a formula  $v$ , let  $v\sigma$  be the formula that results from replacing every occurrence of a term  $t$  as a predicate argument in an atom in  $v$  with  $\sigma(t)$  if the latter is defined

(e.g.,  $P(t, f(t))[t/u] = P(u, f(t))$ ). A *rewriting* (resp. *substitution*) is a functions from  $\mathbf{GT}$  (resp.  $\mathbf{V}$ ) to  $\mathbf{GT}$ .

To guide the renaming that results from applying EGDs, we introduce a strict total order  $\prec$  defined over the set of terms such that  $t \prec u$  for all  $t, u \in \mathbf{T}$  with  $\text{dep}(t) < \text{dep}(u)$ .

**Definition 2** (Rule Applicability). *Consider a rule  $\rho$ , a substitution  $\sigma$ , and an atom set  $\mathcal{A}$ . Then, the tuple  $\langle \rho, \sigma \rangle$  is applicable to  $\mathcal{A}$  if all of the following hold.*

- $\sigma$  is defined and undefined for all of the universally and existentially quantified variables in  $\rho$ , respectively.
- $\beta\sigma \subseteq \mathcal{A}$  with  $\beta$  the body of  $\rho$ .
- If  $\rho$  is of the form (1), then  $\eta\tau \not\subseteq \mathcal{A}$  for all  $\tau \supseteq \sigma$ .
- If  $\rho$  is of the form (2), then  $\sigma(x) \neq \sigma(y)$ .

If  $\langle \rho, \sigma \rangle$  is applicable to  $\mathcal{A}$ , then the application of  $\langle \rho, \sigma \rangle$  on  $\mathcal{A}$ , denoted with  $\mathcal{A}_{\rho, \sigma}$ , is the atom set defined as follows.

- If  $\rho$  is of the form (1), then  $\mathcal{A}_{\rho, \sigma} = \mathcal{A} \cup \text{sk}_\sigma(\eta)$  where  $\text{sk}_\sigma(\eta)$  is the formula that results from replacing every syntactic occurrence of a variable  $x$  in  $\text{sk}(\eta)$  with  $\sigma(x)$ .
- If  $\rho$  is of the form (2); then  $\mathcal{A}_{\rho, \sigma} = \mathcal{A}[\sigma(y)/\sigma(x)]$  if  $\sigma(x) \prec \sigma(y)$ , and  $\mathcal{A}_{\rho, \sigma} = \mathcal{A}[\sigma(x)/\sigma(y)]$  otherwise.

The case distinction in the last item in the previous definition ensures that we merge “deeper” terms into “shallower” ones when applying EGDs. This strategy simplifies some of our formal arguments (e.g., see the proof of Theorem 6), but it is not necessary to define a correct procedure.

**Definition 3** (The Chase Algorithm). *A chase sequence of an ontology  $\mathcal{O} = \langle \mathcal{R}, \mathcal{F} \rangle$  is a (possibly finite) sequence  $\mathcal{A}_0, \mathcal{A}_1, \dots$  of atom sets such that the following hold.*

- $\mathcal{A}_0 = \mathcal{F}$
- For all  $i \geq 1$ , there is some rule  $\rho \in \mathcal{R}$  and some substitution  $\sigma$  such that  $\mathcal{A}_i$  is the application of  $\langle \rho, \sigma \rangle$  on  $\mathcal{A}_{i-1}$ .
- For all rules  $\rho \in \mathcal{R}$  and substitutions  $\sigma$ , there is a  $k \geq 0$  such that, for all  $i \geq k$ , the tuple  $\langle \rho, \sigma \rangle$  is not applicable to  $\mathcal{A}_i$  (i.e., fairness).

A chase of  $\mathcal{O}$  is the atom set that results from taking the union of all atom sets in some chase sequence of  $\mathcal{O}$ .

The chase of an ontology  $\mathcal{O}$  terminates if all of the chases of  $\mathcal{O}$  are finite; that is, if all chase sequences of  $\mathcal{O}$  are finite. The chase of a rule set  $\mathcal{R}$  terminates if, for all fact sets  $\mathcal{F}$ , the chase of  $\langle \mathcal{R}, \mathcal{F} \rangle$  terminates. An atom set is a chase step of  $\mathcal{O}$  if it occurs in a chase sequence of  $\mathcal{O}$ .

An ontology may admit infinitely many different chases. Nevertheless, it is well-known that any chase of an ontology is a universal model for this logical theory; i.e., a model that can be homomorphically embedded into any other model. Therefore, this structure can be directly used to solve BCQ entailment (Deutsch, Nash, and Remmel 2008).

**Fact 1.** *An ontology entails a BCQ iff any chase of this ontology entails this BCQ.*

This result holds because, by our definition, rules and BCQs do not contain constants. If we want to lift this restriction, we would need to modify the definition of the chase so it takes into account the rewriting of terms that occurs when EGDs are applied as it is done in (Motik et al. 2015).

## Handling Equality via Axiomatisation

We present two well-known techniques to compute equality axiomatisations; namely, the *standard axiomatisation* (see Section 2.1 in (Cuenca Grau et al. 2013)) and *singularisation* (see Section 5 in (Marnette 2009)). In the definition of these, we replace  $\approx$  with the fresh predicate  $Eq$  to clarify that these two predicates are to be treated differently.

**Definition 4.** *The standard axiomatisation of a rule set  $\mathcal{R}$ , denoted with  $\text{St}(\mathcal{R})$ , is the TGD set that contains all of the TGDs in  $\mathcal{R}$ , the TGD  $\beta \rightarrow Eq(x, y)$  for every EGD  $\beta \rightarrow x \approx y \in \mathcal{R}$ , an instance of (3) for every  $P \in \mathbf{P}_{\mathcal{R}} \setminus \{\approx\}$ , the TGDs (4) and (5), and an instance of (6) for every  $P \in \mathbf{P}_{\mathcal{R}} \setminus \{\approx\}$  and every  $i \in \{1, \dots, \text{ar}(P)\}$ .*

$$P(\vec{x}_{\text{ar}(P)}) \rightarrow \bigwedge_{1 \leq i \leq \text{ar}(P)} Eq(x_i, x_i) \quad (3)$$

$$Eq(x, y) \rightarrow Eq(y, x) \quad (4)$$

$$Eq(x, y) \wedge Eq(y, z) \rightarrow Eq(x, z) \quad (5)$$

$$P(\vec{x}_{\text{ar}(P)}) \wedge Eq(x_i, x'_i) \rightarrow (P(\vec{x}_{\text{ar}(P)})[x_i/x'_i]) \quad (6)$$

In the above,  $\vec{x}_k = x_1, \dots, x_k$  for all  $k \geq 1$ .

