arXiv:1404.3141v1 [cs.AI] 11 Apr 2014

Datalog Rewritability of Disjunctive Datalog Programs and its Applications to Ontology Reasoning

Mark Kaminski and **Yavor Nenov** and **Bernardo Cuenca Grau**
Department of Computer Science, University of Oxford, UK

Abstract

We study the problem of rewriting a disjunctive datalog program into plain datalog. We show that a disjunctive program is rewritable if and only if it is equivalent to a linear disjunctive program, thus providing a novel characterisation of datalog rewritability. Motivated by this result, we propose *weakly linear disjunctive datalog*—a novel rule-based KR language that extends both datalog and linear disjunctive datalog and for which reasoning is tractable in data complexity. We then explore applications of weakly linear programs to ontology reasoning and propose a tractable extension of OWL 2 RL with disjunctive axioms. Our empirical results suggest that many non-Horn ontologies can be reduced to weakly linear programs and that query answering over such ontologies using a datalog engine is feasible in practice.

1 Introduction

Disjunctive datalog, which extends plain datalog by allowing disjunction in the head of rules, is a prominent KR formalism that has found many applications in the areas of deductive databases, information integration and ontological reasoning (Eiter, Gottlob, and Mannila 1997; Dantsin et al. 2001).[1] Disjunctive datalog is a powerful language, which can model incomplete information. Expressiveness comes, however, at the expense of computational cost: fact entailment is co-NEXPTIME-complete in combined complexity and co-NP-complete w.r.t. data (Eiter, Gottlob, and Mannila 1997). Thus, even with the development of optimised implementations (Leone et al. 2006), robust behaviour of reasoners in data-intensive applications cannot be guaranteed.

Plain datalog offers more favourable computational properties (EXPTIME-completeness in combined complexity and PTIME-completeness w.r.t. data) at the expense of a loss in expressive power (Dantsin et al. 2001). Tractability in data complexity is an appealing property for data-intensive KR; in particular, the RL profile of the ontology language OWL 2 was designed such that each ontology corresponds to a datalog program (Motik et al. 2009). Furthermore, datalog programs obtained from RL ontologies contain rules of a restricted shape, and they can be evaluated in polynomial time also in *combined complexity*, thus providing the ground for robust implementations. The standardisation of OWL 2 RL has spurred the development of reasoning engines within industry and academia, such as OWLim (Bishop et al. 2011), Oracle's Semantic Data Store (Wu et al. 2008), and RDFox (Motik et al. 2014).

We study the problem of rewriting a disjunctive datalog program into an equivalent datalog program (i.e., one that entails the same facts for every dataset). By computing such rewritings, we can ensure tractability w.r.t. data and exploit reasoning infrastructure available for datalog. Not every disjunctive datalog program is, however, datalog rewritable (Afrati, Cosmadakis, and Yannakakis 1995).

Our first contribution is a novel characterisation of datalog rewritability based on *linearity*: a restriction that requires each rule to contain at most one body atom with an IDB predicate (i.e., a predicate occurring in head position). For plain datalog, linearity is known to limit the effect of recursion and lead to reduced data and combined complexity (Dantsin et al. 2001). For disjunctive programs the effects of the linearity restriction are, to the best of our knowledge, unknown. In Section 3, we show that every linear disjunctive program can be polynomially transformed into an equivalent datalog program; conversely, we also provide a polynomial transformation from datalog into linear disjunctive datalog. Thus, linear disjunctive datalog and datalog have the same computational properties, and linearisability of disjunctive programs is equivalent to rewritability into datalog.

Motivated by our characterisation, in Section 4 we propose *weakly linear disjunctive datalog*: a rule language that extends both datalog and linear disjunctive datalog. In a weakly linear (WL for short) program, the linearity requirement is relaxed: instead of applying to all IDB predicates, it applies only to those that "depend" on a disjunctive rule. Analogously to linear disjunctive programs, WL programs can be polynomially rewritten into datalog. Thus, our language captures disjunctive information while leaving the favourable computational properties of datalog intact.

In Section 5, we propose a linearisation procedure based on unfolding transformations. Our procedure picks a non-WL rule and a "culprit" body atom and replaces it with WL rules by "unfolding" the selected atom. Our procedure is incomplete: if it succeeds, it outputs a WL program, which is rewritten into datalog; if it fails, no conclusion can be drawn.

In Section 6, we focus on ontology reasoning. We pro-

[1] Disjunctive datalog typically allows for negation-as-failure, which we don't consider since we focus on monotonic reasoning.

pose an extension of OWL 2 RL with disjunctive axioms such that each ontology in our extended profile maps to a WL program. We show that the resulting programs can be evaluated in polynomial time in *combined* complexity; thus, fact entailment in our language is no harder than in OWL 2 RL. Finally, we argue that the algorithm in (Hustadt, Motik, and Sattler 2007) can be combined with our techniques to rewrite a $\mathcal{SHIQ}$ ontology into a plain datalog program.

We have evaluated our techniques on a large ontology repository. Our results show that many non-Horn ontologies can be rewritten into WL programs, and thus into datalog. We have tested the scalability of query answering using our approach, with promising results. Proofs of our technical results are delegated to the appendix.

2 Preliminaries

We use standard first-order syntax and semantics and assume all formulae to be function-free. We assume that equality $\approx$ is an ordinary predicate and that every set of set of formulae contains the standard explicit axiomatisation of $\approx$ as a congruence relation for its signature.

A *fact* is a ground atom and a *dataset* is a finite set of facts. A *rule* r is a sentence of the form $\forall\vec{x}\forall\vec{z}.[\varphi(\vec{x},\vec{z}) \rightarrow \psi(\vec{x})]$, where tuples of variables $\vec{x}$ and $\vec{z}$ are disjoint, $\varphi(\vec{x},\vec{z})$ is a conjunction of distinct equality-free atoms, and $\psi(\vec{x})$ is a disjunction of distinct atoms. Formula φ is the *body* of r, and ψ is the *head*. Quantifiers in rules are omitted. We assume that rules are *safe*, i.e., all variables in the head occur in the body. A rule is *datalog* if $\psi(\vec{x})$ has at most one atom, and it is *disjunctive* otherwise. A *program* $\mathcal{P}$ is a finite set of rules; it is *datalog* if it consists only of datalog rules, and *disjunctive* otherwise. We assume that rules in $\mathcal{P}$ do not share variables.