The standard axiomatisation of an ontology  $\mathcal{O} = \langle \mathcal{R}, \mathcal{F} \rangle$  is the ontology  $\text{St}(\mathcal{O}) = \langle \text{St}(\mathcal{R}), \mathcal{F} \rangle$ .

The standard axiomatisation of an ontology  $\mathcal{O}$  can be directly used to solve BCQ entailment over  $\mathcal{O}$ .

**Fact 2.** *For an ontology  $\mathcal{O}$  and a BCQ  $\gamma$  that does not contain the predicate  $Eq$ , we have  $\mathcal{O} \models \gamma$  iff  $\text{St}(\mathcal{O}) \models \gamma$ .*

By applying this result, we can decide BCQ entailment over  $\mathcal{O}$  with algorithms (or implementations!) that do not require a special treatment to deal with equality.

**Definition 5.** *Consider a conjunction of atoms  $\beta$ . A singularisation of  $\beta$  is a conjunction of atoms that is constructed by applying the following step-by-step transformation to  $\beta$ .*

1. For every  $x \in \mathbf{V}_\beta$ , choose some  $k_x \in \{1, \dots, n\}$  with  $n$  the number of occurrences of  $x$  in  $\beta$ .
2. Simultaneously, replace the  $i$ -th occurrence of every variable  $x$  in  $\beta$  with a fresh variable  $x_i$  if  $i \neq k_x$ .
3. Add the atom  $Eq(x, x_i)$  for every fresh variable  $x_i$  introduced in the above step.

Let  $\text{Sg}(\beta)$  be the set of all singularisations of  $\beta$ .

The singularisations of a TGD  $\rho = \beta \rightarrow \exists \vec{w}. \eta$ , an EGD  $\mu = \beta \rightarrow x \approx y$ , a rule set  $\mathcal{R}$ , an ontology  $\mathcal{O} = \langle \mathcal{R}, \mathcal{F} \rangle$ , and a BCQ  $\gamma = \exists \vec{w}. \beta$  are defined as follows, respectively.

- $\text{Sg}(\rho) = \{\beta' \rightarrow \exists \vec{w}. \eta \mid \beta' \in \text{Sg}(\beta)\}$
- $\text{Sg}(\mu) = \{\beta' \rightarrow Eq(x, y) \mid \beta' \in \text{Sg}(\beta)\}$
- Let  $\text{Sg}(\mathcal{R})$  be the set of all TGD sets that contain an instance of (3) for every  $P \in \mathbf{P}_{\mathcal{R}} \setminus \{\approx\}$ , the TGDs (4) and (5), and (exactly) one TGD in  $\text{Sg}(\rho)$  for each  $\rho \in \mathcal{R}$ .
- $\text{Sg}(\mathcal{O}) = \{\langle \mathcal{R}', \mathcal{F} \rangle \mid \mathcal{R}' \in \text{Sg}(\mathcal{R})\}$
- $\text{Sg}(\gamma) = \{\exists \vec{x}. \beta' \mid \beta'[\vec{x}] \in \text{Sg}(\beta)\}$

The use of singularisation may yield many axiomatisations for a single ontology  $\mathcal{O}$ , any of which can be used to solve BCQ over  $\mathcal{O}$  (Marnette 2009).

**Fact 3.** For an ontology  $\mathcal{O}$  and a BCQ  $\gamma$  that does not contain the predicate  $Eq$ , we have  $\mathcal{O} \models \gamma$  iff  $\mathcal{O}' \models \gamma'$  for any  $\mathcal{O}' \in \text{Sg}(\mathcal{O})$  and  $\gamma' \in \text{Sg}(\gamma)$ .

Even though the use of equality axiomatisations does not affect the results of BCQ entailment (see Facts 2 and 3), it may influence chase termination. For instance, in the following section, we show examples of rule sets  $\mathcal{R}$  such that  $\mathcal{R}$  is terminating with respect to the chase, and  $\text{Sg}(\mathcal{R})$  or some set in  $\text{St}(\mathcal{R})$  are not (see Theorems 2 and 4).

## Chase Termination and Axiomatisations

In this section, we show that if the chase terminates for the standard axiomatisation or any singularisation of a given rule set  $\mathcal{R}$ , then the chase of  $\mathcal{R}$  also terminates (see Theorems 1 and 3). Moreover, we provide some results stating that the converses of these implications do not hold (see Theorems 2 and 4). Prior to showing these results, we introduce some preliminary definitions.

An atom set  $\mathcal{A}$  satisfies a rule  $\rho$  iff there is no substitution  $\sigma$  such that  $\langle \rho, \sigma \rangle$  is applicable to  $\mathcal{A}$ .

**Definition 6.** An atom set  $\mathcal{A}$  is *Eq-complete* iff it satisfies the rules (4) and (5) introduced in Definition 4, and  $Eq(t, t) \in \mathcal{A}$  for every  $t \in \mathbf{T}_{\mathcal{A}}$ .

**Definition 7.** Consider an Eq-complete atom set  $\mathcal{A}$ . Then, let  $\pi_{\mathcal{A}}$  be the rewriting that maps every  $t \in \mathbf{T}_{\mathcal{A}}$  into the term  $\pi_{\mathcal{A}}(t) \in \mathbf{T}_{\mathcal{A}}$  such that

- $Eq(t, \pi_{\mathcal{A}}(t)) \in \mathcal{A}$  and
- for all  $u \in \mathbf{T}_{\mathcal{A}}$  with  $u \neq \pi_{\mathcal{A}}(t)$  and  $Eq(t, u) \in \mathcal{A}$ , we have that  $\pi_{\mathcal{A}}(t) \prec u$ .

Furthermore, let  $[\mathcal{A}] = \mathcal{A}\pi_{\mathcal{A}} \setminus \{Eq(t, u) \mid t, u \in \mathbf{T}\}$ .

**Theorem 1.** The chase of a rule set  $\mathcal{R}$  terminates if the chase of the TGD set  $\text{St}(\mathcal{R})$  terminates.

*Proof Sketch.* Consider some fact set  $\mathcal{F}$  and the ontology  $\mathcal{O} = \langle \mathcal{R}, \mathcal{F} \rangle$ . Theorem 1 follows if chase termination of  $\text{St}(\mathcal{O}) = \langle \text{St}(\mathcal{R}), \mathcal{F} \rangle$  implies chase termination of  $\mathcal{O}$ . Note that, if  $\text{St}(\mathcal{O})$  terminates, then the number of chase steps of this ontology is necessarily finite. In turn, this claim follows if, for every chase step  $\mathcal{A}$  of  $\mathcal{O}$ , there is some Eq-complete chase step  $\mathcal{B}$  of  $\text{St}(\mathcal{O})$  such that  $[\mathcal{B}] = \mathcal{A}$ . This implication can be proven via induction.  $\square$

A proof of Theorem 1 for the standard chase was presented in Theorem 4.2 (Grahne and Onet 2018). For completeness, we present a proof for the non-oblivious chase with Skolem terms (which is the chase variant considered in this paper) in the extended version of this paper.

We can show with a counter-example that the converse of Theorem 1 does not hold.

**Theorem 2.** Even if the chase of a rule set  $\mathcal{R}$  terminates, the chase of the TGD set  $\text{St}(\mathcal{R})$  may not.

*Proof.* The chase of the rule set  $\mathcal{R} = \{(7), (8)\}$  does terminate whereas the chase of  $\text{St}(\mathcal{R})$  does not.

$$A(x) \rightarrow \exists w.R(x, w) \wedge B(w) \quad (7)$$

$$R(x, y) \wedge R(x, z) \rightarrow y \approx z \quad (8)$$

For instance, the chase of the ontology  $\text{St}(\mathcal{O}) = \langle \text{St}(\mathcal{R}), \mathcal{F} \rangle$  where  $\mathcal{F}$  is the fact set  $\{A(a), R(a, a)\}$  admits infinite chase sequences. Namely,  $\text{St}(\mathcal{O})$  admits (fair and infinite) chase sequences that contain all of the atom sets in the sequence  $\mathcal{F}_1 \cup \mathcal{F}, \mathcal{G}_1, \mathcal{F}_2, \mathcal{G}_2, \mathcal{F}_3, \mathcal{G}_3, \dots$  where  $\mathcal{G}_i$  is the set of all atoms that can be defined using the predicates  $A, B, R$ , and  $Eq$ , and the terms in  $\{f_w^0(a), \dots, f_w^{i-1}(a)\}$  for all  $i \geq 1$ ;

$$\mathcal{F}_1 = \{R(a, f_w^1(a)), B(f_w^1(a)), Eq(a, f_w^1(a)), A(f_w^1(a))\};$$

$$\mathcal{F}_i = \{R(f_w^{i-1}(a), f_w^i(a)), B(f_w^i(a)), Eq(f_w^{i-1}(a), a), \\ R(a, f_w^i(a)), Eq(a, f_w^i(a)), A(f_w^i(a))\} \cup \mathcal{G}_{i-1}$$

for all  $i \geq 2$ ; and  $f_w^0(a) = a, f_w^1(a) = f_w(a), f_w^2(a) = f_w(f_w(a))$ , and so on.  $\square$

The fact that the chase does not terminate for the standard axiomatisation of a rule set as simple as the one described above illustrates why the use of this axiomatisation technique often causes the acyclicity tests to fail. We empirically verify this insight in the evaluation section.

As per Definition 5, a rule set may admit many different singularisations. If the chase of at least one of these terminates, then so does the chase of the rule set itself.

**Theorem 3.** The chase of a rule set  $\mathcal{R}$  terminates if the chase of some TGD set in  $\text{Sg}(\mathcal{R})$  terminates.

*Proof Sketch.* Assume that there is some rule set  $\mathcal{R}' \in \text{Sg}(\mathcal{R})$  for which the chase terminates. Then, let  $\mathcal{F}$  be some fact set, let  $\mathcal{O} = \langle \mathcal{R}, \mathcal{F} \rangle$  and  $\mathcal{O}' = \langle \mathcal{R}', \mathcal{F} \rangle$ , and let  $\mathcal{A}_0, \mathcal{A}_1, \dots$  be a chase sequence of  $\mathcal{O}$ . We can show via induction that there is a sequence of atom sets  $\mathcal{B}_0, \mathcal{B}_1, \dots$  and a sequence of rewritings  $\pi_0, \pi_1, \dots$  such that, for all  $i \geq 0$ ,

1.  $\mathcal{B}_i$  is a chase step of  $\mathcal{O}'$ ,
2.  $\mathcal{B}_i\pi_i \setminus \{Eq(t, u) \mid t, u \in \mathbf{T}\} = \mathcal{A}_i$ , and
3.  $Eq(t, u) \in \mathcal{B}_i$  for all  $t, u \in \mathbf{T}$  in  $\mathcal{B}_i$  with  $\pi_i(t) = \pi_i(u)$ .

Since (1), we conclude that the sequence  $\mathcal{B}_0, \mathcal{B}_1, \dots$  is finite and hence, the sequence  $\mathcal{A}_0, \mathcal{A}_1, \dots$  is also finite by (2). In turn, this implies the chase of  $\mathcal{R}$  terminates. Item (3) is an auxiliary claim used to structure the induction.  $\square$

We can show with a counter-example that the converse of Theorem 3 does not hold.

**Theorem 4.** Even if the chase of a rule set  $\mathcal{R}$  terminates, the chase of every TGD set in  $\text{Sg}(\mathcal{R})$  may not.

*Proof.* The chase of the rule  $\mathcal{R} = \{(9), (10), (11)\}$  does terminate whereas the chase every  $\mathcal{R}' \in \text{Sg}(\mathcal{R})$  does not.

$$B(x) \wedge C(x) \rightarrow \exists y.R(x, y) \wedge B(y) \quad (9)$$

$$B(x) \wedge C(x) \rightarrow \exists z.R(x, z) \wedge C(z) \quad (10)$$

$$R(x, y) \rightarrow x \approx y \quad (11)$$

The chase of rule set  $\mathcal{R}$  does terminate as we have that, for any given fact set  $\mathcal{F}$ , the chase of  $\langle \mathcal{R}, \mathcal{F} \rangle$  may only contain terms of depth at most 1.