For convenience, we treat $\top$ and $\bot$ in a non-standard way as a unary and a nullary predicate, respectively. Given a program $\mathcal{P}$, $\mathcal{P}_\top$ is the program with a rule $Q(x_1,\ldots,x_n) \rightarrow \top(x_i)$ for each predicate Q in $\mathcal{P}$ and each $1 \leq i \leq n$, and a rule $\rightarrow \top(a)$ for each constant a in $\mathcal{P}$. We assume that $\mathcal{P}_\top \subseteq \mathcal{P}$ and $\top$ does not occur in head position in $\mathcal{P} \setminus \mathcal{P}_\top$. We define $\mathcal{P}_\bot$ as consisting of a rule with $\bot$ as body and empty head. We assume $\mathcal{P}_\bot \subseteq \mathcal{P}$ and no rule in $\mathcal{P} \setminus \mathcal{P}_\bot$ has an empty head or $\bot$ in the body. Thus, $\mathcal{P} \cup \mathcal{D} \models \top(a)$ for every a in $\mathcal{P} \cup \mathcal{D}$, and $\mathcal{P} \cup \mathcal{D}$ is unsatisfiable iff $\mathcal{P} \cup \mathcal{D} \models \bot$.

Head predicates in $\mathcal{P} \setminus \mathcal{P}_\top$ are *intensional* (or *IDB*) in $\mathcal{P}$. All other predicates (including $\top$) are *extensional* (*EDB*). An atom is intensional (extensional) if so is its predicate. A rule is *linear* if it has at most one IDB body atom. A program $\mathcal{P}$ is linear if all its rules are. In contrast to KR, in logic programming it is often assumed that IDB predicates do not occur in datasets. This assumption can be lifted (see, e.g., (Bry et al. 2007)): for every $\mathcal{P}$ and IDB predicate Q in $\mathcal{P}$, let Q' be a fresh predicate; the *IDB expansion* $\mathcal{P}^e$ of $\mathcal{P}$ is obtained from $\mathcal{P}$ by renaming each IDB predicate Q in $\mathcal{P}$ with Q' and adding a rule $Q(\vec{x}) \rightarrow Q'(\vec{x})$, with $\vec{x}$ distinct variables. Then, for each $\mathcal{D}$ and each fact α over the signature of $\mathcal{P}$ we have $\mathcal{P} \cup \mathcal{D} \models \alpha$ iff $\mathcal{P}^e \cup \mathcal{D} \models \alpha\theta$, where θ is the predicate substitution mapping each IDB predicate Q to Q'.

The *evaluation* of $\mathcal{P}$ over a dataset $\mathcal{D}$ is the set $\mathsf{Eval}(\mathcal{P},\mathcal{D})$ which comprises $\bot$ if $\mathcal{P} \cup \mathcal{D}$ is unsatisfiable and all facts entailed by $\mathcal{P} \cup \mathcal{D}$ otherwise. For a set of predicates S, $\mathsf{Eval}(\mathcal{P},\mathcal{D})|_S$ consists of those facts in $\mathsf{Eval}(\mathcal{P},\mathcal{D})$ involving predicates in $S \cup \{\bot\}$. Program $\mathcal{P}'$ is a *rewriting* of $\mathcal{P}$ w.r.t. a set of predicates S if there is an injective predicate renaming θ such that $(\mathsf{Eval}(\mathcal{P},\mathcal{D})|_S)\theta = \mathsf{Eval}(\mathcal{P}',\mathcal{D})|_{S\theta}$ for every dataset $\mathcal{D}$ over the signature of $\mathcal{P}$. The program $\mathcal{P}'$ is a rewriting of $\mathcal{P}$ if $\mathcal{P}'$ is a rewriting of $\mathcal{P}$ w.r.t. the set of all predicates in $\mathcal{P}$. Clearly, $\mathcal{P}^e$ is a rewriting of $\mathcal{P}$.

3 Characterisation of Datalog Rewritability

In this section, we establish a strong correspondence between linear disjunctive datalog and plain datalog. We show that every linear disjunctive program can be polynomially rewritten into datalog and, conversely, every datalog program is polynomially rewritable to a linear disjunctive program. Consequently, we not only can conclude that fact entailment over linear programs has exactly the same data and combined complexity as over plain datalog programs, but also that a disjunctive program is datalog rewritable if and only if it is linearisable. Thus, datalog rewritability and linearisability of disjunctive programs are equivalent problems.

From Linear Programs to Datalog We first show that linear disjunctive programs can be polynomially rewritten into datalog. Let us consider the following program $\mathcal{P}_1$, which we want to rewrite into a datalog program $\Xi(\mathcal{P}_1)$:

$$\mathcal{P}_1 = \{\, V(x) \rightarrow B(x) \vee G(x) \quad (1)$$

$$G(y) \wedge E(x,y) \rightarrow B(x) \quad (2)$$

$$B(y) \wedge E(x,y) \rightarrow G(x) \,\} \quad (3)$$

Predicates V and E are EDB, so their extension depends solely on $\mathcal{D}$. To prove facts about IDB predicates G and B we introduce fresh binary predicates B^G, B^B, G^B, and G^G. Intuitively, if a fact $B^G(c,d)$ holds in $\Xi(\mathcal{P}_1) \cup \mathcal{D}$ then proving $B(c)$ suffices for proving $G(d)$ in $\mathcal{P}_1 \cup \mathcal{D}$. To "initialise" the extension of these fresh predicates we need rules $\top(x) \rightarrow X^X(x,x)$ with $X \in \{G,B\}$. The key step is then to "flip" the direction of all rules in $\mathcal{P}_1$ involving G or B by moving all IDB atoms from the head to the body and vice-versa while at the same time replacing their predicates with the relevant auxiliary predicates. Thus, Rule (2) leads to the following rules in $\Xi(\mathcal{P}_1)$ for each IDB predicate X:

$$B^X(x,z) \wedge E(x,y) \rightarrow G^X(y,z)$$

These rules are natural consequences of Rule (2) under the intended meaning of the auxiliary predicates: if we can prove a goal $X(z)$ by proving first $B(x)$, and $E(x,y)$ holds, then by Rule (2) we deduce that proving $G(y)$ suffices to prove $X(z)$. In contrast to (2), Rule (1) contains no IDB body atoms. We "flip" this rule as follows, with X IDB:

$$V(x) \wedge B^X(x,z) \wedge G^X(x,z) \rightarrow X(z)$$

Similarly to the previous case, this rule follows from Rule (1): if $V(x)$ holds and we can establish that $X(z)$ can be proved from $B(x)$ and also from $G(x)$, then $X(z)$ must hold. Finally, we introduce rules that allow us to derive facts about the IDB predicates G and B from facts derived about the auxiliary predicates. For example, the rule $B(x) \wedge B^X(x,z) \rightarrow X(z)$ states that if $B(x)$ holds and is sufficient to prove $X(z)$, then $X(z)$ must also hold.