None of the four different rule sets in  $\text{Sg}(\mathcal{R})$  does terminate. More precisely, for some  $\mathcal{R}' \in \text{Sg}(\mathcal{R})$ , the chase of  $\langle \mathcal{R}', \{B(a), C(a)\} \rangle$  does not terminate.  $\square$

## Extending MFA with Equality

In this section, we present *equality model-faithful acyclicity* (EMFA), an acyclicity notion based on model-faithful acyclicity (MFA) (Cuenca Grau et al. 2013) that can be directly applied to existential rule sets with equality.

**Definition 8.** A term  $t$  is cyclic if it is of the form  $f(\vec{u})$  and the function symbol  $f$  occurs in some term in  $\vec{u}$ .

**Definition 9 (MFA/EMFA).** For a rule set  $\mathcal{R}$ , let  $E_{\mathcal{R}}$  be the minimal atom set that satisfies the following.

- $\mathcal{F}_{\mathcal{R}}^* \subseteq E_{\mathcal{R}}$  with  $\mathcal{F}_{\mathcal{R}}^*$  the critical instance for  $\mathcal{R}$ ; that is, the set of all facts that can be defined using the predicates in  $\mathcal{P}_{\mathcal{R}}$  and the special constant  $\star$ .
- If there is a TGD  $\beta \rightarrow \exists \vec{u}. \eta \in \mathcal{R}$  and a substitution  $\sigma$  with  $\beta\sigma \subseteq E_{\mathcal{R}}$ , then  $sk_{\sigma}(\eta) \subseteq E_{\mathcal{R}}$ .
- If there is an EGD  $\beta \rightarrow x \approx y \in \mathcal{R}$  and a substitution  $\sigma$  with  $\beta\sigma \subseteq E_{\mathcal{R}}$ ; then  $E_{\mathcal{R}}[\sigma(y)/\sigma(x)] \subseteq E_{\mathcal{R}}$  if  $dep(\sigma(x)) \leq dep(\sigma(y))$ , and  $E_{\mathcal{R}}[\sigma(x)/\sigma(y)] \subseteq E_{\mathcal{R}}$  if  $dep(\sigma(y)) \leq dep(\sigma(x))$ .

A rule set  $\mathcal{R}$  is EMFA if  $E_{\mathcal{R}}$  does not contain any cyclic terms. A TGD set is MFA if it is EMFA. An ontology  $\langle \mathcal{R}, \mathcal{F} \rangle$  is MFA (resp. EMFA) if  $\mathcal{R}$  is MFA (resp. EMFA).

Even though our definition of MFA may seem different from its original definition, these two coincide. Namely, a TGD set  $\mathcal{R}$  is MFA with respect to Definition 3 in (Cuenca Grau et al. 2013) iff it is MFA with respect to the above definition. This equivalence readily follows from Proposition 5 in (Cuenca Grau et al. 2013).

In the remainder of the section, we show some results about the complexity of checking EMFA membership and of reasoning with EMFA rule sets. We apply the following preliminary lemma in the proofs of some of these results.

**Lemma 1.** For a rule set  $\mathcal{R}$ , the number of atoms without cyclic terms that can be defined using (a) a finite set of constants  $C$ , (b) the (Skolem) function symbols in  $\mathbf{F}_{sk(\mathcal{R})}$ , and (c) the predicates in  $\mathcal{P}_{\mathcal{R}}$  is doubly exponential in  $\mathcal{R}$ .

*Proof.* Firstly, we determine that the size of the set of all non-cyclic terms  $T$  that can be defined using the symbols in  $C$  and  $\mathbf{F}_{sk(\mathcal{R})}$  is doubly exponential in  $\mathcal{R}$ . Let  $n = |\mathbf{F}_{sk(\mathcal{R})}|$  and let  $m$  be the maximal arity of a function in  $\mathbf{F}_{sk(\mathcal{R})}$ . By assumption ( $\dagger$ ) in the second section, we have that  $|\mathbf{F}_{sk(\mathcal{R})}| \leq |\mathbf{V}_{\mathcal{R}}|$ . For every  $t \in \mathbf{T}$ , let  $r_t = t$  if  $t \in C$  and  $r_t = f$  if  $t$  is of the form  $f(u_1, \dots, u_n)$ . Moreover, let  $G_t = (V_t, E_t)$  be a directed tree defined as follows: if  $t \in C$ , then  $G_t = (\{t\}, \emptyset)$ ; if  $t$  is of the form  $f(u_1, \dots, u_n)$ , then

$$G_t = (\{f\} \cup \bigcup_{1 \leq i \leq n} V_{u_i}, \{\langle f, r_{u_i} \rangle \mid 1 \leq i \leq n\} \cup \bigcup_{1 \leq i \leq n} E_{u_i}).$$

For every  $t \in \mathbf{T}$ , the maximal length of a path in  $G_t$  is at most  $n$ , since a greater length implies the repetition of a function symbol, which in turn would imply that  $t$  is cyclic. Hence,  $G_t$  is of depth at most  $n$ , fan-out at most  $m$ , and with leafs labelled with constants in  $C$ . Such trees have at most  $m^n$  leafs and at most  $n \cdot m^n$  nodes in total. As each node is assigned a constant or function symbol, there are at most  $(|C|+n)^{(n \cdot m^n)}$  such trees and hence, non-cyclic terms, overall. Therefore,  $|T|$  is doubly exponential in  $\mathcal{R}$ .

Secondly, we determine the number of atoms without cyclic terms. Let  $k = |\mathcal{P}_{\mathcal{R}}|$  and let  $\ell$  be the maximal arity of a predicate in  $\mathcal{R}$ . Then, the number of atoms without cyclic terms is at most  $k|T|^\ell = k(|C|+n)^{\ell \cdot n \cdot m^n}$ .  $\square$

Despite the fact that EMFA can be applied to rule sets with equality, checking membership with respect to this notion is not harder than deciding MFA membership.

**Theorem 5.** Deciding whether a rule set  $\mathcal{R}$  is EMFA is 2EXPTIME-complete.

*Proof.* (Membership) To decide if  $\mathcal{R}$  is EMFA, it suffices to compute the atom set  $E_{\mathcal{R}}$  up until the occurrence of an atom containing a cyclic term. By Lemma 1, we may have to compute  $E_{\mathcal{R}}$  up until it contains doubly exponential many atoms. To decide whether to include each of these atoms we have to determine whether there is a pair  $\langle \rho, \sigma \rangle$  with  $\rho \in \mathcal{R}$  that is applicable to some subset of  $E_{\mathcal{R}}$ . Checking if this is the case can be done in 2EXPTIME and hence, the EMFA membership check can be performed in 2EXPTIME.

(Hardness) An equality-free rule set (i.e., a TGD set) is EMFA iff it is MFA. Hence, an algorithm that solves EMFA membership can be used to decide MFA membership. Therefore, the hardness of the EMFA check follows from the hardness of the MFA membership check, which was shown to be 2EXPTIME-hard by Cuenca Grau et al. (2013) (see Theorem 8).  $\square$

In the proof of the following result, we show that the chase is a decision procedure for BCQ entailment over EMFA ontologies that runs in 2EXPTIME.

**Theorem 6.** Deciding BCQ entailment over an EMFA ontology is in 2EXPTIME.

*Proof.* The above result follows from the fact that, for any (arbitrarily chosen) chase  $\mathcal{C}$  of an EMFA ontology  $\mathcal{O} = \langle \mathcal{R}, \mathcal{F} \rangle$ , the atom set  $\mathcal{C}$  does not contain any cyclic terms. Hence, we can show that this atom set  $\mathcal{C}$  can be computed in 2EXPTIME with an analogous argument to the one that is used in the ‘‘Membership’’ part of the proof of Theorem 5 to show that  $E_{\mathcal{R}}$  can be computed in 2EXPTIME. Note that, once an atom is removed from a chase sequence due to the application of a tuple with an EGD, it may never be reintroduced in any descendant in the sequence by Definition 3. We show that  $\mathcal{C}$  does not contain cyclic terms via induction.

By Definition 3,  $\mathcal{C}$  is the union of all of the sets in some chase sequence  $\mathcal{A}_0, \mathcal{A}_1, \dots$  of  $\mathcal{O}$ . The fact that  $\mathcal{C}$  does not contain cyclic terms follows from the following claim: for all  $i \geq 0$ ,  $\mathcal{A}_i \pi_{\star} \subseteq E_{\mathcal{R}}$  with  $\pi_{\star}$  the rewriting that maps every ground term  $t$  to the term that results from replacing every syntactic occurrence of a constant with  $\star$  (e.g.,  $A(a, f(a)) \pi_{\star} = A(\star, f(\star))$  where  $a \in C$ ). Note that, if  $\mathcal{R}$  is EMFA, then  $E_{\mathcal{R}}$  does not contain cyclic terms.

(Base case) By Definition 3,  $\mathcal{A}_0 = \mathcal{F}$ . By Definition 9,  $\mathcal{F}_{\mathcal{R}}^* \subseteq E_{\mathcal{R}}$  and hence,  $\mathcal{A}_0 \pi_{\star} \subseteq E_{\mathcal{R}}$  since  $\mathcal{A}_0 \pi_{\star} \subseteq \mathcal{F}_{\mathcal{R}}^*$ .

(Inductive step) Let  $i \geq 1$ . Then, there is a rule  $\rho \in \mathcal{R}$  and a substitution  $\sigma$  such that  $\langle \rho, \sigma \rangle$  is applicable to  $\mathcal{A}_{i-1}$  and  $\mathcal{A}_i$  is the application of  $\langle \rho, \sigma \rangle$  on  $\mathcal{A}_{i-1}$ . By induction hypothesis, we have that  $\mathcal{A}_{i-1} \pi_{\star} \subseteq E_{\mathcal{R}}$ . We consider two different cases depending on whether  $\rho$  is a TGD or an EGD.

- Let  $\rho$  be a TGD; that is, this rule is of the form  $\beta \rightarrow \exists \vec{w}. \eta$ . Then,  $\beta\sigma \subseteq \mathcal{A}_{i-1}$  and  $\mathcal{A}_i = \mathcal{A}_{i-1} \cup \{sk_\sigma(\eta)\}$ . Moreover,  $(\beta\sigma)\pi_\star \subseteq \mathbf{E}_{\mathcal{R}}$  since  $\mathcal{A}_{i-1}\pi_\star \subseteq \mathbf{E}_{\mathcal{R}}$ . Hence,  $(sk_\sigma(\eta))\pi_\star \subseteq \mathbf{E}_{\mathcal{R}}$ , and  $\mathcal{A}_i\pi_\star$  is a subset of  $\mathbf{E}_{\mathcal{R}}$ .
- Let  $\rho$  be an EGD; that is, this rule is of the form  $\beta \rightarrow x \approx y$ . Then,  $\beta\sigma \subseteq \mathcal{A}_{i-1}$ , and  $\beta\sigma\pi_\star \subseteq \mathbf{E}_{\mathcal{R}}$  since  $\mathcal{A}_{i-1}\pi_\star \subseteq \mathbf{E}_{\mathcal{R}}$ . We consider two different cases.
  - $\sigma(x) \prec \sigma(y)$  with  $\prec$  the strict total order introduced before Definition 2. Then,  $\mathcal{A}_i = \mathcal{A}_i[\sigma(y)/\sigma(x)]$ . Moreover,  $\mathbf{E}_{\mathcal{R}}[\pi_\star(\sigma(y))/\pi_\star(\sigma(x))] \subseteq \mathbf{E}_{\mathcal{R}}$  since  $\sigma(x) \prec \sigma(y)$  implies  $dep(\sigma(x)) \leq dep(\sigma(y))$ .
  - $\sigma(y) \prec \sigma(x)$ . Analogous to the previous case.

In either case,  $\mathcal{A}_i\pi_\star$  is a subset of  $\mathbf{E}_{\mathcal{R}}$ .  $\square$

As implied by the following result, the chase is a worst-case optimal procedure to reason with EMFA ontologies.

**Theorem 7.** *Deciding BCQ entailment for ontologies  $\langle \mathcal{R}, \mathcal{F} \rangle$  with  $\mathcal{R}$  an EMFA rule set is 2EXPTIME-hard.*

*Proof.* Hardness is established by modifying the construction of a 2EXPTIME Turing machine given for weakly acyclic rules by (Cali, Gottlob, and Pieris 2010). For a more detailed explanation of this argument, see the proof of Theorem 3 in (Carral, Dragoste, and Krötzsch 2017).  $\square$

In the remainder of the section, we present some results and examples that illustrate the expressivity of EMFA compared to that of using MFA over axiomatised rule sets.