Definition 1. Let $\mathcal{P}$ be a linear program and let Σ be the set of IDB predicates in $\mathcal{P} \setminus \mathcal{P}_\top$. For each $(P, Q) \in \Sigma^2$, let P^Q be a fresh predicate unique to (P, Q) where $\mathsf{arity}(P^Q) = \mathsf{arity}(P) + \mathsf{arity}(Q)$. Then $\Xi(\mathcal{P})$ is the datalog program containing the rules given next, where φ is the conjunction of all EDB atoms in a rule, $\varphi_\top$ is the least conjunction of $\top$-atoms needed to make a rule safe, all predicates P_i are in Σ, and $\vec{y}$, $\vec{z}$ are disjoint vectors of distinct fresh variables:

1. a rule $\varphi_\top \to R^R(\vec{y}, \vec{y})$ for every $R \in \Sigma$;
2. a rule $\varphi_\top \wedge \varphi \wedge \bigwedge_{i=1}^n P_i^R(\vec{s}_i, \vec{y}) \to Q^R(\vec{t}, \vec{y})$ for every rule $\varphi \wedge Q(\vec{t}) \to \bigvee_{i=1}^n P_i(\vec{s}_i) \in \mathcal{P} \setminus \mathcal{P}_\top$ and every $R \in \Sigma$;
3. a rule $\varphi \wedge \bigwedge_{i=1}^n P_i^R(\vec{s}_i, \vec{y}) \to R(\vec{y})$ for every rule $\varphi \to \bigvee_{i=1}^n P_i(\vec{s}_i) \in \mathcal{P} \setminus \mathcal{P}_\top$ and every $R \in \Sigma$;
4. a rule $Q(\vec{z}) \wedge Q^R(\vec{z}, \vec{y}) \to R(\vec{y})$ for every $(Q, R) \in \Sigma^2$. $\diamond$

This transformation is quadratic and the arity of predicates is at most doubled. For $\mathcal{P}_1$, we obtain the following datalog program, where each rule mentioning X stands for one rule where $X = B$ and one where $X = G$:

$$\Xi(\mathcal{P}_1) = \{\, V(x) \wedge B^X(x, z) \wedge G^X(x, z) \to X(z) \quad (1')$$
$$B^X(x, z) \wedge E(x, y) \to G^X(y, z) \quad (2')$$
$$G^X(x, z) \wedge E(x, y) \to B^X(y, z) \quad (3')$$
$$\top(x) \to X^X(x, x) \quad (4)$$
$$B(x) \wedge B^X(x, z) \to X(z) \quad (5)$$
$$G(x) \wedge G^X(x, z) \to X(z) \,\} \quad (6)$$

Correctness of Ξ is established by the following theorem.

Theorem 2. *If $\mathcal{P}$ is linear, then $\Xi(\mathcal{P})$ is a polynomial datalog rewriting of $\mathcal{P}$.*

Thus, fact entailment over linear programs is no harder than in datalog: PTIME w.r.t. data and EXPTIME in combined complexity. Formally, Theorem 2 is shown by induction on hyperresolution derivations of facts entailed by the rules in $\mathcal{P}$ from a given dataset $\mathcal{D}$ (see Appendix). We next sketch the intuitions on $\mathcal{P}_1$ and $\mathcal{D}_1 = \{V(a), V(b), V(c), E(a, b), E(b, c), E(a, c)\}$.

Figure 1, Part (a) shows a linear (hyperresolution) derivation ρ_1 of $B(a)$ from $\mathcal{P}_1 \cup \mathcal{D}_1$ while Part (b) shows a derivation ρ_2 of the same fact from $\Xi(\mathcal{P}_1) \cup \mathcal{D}_1$. We represent derivations as trees whose nodes are labeled with disjunctions of facts and where every inner node is derived from its children using a rule of the program (initialisation rules in ρ_2 are omitted for brevity). We first show that if $B(a)$ is provable in $\mathcal{P}_1 \cup \mathcal{D}_1$, then it is entailed by $\Xi(\mathcal{P}_1) \cup \mathcal{D}_1$. From the premise, a linear derivation such as ρ_1 exists. The crux of the proof is to show that each disjunction of facts in ρ_1 corresponds to a set of facts over the auxiliary predicates entailed by $\Xi(\mathcal{P}_1) \cup \mathcal{D}_1$. Furthermore, these facts must be of the form $X^B(u, a)$, where $B(a)$ is the goal, u is a constant, and $X \in \{B, G\}$. For example, $B(c) \vee G(c)$ in ρ_1 corresponds to facts $B^B(c, a)$ and $G^B(c, a)$, which are provable from $\Xi(\mathcal{P}_1) \cup \mathcal{D}_1$, as witnessed by ρ_2. Since ρ_1 is linear, it has a unique rule application that has only EDB atoms as premises, i.e., the application of (1), which generates $B(c) \vee G(c)$. Since $B^B(c, a)$ and $G^B(c, a)$ are provable from $\Xi(\mathcal{P}_1) \cup \mathcal{D}_1$, we can apply (1') to derive $B(a)$.

Finally, we show the converse: if $B(a)$ is provable from $\Xi(\mathcal{P}_1) \cup \mathcal{D}_1$ then it follows from $\mathcal{P}_1 \cup \mathcal{D}_1$. For this, we take a derivation such as ρ_2, and show that each fact in ρ_2 about an auxiliary predicate carries the intended meaning, e.g., for $G^B(b, a)$ we must have $\mathcal{P}_1 \cup \mathcal{D}_1 \models G(b) \to B(a)$.

From Datalog to Linear Programs The transformation from datalog to linear disjunctive datalog is based on the same ideas, but it is simpler in that we no longer distinguish between EDB and IDB atoms: a rule in $\mathcal{P}$ is now "flipped" by moving *all* its atoms from the head to the body and vice-versa. Moreover, we make use of the IDB expansion $\mathcal{P}^e$ of $\mathcal{P}$ to ensure linearity of the resulting disjunctive program.

Definition 3. Let $\mathcal{P}$ be a datalog program. For each pair (P, Q) of predicates in $\mathcal{P}$, let P^Q be a fresh predicate unique to (P, Q) where $\mathsf{arity}(P^Q) = \mathsf{arity}(P) + \mathsf{arity}(Q)$. Furthermore, let $\mathcal{P}^e$ be the IDB expansion of $\mathcal{P}$. Then, $\Psi(\mathcal{P})$ is the linear disjunctive program containing, for each IDB predicate R in $\mathcal{P}^e$ the rules given next, where $\varphi_\top$ is the least conjunction of $\top$-atoms making a rule safe and $\vec{y} = y_1 \dots y_{\mathsf{arity}(R)}$ is a vector of distinct fresh variables:

1. a rule $\varphi_\top \wedge Q^R(\vec{t}, \vec{y}) \to \bigvee_{i=1}^n P_i^R(\vec{s}_i, \vec{y})$ for every rule $\bigwedge_{i=1}^n P_i(\vec{s}_i) \to Q(\vec{t}) \in \mathcal{P}^e \setminus \mathcal{P}^e_\top$, where $Q(\vec{t}) \neq \bot$ and $\bigvee_{i=1}^n P_i^R(\vec{s}_i, \vec{y})$ is interpreted as $\bot$ if $n = 0$;
2. a rule $\varphi_\top \to \bigvee_{i=1}^n P_i^R(\vec{s}_i, \vec{y})$ for every $\bigwedge_{i=1}^n P_i(\vec{s}_i) \to \bot \in \mathcal{P}^e \setminus \mathcal{P}^e_\top$;
3. a rule $\varphi_\top \to R^R(\vec{y}, \vec{y})$;
4. a rule $Q(\vec{z}) \wedge Q^R(\vec{z}, \vec{y}) \to R(\vec{y})$ for every EDB predicate Q in $\mathcal{P}^e$, where $\vec{z}$ is a vector of distinct fresh variables. $\diamond$

Again, the transformation is quadratic and the arity of predicates is at most doubled.

Example 4. Consider $\mathcal{P}_2$, which encodes path system accessibility (a canonical PTIME-complete problem):

$$\mathcal{P}_2 = \{\, R(x, y, z) \wedge A(y) \wedge A(z) \to A(x) \,\} \quad (7)$$

Linear datalog is NLOGSPACE, and cannot capture $\mathcal{P}_2$. However, we can rewrite $\mathcal{P}_2$ into linear disjunctive datalog:

$$\Psi(\mathcal{P}_2) = \{\, \top(y) \wedge \top(z) \wedge A'^{A'}(x, u) \quad (7')$$
$$\to R^{A'}(x, y, z, u) \vee A'^{A'}(y, u) \vee A'^{A'}(z, u)$$
$$A'^{A'}(x, y) \to A^{A'}(x, y) \quad (8)$$
$$\top(x) \to A'^{A'}(x, x) \quad (9)$$
$$A(x) \wedge A^{A'}(x, y) \to A'(y) \quad (10)$$
$$R(x, y, z) \wedge R^{A'}(x, y, z, u) \to A'(u) \,\} \quad (11)$$

Rule (7) yields Rule (7') in $\Psi(\mathcal{P}_2)$. Rule (8) is obtained from $A(x) \to A'(x) \in \mathcal{P}^e$. To see why we need the IDB expansion $\mathcal{P}^e$, suppose we replaced $\mathcal{P}^e$ by $\mathcal{P}$ in Definition 3. Rule (8) would not be produced and A' would be replaced by A elsewhere. Then the rule $A(x) \wedge A^A(x, y) \to A(y)$ would not be linear since both A and A^A would be IDB. $\diamond$

Correctness of Ψ is established by the following theorem.

Theorem 5. *If $\mathcal{P}$ is datalog, then $\Psi(\mathcal{P})$ is a polynomial rewriting of $\mathcal{P}$ into a linear disjunctive program.*

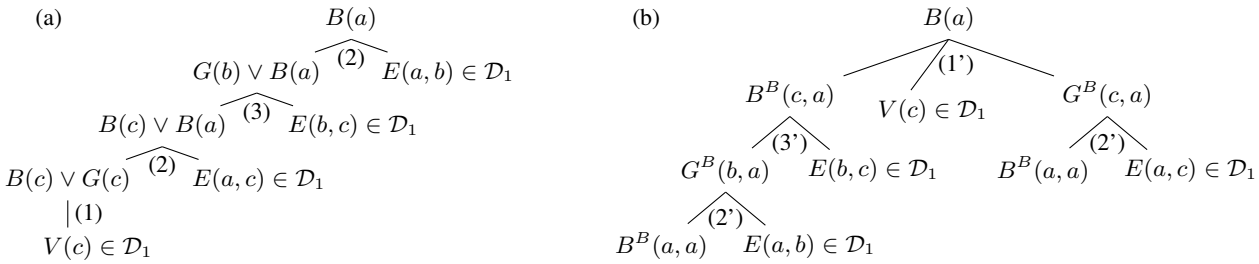


Figure 1: (a) derivation of $B(a)$ from $\mathcal{P}_1 \cup \mathcal{D}_1$; (b) derivation of $B(a)$ from $\Xi(\mathcal{P}_1) \cup \mathcal{D}_1$

From Theorems 2 and 5 we obtain the following results.

Corollary 6. *A disjunctive program $\mathcal{P}$ is datalog rewritable iff it is rewritable into a linear disjunctive program.*

Corollary 7. *Checking $\mathcal{P} \cup \mathcal{D} \models \alpha$ for $\mathcal{P}$ a linear program, $\mathcal{D}$ a dataset and α a fact is* PTIME*-complete w.r.t. data complexity and* EXPTIME*-complete w.r.t. combined complexity.*

4 Weakly Linear Disjunctive Datalog

In this section, we introduce weakly linear programs: a new class of disjunctive datalog programs that extends both datalog and linear disjunctive datalog. The main idea is simple: instead of requiring the body of each rule to contain at most one occurrence of an IDB predicate, we require at most one occurrence of a *disjunctive* predicate—a predicate whose extension for some dataset could depend on the application of a disjunctive rule. This intuition is formalised as given next.