**Theorem 8.** *A rule set  $\mathcal{R}$  is EMFA if  $\text{St}(\mathcal{R})$  is MFA.*

*Proof.* Let  $\mathbf{E}_{\mathcal{R}}^0 = \mathcal{F}_{\mathcal{R}}^*$ ,  $\mathbf{E}_{\mathcal{R}}^1, \dots$  be a sequence consisting of all the intermediate sets that are computed to construct the set  $\mathbf{E}_{\mathcal{R}}$  by applying the rules defined in Definition 9. We show that  $\mathbf{E}_{\mathcal{R}}^i \subseteq \mathbf{E}_{\text{St}(\mathcal{R})}$  for all  $i \geq 1$  via induction. Hence,  $\mathbf{E}_{\mathcal{R}} \subseteq \mathbf{E}_{\text{St}(\mathcal{R})}$  and the theorem follows.

(Base Case) By Definition 9,  $\mathcal{F}_{\text{St}(\mathcal{R})}^* \subseteq \mathbf{E}_{\text{St}(\mathcal{R})}$ . Therefore, we have that  $\mathbf{E}_{\mathcal{R}}^0 \subseteq \mathbf{E}_{\text{St}(\mathcal{R})}$  since  $\mathcal{F}_{\mathcal{R}}^* \subseteq \mathcal{F}_{\text{St}(\mathcal{R})}^*$ .

(Inductive Step) Let  $i \geq 1$ . Then, by IH we have that  $\mathbf{E}_{\mathcal{R}}^{i-1} \subseteq \mathbf{E}_{\text{St}(\mathcal{R})}$ . We consider the following cases.

- $\mathbf{E}_{\mathcal{R}}^i = sk_\sigma(\eta) \cup \mathbf{E}_{\mathcal{R}}^{i-1}$  where  $\eta$  is the head of some TGD  $\rho = \beta \rightarrow \exists \vec{w}. \eta \in \mathcal{R}$  and  $\sigma$  is some substitution such that  $\beta\sigma \subseteq \mathbf{E}_{\mathcal{R}}^{i-1}$ . By Definition 4,  $\rho \in \text{St}(\mathcal{R})$ . Since  $\beta\sigma \subseteq \mathbf{E}_{\text{St}(\mathcal{R})}$  by IH,  $sk_\sigma(\eta) \subseteq \mathbf{E}_{\text{St}(\mathcal{R})}$ .
- $\mathbf{E}_{\mathcal{R}}^i = \mathbf{E}_{\mathcal{R}}^{i-1}[\sigma(x)/\sigma(y)] \cup \mathbf{E}_{\mathcal{R}}^{i-1}$ , there is an EGD  $\beta \rightarrow x \approx y \in \mathcal{R}$  and a substitution  $\sigma$  with  $\beta\sigma \subseteq \mathbf{E}_{\mathcal{R}}^{i-1}$ , and  $dep(\sigma(x)) \leq dep(\sigma(y))$ . By Definition 4,  $\beta \rightarrow Eq(x, y) \in \text{St}(\mathcal{R})$ . Since  $\beta\sigma \subseteq \mathbf{E}_{\text{St}(\mathcal{R})}$  by IH,  $Eq(x, y)\sigma \in \mathbf{E}_{\text{St}(\mathcal{R})}$ . Since TGD (4)  $\in \text{St}(\mathcal{R})$ ,  $Eq(\sigma(y), \sigma(x)) \in \mathbf{E}_{\text{St}(\mathcal{R})}$ . Because of the rules of type (6) in  $\text{St}(\mathcal{R})$ ,  $\varphi[\sigma(x)/\sigma(y)] \in \mathbf{E}_{\text{St}(\mathcal{R})}$  for all  $\varphi \in \mathbf{E}_{\text{St}(\mathcal{R})}$ .
- $\mathbf{E}_{\mathcal{R}}^i = \mathbf{E}_{\mathcal{R}}^{i-1}[\sigma(y)/\sigma(x)] \cup \mathbf{E}_{\mathcal{R}}^{i-1}$ , there is an EGD  $\beta \rightarrow x \approx y \in \mathcal{R}$  and a substitution  $\sigma$  with  $\beta\sigma \subseteq \mathbf{E}_{\mathcal{R}}^{i-1}$ , and  $dep(\sigma(y)) \leq dep(\sigma(x))$ . Analogous to the previous case.

In either case,  $\mathbf{E}_{\mathcal{R}}^i \subseteq \mathbf{E}_{\text{St}(\mathcal{R})}$ .  $\square$

As shown by the following example, the converse of the above theorem does not hold.

**Example 1.** *The rule set  $\mathcal{R}$  from Example 2 is EMFA, but the TGD set  $\text{St}(\mathcal{R})$  is not MFA.*

To conclude the section, we introduce some examples that illustrate the generality of EMFA versus that of applying MFA over singularised rule sets. For instance, there are rule sets that are EMFA, but no singularisation of these are MFA.

**Example 2.** *The rule set  $\mathcal{R}$  containing all of the following rules is EMFA, but no TGD set in  $\text{Sg}(\mathcal{R})$  is MFA.*

$$A(x) \rightarrow \exists v. R(x, v) \wedge B(v) \quad (12)$$

$$A(x) \rightarrow \exists w. S(x, w) \wedge C(w) \quad (13)$$

$$C(x) \wedge B(x) \rightarrow A(x) \quad (14)$$

$$R(x, y) \rightarrow x \approx y \quad (15)$$

$$S(x, y) \rightarrow x \approx y \quad (16)$$

For example, the chase of an ontology  $\langle \mathcal{R}, \mathcal{F} \rangle$  with  $\mathcal{R} \in \text{Sg}(\mathcal{R})$  and  $\mathcal{F} = \{A(a), R(a, a), S(a, a)\}$  does not terminate irrespectively of  $\mathcal{R}$ . Note that, this is the case even though all such ontologies admit finite chases. Therefore, neither TGD set in  $\text{Sg}(\mathcal{R})$  is MFA.

Furthermore, there are rule sets that are not EMFA, but all of their singularisations are MFA.