Definition 8. The *dependency graph* $G_{\mathcal{P}} = (V, E, \mu)$ of a program $\mathcal{P}$ is the smallest edge-labeled digraph such that:

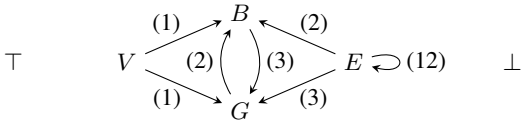
1. V contains every predicate occurring in $\mathcal{P}$;
2. $r \in \mu(P, Q)$ whenever $P, Q \in V$, $r \in \mathcal{P} \setminus \mathcal{P}_\top$, P occurs in the body of r, and Q occurs in the head of r; and
3. $(P, Q) \in E$ whenever $\mu(P, Q)$ is nonempty.

A predicate Q *depends on a rule* $r \in \mathcal{P}$ if $G_{\mathcal{P}}$ has a path that ends in Q and involves an r-labeled edge. Predicate Q is *datalog* if it only depends on datalog rules; otherwise, Q is *disjunctive*. Program $\mathcal{P}$ is *weakly linear* (WL for short) if every rule in $\mathcal{P}$ has at most one occurrence of a disjunctive predicate in the body. ◇

Checking whether $\mathcal{P}$ is WL is clearly feasible in polynomial time. If $\mathcal{P}$ is datalog, then all its predicates are datalog and $\mathcal{P}$ is WL. Furthermore, every disjunctive predicate is IDB and hence every linear program is also WL. There are, however, WL programs that are neither datalog nor linear. Consider $\mathcal{P}_3$, which extends $\mathcal{P}_1$ with the following rule:

$$E(y, x) \rightarrow E(x, y) \tag{12}$$

Since E is IDB in $\mathcal{P}_3$, Rules (2) and (3) have two IDB atoms. Thus, $\mathcal{P}_3$ is not linear. The graph $G_{\mathcal{P}_3}$ looks as follows.



Predicate V is EDB and hence does not depend on any rule. Predicates B and G depend on Rule (1) and hence are disjunctive. Finally, predicate E depends only on Rule (12) and hence it is a datalog predicate. Thus, $\mathcal{P}_3$ is WL.

Definition 9. For $\mathcal{P}$ WL, let $\Xi'(\mathcal{P})$ be defined as $\Xi(\mathcal{P})$ in Definition 1 but where: *(i)* Σ is the set of all disjunctive predicates in $\mathcal{P} \setminus \mathcal{P}_\top$; *(ii)* φ denotes the conjunction of all datalog atoms in a rule; and *(iii)* in addition to rules (1)–(4), $\Xi'(\mathcal{P})$ contains every rule in $\mathcal{P}$ with no disjunctive predicates. ◇

By adapting the proof of Theorem 2 we obtain:

Theorem 10. *If $\mathcal{P}$ is WL, then $\Xi'(\mathcal{P})$ is a polynomial datalog rewriting of $\mathcal{P}$.*

Thus, fact entailment over WL programs has the same data and combined complexity as for datalog. Furthermore, $\Xi'(\mathcal{P})$ is a rewriting of $\mathcal{P}$ and hence it preserves the extension of all predicates. If, however, we want to query a specific predicate Q, we can compute a smaller program, which is linear in the size of $\mathcal{P}$ and preserves the extension of Q. Indeed, if Q is datalog, each proof in $\mathcal{P}$ of a fact about Q involves only datalog rules, and if Q is disjunctive, each such proof involves only auxiliary predicates X^Q. Thus, in Ξ' we can dispense with all rules involving auxiliary predicates X^R for $R \neq Q$. In particular, if Q is datalog, the rewriting contains no auxiliary predicates.

Theorem 11. *Let $\mathcal{P}$ be WL, S a set of predicates in $\mathcal{P}$, and $\mathcal{P}'$ obtained from $\Xi'(\mathcal{P})$ by removing all rules with a predicate X^R for $R \notin S$. Then $\mathcal{P}'$ is a rewriting of $\mathcal{P}$ w.r.t. S.*

5 Rewriting Programs via Unfolding

Although WL programs can be rewritten into datalog, not all datalog rewritable programs are WL. Let $\mathcal{P}_4$ be as follows:

$$\mathcal{P}_4 = \{\, A(x) \wedge B(x) \rightarrow C(x) \vee D(x) \tag{13}$$

$$E(x) \rightarrow A(x) \vee F(x) \tag{14}$$

$$C(x) \wedge R(x, y) \rightarrow B(y)\,\} \tag{15}$$

Program $\mathcal{P}_4$ is not WL since both body atoms in (13) are disjunctive. However, $\mathcal{P}_4$ is datalog rewritable.

We now present a rewriting procedure that combines our results in Section 4 with the work of Gergatsoulis (1997) on program transformation for disjunctive logic programs. Our procedure iteratively eliminates non-WL rules by "unfolding" the culprit atoms w.r.t. the other rules in the program. It stops when the program becomes WL, and outputs a datalog program as in Section 4. The procedure is sound: if it

Procedure 1 Rewrite

Input: $\mathcal{P}$: a disjunctive program
Output: a datalog rewriting of $\mathcal{P}$

1: $\mathcal{P}' := \mathcal{P}^e$
2: **while** $\mathcal{P}'$ not WL **do**
3: select $r \in \mathcal{P}'$ with more than one disjunctive body atom
4: select a disjunctive body atom $\alpha \in r$
5: $\mathcal{P}' := \mathsf{Unfold}(\mathcal{P}', r, \alpha)$
6: **return** $\Xi'(\mathcal{P}')$

succeeds, the output is a datalog rewriting. It is, however, both incomplete (linearisability cannot be semi-decided just by unfolding) and non-terminating. Nevertheless, our experiments suggest that unfolding can be effective in practice since some programs obtained from realistic ontologies can be rewritten into datalog after a few unfolding steps.