**Example 3.** *Even though the rule set  $\mathcal{R}$  with all of the following rules is not EMFA, all of the sets in  $\text{Sg}(\mathcal{R})$  are MFA.*

$$A(x) \rightarrow \exists v. R(x, v) \wedge B(v) \quad (17)$$

$$B(x) \rightarrow \exists w. R(x, w) \wedge C(w) \quad (18)$$

$$R(x, y) \wedge R(x, z) \rightarrow y \approx z \quad (19)$$

Note that, the two TGD sets in  $\text{Sg}(\mathcal{R})$  are equivalent.

## Evaluation

We performed some experiments to verify, from an empirical perspective, claims (I–IV) stated in the introduction.

To verify (I), we implemented the “renaming” chase variant presented in Definition 3 in VLog (Urbani, Jacobs, and Krötzsch 2016), which is an efficient rule engine for existential rules (Urbani et al. 2018). Then, we checked if using this procedure to compute the chase over an ontology  $\langle \mathcal{R}, \mathcal{F} \rangle$  is more efficient than computing the chase of an ontology  $\langle \mathcal{R}', \mathcal{F} \rangle$  with  $\mathcal{R}'$  some arbitrarily chosen rule set in  $\text{Sg}(\mathcal{R})$ . For this experiment we considered two ontologies—DBPedia (Bizer et al. 2009) and Claros (Rahtz et al. 2011)—that we obtained from the evaluation in (Motik et al. 2015). In either case, the performance of the “renaming” chase was far superior: we can compute the chase of DBPedia in 27.5s when using renaming to deal with equality; computing the chase of a (randomly selected) singularisation of this ontology takes 590s. We get similarly lopsided results for Claros: 11.7s when using renaming; 67.4s with singularisation.

Motik et al. (2015) showed that using renaming to deal with equality is more efficient than applying the standard axiomatisation. Therefore, we conclude that the use of axiomatisations (singularisation and standard) results in poor performance; a fact that validates the practical usefulness of

#TGDs \ #EGDs	[1, 2]	[3, 7]	$\geq 8$
[1, 2]	106 / 101 / 101 6ms / 9ms	56 / 50 / 50 13ms / 21ms	8 / 7 / 7 28ms / 46ms
[3, 15]	89 / 65 / 65 25ms / 51ms	92 / 74 / 74 31ms / 71ms	87 / 66 / 66 41ms / 65ms
$\geq 16$	13 / 7 / 7 404ms / 606ms	12 / 3 / 3 56ms / 94ms	<b>102 / 14 / 11</b> 2.7s / 7.6s

#TGDs \ #EGDs	[1]	[2, 3]	$\geq 4$
[1]	5 / 5 / 5 2ms / 4ms	1 / 1 / 1 5ms / 11ms	1 / 1 / 1 6ms / 10ms
[2, 5]	6 / 5 / 5 10ms / 18ms	9 / 8 / 8 5ms / 9ms	5 / 5 / 5 10ms / 15ms
$\geq 6$	82 / 9 / 9 62s / 78s	2 / 2 / 2 16ms / 27ms	20 / 2 / 2 <b>15s / 124s</b>

Figure 1: MOWLCorp (left) and Oxford Ontology Repository (right) results, see Summary 1 for an explanation of the above

Theorems 1 and 3. As discussed in the introduction, the application of these results allows us to check acyclicity with respect to some axiomatisation of a rule set, and then use the original rule set for computing the chase.

To verify claims (II–IV), we use Description Logics TBoxes (Baader et al. 2003) from MOWLCorp (Matentzoglou, Bail, and Parsia 2013) and the Oxford Ontology Repository<sup>2</sup> (OOR). First, we normalise these TBoxes by structural decomposition of complex axioms into the normal form considered in (Carral et al. 2014) and subsequently filter all TBoxes with non-deterministic features or nominals. Then, we apply a standard translation into FOL to obtain equivalent existential rule sets (see Section 6 in (Cuenca Grau et al. 2013)). Finally, we discard rule sets that do not contain at least one EGD and at least one TGD with existential quantified variables.

We implemented the EMFA and MFA checks in VLog and applied them to the rule sets from MOWLCorp and OOR. In summary, we obtained the following results:

- MOWLCorp: out of a total of 565 rule sets after preprocessing, we found that 387 rule sets are EMFA. Moreover, 73 and 384 rule sets are MFA if we apply the standard axiomatisation and singularisation, respectively.
- OOR: out of 131 rule sets, we found that that 38 rule sets are EMFA. Also, 38 are MFA if singularisation is applied.

For each rule set  $\mathcal{R}$ , we compute a single (arbitrarily chosen) set in  $\text{Sg}(\mathcal{R})$  when testing singularisation. Given the poor performance of the MFA check with the standard axiomatisation for MOWLCorp, we did not consider this technique again when we repeated the experiment with OOR.

From the above results, we can readily verify claim (II). All rule sets classified as MFA when applying either axiomatisation technique were also found to be EMFA and therefore, we consider that claim (III) was also validated.

To verify (IV), we measured the time that took to perform each check. On average, the EMFA check takes 35% and 60% of the time taken by the “MFA + singularisation” check for the rule sets in MOWLCorp and OOR, respectively. We present a more detailed analysis in Figure 1.

**Summary 1.** *In each cell in Figure 1, we include information about some subset of the rule sets in MOWLCorp (left table) or OOR (right table). For example, the upper right cell in the left table contains the counts and average times for the rule sets in MOWLCorp that contain at least 8 EGDs and between 1 and 2 TGDs with existential quantifiers. Each cell*

*contains two lines: the first features the total count of rule sets included, as well as the number of EMFA and “MFA + singularisation” successful checks; the second one includes two values which indicate the average times taken by the EMFA and the “MFA + singularisation” tests.*

Note that, EMFA outperforms “MFA + singularisation” by almost an order of magnitude on average for the hardest rule sets considered (i.e., lower right cell in the right table). Moreover, in a small number of cases, EMFA succeeded when MFA failed (lower right cell in the left table).

## Related Work and Conclusions

A previously existing acyclicity notion that can be directly applied to rule sets with EGDs is *weak acyclicity* (Fagin et al. 2005). Alas, this notion is significantly less general than checking MFA membership over singularised rule sets in practice (see Section 7 of (Cuenca Grau et al. 2013)).

As for future work, we plan to extend *restricted model-faithful acyclicity* (RMFA) (Carral, Dragoste, and Krötzsch 2017), an acyclicity notion for the *Datalog-first restricted chase*, so it can be applied to rule sets with equality. Since RMFA is more general than MFA, this extension can yield an even more general condition applicable for rule sets with EGDs. To verify that this notion captures all possible rule sets with a terminating chase, we plan to develop a *cyclicity notion* such as the one presented in (Carral, Dragoste, and Krötzsch 2017). That is, a sufficient condition that can detect if the chase does not terminate for a given rule set.

In this paper, we have presented several results that we believe are of theoretical interest and of practical usefulness regarding chase termination of rule sets with EGDs. In particular, we believe that Theorems 1 and 3 are very useful, as they enable the application of all existing acyclicity notions to existential rule sets with equality.

**Acknowledgments** This work is funded by Deutsche Forschungsgemeinschaft (DFG) grant 389792660 as part of TRR 248 (see [www.perspicuous-computing.science](http://www.perspicuous-computing.science)) and by the NWO research programme 400.17.605 (VWData). We also thank Irina Dragoste for her useful comments.

## References

Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D.; and Patel-Schneider, P. F., eds. 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.

<sup>2</sup><https://www.cs.ox.ac.uk/isg/ontologies/>

- Baget, J.; Mugnier, M.; and Thomazo, M. 2011. Towards Farsighted Dependencies for Existential Rules. In *Proc. of the 5th Int. Joint Conf. on Web Reasoning and Rule Systems (RR)*, volume 6902, 30–45. Springer.
- Beeri, C., and Vardi, M. Y. 1981. The implication problem for data dependencies. In *Proc. of the 8th Int. Colloquium on Automata, Languages and Programming (ICALP)*, 73–85. Springer-Verlag.
- Bizer, C.; Lehmann, J.; Kobilarov, G.; Auer, S.; Becker, C.; Cyganiak, R.; and Hellmann, S. 2009. DBpedia - A crystallization point for the web of data. *J. Web Semant.* 7(3):154–165.
- Calì, A.; Gottlob, G.; and Pieris, A. 2010. Query Answering under Non-guarded Rules in Datalog+/- . In *Proc. of the 4th Int. Joint Conf. on Web Reasoning and Rule Systems (RR)*, volume 6333, 1–17. Springer.
- Carral, D., and Urbani, J. 2019. How to check chase termination for existential rules with equality: Extended technical report. Available at <https://iccl.inf.tu-dresden.de/web/Article3057/en>.
- Carral, D.; Feier, C.; Cuenca Grau, B.; Hitzler, P.; and Horrocks, I. 2014.  $\mathcal{EL}$ -ifying Ontologies. In *Proc. of the 7th Int. Joint Conf. on Automated Reasoning (IJCAR)*, 464–479.
- Carral, D.; Dragoste, I.; and Krötzsch, M. 2017. Restricted Chase (Non)Termination for Existential Rules with Disjunctions. In *Proc. of the 26th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 922–928.
- Carral, D.; Feier, C.; and Hitzler, P. 2016. A Practical Acyclicity Notion for Query Answering Over Horn-*SRIQ* Ontologies. In *Proc. of the 15th Int. Semantic Web Conf. (ISWC)*, 70–85.
- Cuenca Grau, B.; Horrocks, I.; Krötzsch, M.; Kupke, C.; Magka, D.; Motik, B.; and Wang, Z. 2013. Acyclicity notions for existential rules and their application to query answering in ontologies. *J. Artificial Intelligence Research* 47:741–808.
- Deutsch, A.; Nash, A.; and Rummel, J. B. 2008. The Chase Revisited. In *Proc. of the 27th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, 149–158.
- Fagin, R.; Kolaitis, P. G.; Miller, R. J.; and Popa, L. 2005. Data exchange: semantics and query answering. *J. Theoretical Computer Science* 336(1):89–124.
- Gogacz, T., and Marcinkowski, J. 2014. All-Instances Termination of Chase is Undecidable. In *Proc. of the 41st Int. Colloquium on Automata, Languages and Programming (ICALP)*, 293–304.
- Grahne, G., and Onet, A. 2018. Anatomy of the chase. *Fundam. Inform.* 157(3):221–270.
- Karimi, A.; Zhang, H.; and You, J. 2018. Restricted Chase Termination: A Hierarchical Approach and Experimentation. In *Proc. of the 2th Int. Joint Conf. on Rules and Reasoning (RuleML+RR)*, volume 11092, 98–114. Springer.
- Krötzsch, M., and Rudolph, S. 2011. Extending Decidable Existential Rules by Joining Acyclicity and Guardedness. In *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 963–968. IJCAI/AAAI.
- Maier, D.; Mendelzon, A. O.; and Sagiv, Y. 1979. Testing implications of data dependencies. *J. ACM Transactions of Database Systems* 4(4):455–469.
- Marnette, B. 2009. Generalized Schema-Mappings: from Termination to Tractability. In *Proc. of the 28th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, 13–22.
- Matentzoglou, N.; Bail, S.; and Parsia, B. 2013. A Snapshot of the OWL Web. In *Proc. of the 12th Int. Semantic Web Conf. (ISWC)*, 331–346.
- Motik, B.; Nenov, Y.; Piro, R. E. F.; and Horrocks, I. 2015. Handling OWL: sameAs via Rewriting. In *Proc. of the 29th AAAI Conf. on Artificial Intelligence (AAAI)*, 231–237. AAAI Press.
- Rahtz, S.; Dutton, A.; Kurtz, D.; Klyne, G.; Zisserman, A.; and Arandjelovic, R. 2011. CLAROS – Collaborating on Delivering the Future of the Past. In *Digital Humanities 2011, Conference Abstracts (DH)*, 355–357.
- Urbani, J.; Krötzsch, M.; Jacobs, C. J. H.; Dragoste, I.; and Carral, D. 2018. Efficient Model Construction for Horn Logic with VLog – System Description. In *Proc. of the 9th Int. Joint Conf. on Automated Reasoning (IJCAR)*, 680–688.
- Urbani, J.; Jacobs, C. J. H.; and Krötzsch, M. 2016. Column-Oriented Datalog Materialization for Large Knowledge Graphs. In *Proc. of the 30th AAAI Conf. on Artificial Intelligence (AAAI)*, 258–264. AAAI Press.