Unfolding We start by recapitulating (Gergatsoulis 1997). Given a disjunctive program $\mathcal{P}$, a rule r in $\mathcal{P}$, and a body atom α of r, Gergatsoulis defines the *unfolding* of r at α in $\mathcal{P}$ as a transformation of $\mathcal{P}$ that replaces r with a set of resolvents of r with rules in $\mathcal{P}$ at α (see Appendix). We denote the resulting program by $\mathsf{Unfold}(\mathcal{P}, r, \alpha)$. Unfolding preserves all entailed disjunctions φ of facts: $\mathcal{P} \models \varphi$ iff $\mathsf{Unfold}(\mathcal{P}, r, \alpha) \models \varphi$ for all $\mathcal{P}$, r, α, and φ. However, to ensure that unfolding produces a rewriting we need a stronger correctness result that is dataset independent.

Theorem 12. *Let $\mathcal{P}_0$ be a disjunctive program and let $\mathcal{P}$ be a rewriting of $\mathcal{P}_0$ such that no IDB predicate in $\mathcal{P}$ occurs in $\mathcal{P}_0$. Let r be a rule in $\mathcal{P}$ and α be an IDB body atom of r. Then $\mathsf{Unfold}(\mathcal{P}, r, \alpha)$ is a rewriting of $\mathcal{P}_0$. Moreover, no IDB predicate in $\mathsf{Unfold}(\mathcal{P}, r, \alpha)$ occurs in $\mathcal{P}_0$.*

The Rewriting Procedure Procedure 1 attempts to eliminate rules with several disjunctive body atoms by unfolding one such atom. Note that to satisfy the premise of Theorem 12, unfolding is applied to $\mathcal{P}^e$ rather than $\mathcal{P}$. Correctness of Procedure 1 is established by the following theorem.

Theorem 13. *Let $\mathcal{P}$ be a disjunctive program. If* Rewrite *terminates on $\mathcal{P}$ with output $\mathcal{P}'$, then $\mathcal{P}'$ is a rewriting of $\mathcal{P}$.*

Rewrite first transforms our example program $\mathcal{P}_4$ to

$$\mathcal{P}'_4 = \{\, A'(x) \wedge B'(x) \to C'(x) \vee D'(x) \tag{16}$$

$$E(x) \to A'(x) \vee F'(x) \tag{17}$$

$$C'(x) \wedge R(x,y) \to B'(y) \,\} \cup \mathcal{P}_{\text{aux}} \tag{18}$$

where $\mathcal{P}_{\text{aux}} = \{\, P(x) \to P'(x) \mid P \in \{A, B, C, D, F\} \,\}$ and A', B', C', D', F' are fresh. Rule (16) is not WL in $\mathcal{P}'_4$, and needs to be unfolded. We choose to unfold (16) on $A'(x)$. Thus, in Step 5, Rule (16) is replaced by the rules

$$A(x) \wedge B'(x) \to C'(x) \vee D'(x) \tag{19}$$

$$E(x) \wedge B'(x) \to C'(x) \vee D'(x) \vee F'(x) \tag{20}$$

The resulting $\mathcal{P}''_4$ is WL, and Rewrite returns $\Xi'(\mathcal{P}''_4)$.

6 Application to OWL Ontologies

The RL profile is a fragment of OWL 2 for which reasoning is tractable and practically realisable by means of rule-based technologies. RL is also a fragment of datalog: each RL ontology can be normalised to a datalog program.

1.	$A \sqsubseteq \leq 1\, R.B$	$A(z) \wedge R(z,x_1) \wedge B(x_1) \wedge R(z,x_2) \wedge B(x_2) \to x_1 \approx x_2$
2.	$A \sqcap B \sqsubseteq C$	$A(x) \wedge B(x) \to C(x)$
3.	$\exists R.A \sqsubseteq B$	$R(x,y) \wedge A(y) \to B(x)$
4.	$R \sqsubseteq S$	$R(x_1,x_2) \to S(x_1,x_2)$
5.	$R \circ S \sqsubseteq T$	$R(x_1,z) \wedge S(z,x_2) \to T(x_1,x_2)$
6.	$A \sqsubseteq \mathsf{Self}(R)$	$A(x) \to R(x,x)$
7.	$\mathsf{Self}(R) \sqsubseteq A$	$R(x,x) \to A(x)$
8.	$R \sqsubseteq S^-$	$R(x,y) \to S(y,x)$
9.	$A \sqsubseteq \{a\}$	$A(x) \to x \approx a$
10.	$\{a\} \sqsubseteq A$	$A(a)$
11.	$A \sqsubseteq B \sqcup C$	$A(x) \to B(x) \vee C(x)$

Table 1: Normalised $\text{RL}^{(\sqcup)}$ axioms, with A, B atomic or $\top$, C atomic or $\bot$, R, S, T atomic roles, a an individual.

We next show how to extend RL with disjunctions while retaining tractability of consistency checking and fact entailment in *combined complexity*. We first recapitulate the kinds of normalised axioms that can occur in an RL ontology. We assume familiarity with Description Logic (DL) notation.

A (normalised) *RL ontology* is a finite set of DL axioms of the form 1-10 in Table 1. The table also provides the translation of DL axioms into rules. We define $\text{RL}^{\sqcup}$ as the extension of RL with axioms capturing disjunctive knowledge.

Definition 14. An *$\text{RL}^{\sqcup}$ ontology* is a finite set of DL axioms of the form 1-11 in Table 1. ◇

Fact entailment in $\text{RL}^{\sqcup}$ is co-NP-hard since $\text{RL}^{\sqcup}$ can encode non-3-colourability. Membership in co-NP holds since rules have bounded number of variables, and hence programs can be grounded in polynomial time (see Appendix). Tractability can be regained if we restrict ourselves to $\text{RL}^{\sqcup}$ ontologies corresponding to WL programs. WL programs $\mathcal{P}$ obtained from $\text{RL}^{\sqcup}$ ontologies have bounded number of variables, and thus variables in $\Xi'(\mathcal{P})$ are also bounded.

Theorem 15. *Checking $\mathcal{O} \cup \mathcal{D} \models \alpha$, for $\mathcal{O}$ an $\text{RL}^{\sqcup}$ ontology that corresponds to a WL program, is* PTIME*-complete.*

Thus, fact entailment in $\text{RL}^{\sqcup}$ is no harder than in RL, and one can use scalable engines such as RDFox. Our experiments indicate that many ontologies captured by $\text{RL}^{\sqcup}$ are either WL or can be made WL via unfolding, which makes data reasoning over such ontologies feasible.[2]

Dealing with Expressive Ontology Languages Hustadt, Motik, and Sattler (2007) developed an algorithm for transforming $\mathcal{SHIQ}$ ontologies into an equivalent disjunctive datalog program. Cuenca Grau et al. (2013) combined this algorithm with a knowledge compilation procedure (called Compile-Horn) obtaining a sound but incomplete and non-terminating datalog rewriting procedure for $\mathcal{SHIQ}$. Our procedure Rewrite provides an alternative to Compile-Horn

[2]For CQ answering, our language becomes co-NP-hard w.r.t. data, whereas RL is tractable. This follows from (Lutz and Wolter 2012) already for a single axiom of type 11.

for $\mathcal{SHIQ}$. The classes of ontologies rewritable by the two procedures can be shown incomparable (e.g., Compile-Horn may not terminate on WL programs).

7 Related Work

Complexity of disjunctive datalog with negation as failure has been extensively studied (Ben-Eliyahu-Zohary and Palopoli 1997; Eiter, Gottlob, and Mannila 1997). The class of *head-cycle* free programs was studied in Ben-Eliyahu-Zohary and Palopoli; Ben-Eliyahu-Zohary, Palopoli, and Zemlyanker (1997; 2000), where it was shown that certain reasoning problems are tractable for such programs (fact entailment, however, remains intractable w.r.t. data).

Gottlob et al. (2012) investigated complexity of disjunctive TGDs and showed tractability (w.r.t. data complexity) of fact entailment for a class of linear disjunctive TGDs. Such rules allow for existential quantifiers in the head, but require single-atom bodies; thus, they are incomparable to WL rules. Artale et al. (2009) showed tractability of fact entailment w.r.t. data for DL-Lite$_{\mathsf{bool}}$ logics. This result is related to (Gottlob et al. 2012) since certain DL-Lite$_{\mathsf{bool}}$ logics can be represented as linear disjunctive TGDs. Finally, combined complexity of CQ answering for disjunctive TGDs was studied by Bourhis, Morak, and Pieris (2013).

Lutz and Wolter (2012) investigated non-uniform data complexity of CQ answering w.r.t. extensions of $\mathcal{ALC}$, and related CQ answering to constraint satisfaction problems. This connection was explored by Bienvenu et al. (2013), who showed NEXPTIME-completeness of first-order and datalog rewritability of instance queries for $\mathcal{SHI}$.

The procedure in (Cuenca Grau et al. 2013), mentioned in Section 6, is used by Kaminski and Cuenca Grau (2013) to show first-order/datalog rewritability of two fragments of $\mathcal{ELU}$. Notably, both fragments yield linear programs. Finally, our unfolding-based rewriting procedure is motivated by the work of Afrati, Gergatsoulis, and Toni (2003) on linearisation of plain datalog programs by means of program transformation techniques (Tamaki and Sato 1984; Proietti and Pettorossi 1993; Gergatsoulis 1997).

8 Evaluation

Rewritability Experiments. We have evaluated whether realistic ontologies can be rewritten to WL (and hence to datalog) programs. We analysed 118 non-Horn ontologies from BioPortal, the Protégé library, and the corpus in (Gardiner, Tsarkov, and Horrocks 2006). To transform ontologies into disjunctive datalog we used KAON2 (Motik 2006).[3] KAON2 succeeded to compute disjunctive programs for 103 ontologies. On these, Rewrite succeeded in 35 cases: 8 programs were already datalog after CNF normalisation, 12 were linear, 12 were WL, and 3 required unfolding. Rewrite was limited to 1,000 unfolding steps, but all successful cases required at most 11 steps. On average, 73% of the predicates in ontologies were datalog, and so could be queried using a datalog engine (even if the disjunctive program could not be rewritten). We identified 15 RL$^{\sqcup}$ ontologies and obtained WL programs for 13 of them. For comparison, we implemented the procedure Compile-Horn in (Cuenca Grau et al. 2013), which succeeded on 18 ontologies, only one of which could not be rewritten by our approach.

	Our approach			HermiT			Pellet		
	dlog	disj	err	dlog	disj	err	dlog	disj	err
U01	<1s	8s		6s	107s		146s	172s	
U04	<1s	55s		50s	50s	2	—	—	—
U07	<1s	62s	3	107s	122s	2	—	—	—
U10	<1s	66s	5	176s	182s	2	—	—	—

Table 2: Average query answering times

Query Answering. We tested scalability of instance query answering using datalog programs obtained by our approach. For this, we used UOBM and DBpedia, which come with large datasets. UOBM (Ma et al. 2006) is a standard benchmark for which synthetic data is available (Zhou et al. 2013). We denote the dataset for k universities by Uk. We considered the RL$^{\sqcup}$ subset of UOBM (which is rewritable using Rewrite but not using Compile-Horn), and generated datasets U01, U04, U07, U10. DBpedia[4] is a realistic ontology with a large dataset from Wikipedia. Since DBpedia is Horn, we extended it with reasonable disjunctive axioms. We used RDFox as a datalog engine. Performance was measured against HermiT (Motik, Shearer, and Horrocks 2009) and Pellet (Sirin et al. 2007). We used a server with two Intel Xeon E5-2643 processors and 128GB RAM. Timeouts were 10min for one query and 30min for all queries; a limit of 100GB was allocated to each task. We ran RDFox on 16 threads. Systems were compared on individual queries, and on precomputing answers to all queries. All systems succeeded to answer all queries for U01: HermiT required 890s, Pellet 505s, and we 52s. Table 2 depicts average times for datalog and disjunctive predicates, and number of queries on which a system failed.[5] Pellet only succeeded to answer queries on U01. HermiT's performance was similar for datalog and disjunctive predicates. In our case, queries over the 130 datalog predicates in UOBM (88% of all predicates) were answered instantaneously (<1s); queries over disjunctive predicates were harder, since the rewritings expanded the dataset quadratically in some cases. Finally, due to its size, DBpedia's dataset cannot even be loaded by HermiT or Pellet. Using RDFox, our rewriting precomputed the answers for all DBpedia predicates in 48s.

9 Conclusion

We have proposed a characterisation of datalog rewritability for disjunctive datalog programs, as well as tractable fragments of disjunctive datalog. Our techniques can be applied to rewrite OWL ontologies into datalog, which enables the use of scalable datalog engines for data reasoning. Furthermore, our approach is not "all or nothing": even if an ontology cannot be rewritten, we can still answer queries over most (i.e., datalog) predicates using a datalog reasoner.

[3] We doctored the ontologies to remove constructs outside $\mathcal{SHIQ}$, and hence not supported by KAON2. The modified ontologies can be found on http://csu6325.cs.ox.ac.uk/WeakLinearity/

[4] http://dbpedia.org/About

[5] Average times do not reflect queries on which a system failed.

Acknowledgements

This work was supported by the Royal Society, the EPSRC projects Score!, Exoda, and MaSI^3, and the FP7 project OPTIQUE.

References

[Afrati, Cosmadakis, and Yannakakis 1995] Afrati, F.; Cosmadakis, S. S.; and Yannakakis, M. 1995. On datalog vs. polynomial time. *J. Comput. System Sci.* 51(2):177–196.

[Afrati, Gergatsoulis, and Toni 2003] Afrati, F.; Gergatsoulis, M.; and Toni, F. 2003. Linearisability of datalog programs. *Theor. Comput. Sci.* 308(1-3):199–226.

[Artale et al. 2009] Artale, A.; Calvanese, D.; Kontchakov, R.; and Zakharyaschev, M. 2009. The DL-Lite family and relations. *J. Artif. Intell. Res.* 36:1–69.

[Ben-Eliyahu-Zohary and Palopoli 1997] Ben-Eliyahu-Zohary, R., and Palopoli, L. 1997. Reasoning with minimal models: Efficient algorithms and applications. *Artif. Intell.* 96(2):421–449.

[Ben-Eliyahu-Zohary, Palopoli, and Zemlyanker 2000] Ben-Eliyahu-Zohary, R.; Palopoli, L.; and Zemlyanker, V. 2000. More on tractable disjunctive datalog. *J. Log. Programming* 46(1-2):61–101.

[Bienvenu et al. 2013] Bienvenu, M.; ten Cate, B.; Lutz, C.; and Wolter, F. 2013. Ontology-based data access: A study through disjunctive datalog, CSP, and MMSNP. In *PODS*, 213–224. arXiv:1301.6479.

[Bishop et al. 2011] Bishop, B.; Kiryakov, A.; Ognyanoff, D.; Peikov, I.; Tashev, Z.; and Velkov, R. 2011. OWLim: A family of scalable semantic repositories. *Semantic Web J.* 2(1):33–42.

[Bourhis, Morak, and Pieris 2013] Bourhis, P.; Morak, M.; and Pieris, A. 2013. The impact of disjunction on query answering under guarded-based existential rules. In *IJCAI*.

[Bry et al. 2007] Bry, F.; Eisinger, N.; Eiter, T.; Furche, T.; Gottlob, G.; Ley, C.; Linse, B.; Pichler, R.; and Wei, F. 2007. Foundations of rule-based query answering. In *Reasoning Web*, 1–153.

[Cuenca Grau et al. 2013] Cuenca Grau, B.; Motik, B.; Stoilos, G.; and Horrocks, I. 2013. Computing datalog rewritings beyond Horn ontologies. In *IJCAI*. arXiv:1304.1402.

[Dantsin et al. 2001] Dantsin, E.; Eiter, T.; Gottlob, G.; and Voronkov, A. 2001. Complexity and expressive power of logic programming. *ACM Comput. Surv.* 33(3):374–425.

[Eiter, Gottlob, and Mannila 1997] Eiter, T.; Gottlob, G.; and Mannila, H. 1997. Disjunctive datalog. *ACM Trans. Database Syst.* 22(3):364–418.

[Gardiner, Tsarkov, and Horrocks 2006] Gardiner, T.; Tsarkov, D.; and Horrocks, I. 2006. Framework for an automated comparison of description logic reasoners. In *ISWC*, 654–667.

[Gergatsoulis 1997] Gergatsoulis, M. 1997. Unfold/fold transformations for disjunctive logic programs. *Inf. Process. Lett.* 62(1):23–29.

[Gottlob et al. 2012] Gottlob, G.; Manna, M.; Morak, M.; and Pieris, A. 2012. On the complexity of ontological reasoning under disjunctive existential rules. In *MFCS*, 1–18.

[Hustadt, Motik, and Sattler 2007] Hustadt, U.; Motik, B.; and Sattler, U. 2007. Reasoning in Description Logics by a Reduction to Disjunctive Datalog. *J. Autom. Reasoning* 39(3):351–384.

[Kaminski and Cuenca Grau 2013] Kaminski, M., and Cuenca Grau, B. 2013. Sufficient conditions for first-order and datalog rewritability in ELU. In *DL*, 271–293.

[Leone et al. 2006] Leone, N.; Pfeifer, G.; Faber, W.; Eiter, T.; Gottlob, G.; Perri, S.; and Scarcello, F. 2006. The DLV system for knowledge representation and reasoning. *ACM Trans. Comput. Log.* 7(3):499–562.

[Lutz and Wolter 2012] Lutz, C., and Wolter, F. 2012. Non-uniform data complexity of query answering in description logics. In *KR*.

[Ma et al. 2006] Ma, L.; Yang, Y.; Qiu, Z.; Xie, G. T.; Pan, Y.; and Liu, S. 2006. Towards a complete OWL ontology benchmark. In *ESWC*, 125–139.

[Motik et al. 2009] Motik, B.; Cuenca Grau, B.; Horrocks, I.; Wu, Z.; Fokoue, A.; and Lutz, C. 2009. OWL 2 Web Ontology Language Profiles. *W3C Recommendation*.

[Motik et al. 2014] Motik, B.; Nenov, Y.; Piro, R.; Horrocks, I.; and Olteanu, D. 2014. Parallel materialisation of datalog programs in centralised, main-memory rdf systems. In *AAAI*.

[Motik, Shearer, and Horrocks 2009] Motik, B.; Shearer, R.; and Horrocks, I. 2009. Hypertableau Reasoning for Description Logics. *J. Artif. Intell. Res.* 36:165–228.

[Motik 2006] Motik, B. 2006. *Reasoning in Description Logics using Resolution and Deductive Databases*. Ph.D. Dissertation, Univesität Karlsruhe (TH), Karlsruhe, Germany.

[Proietti and Pettorossi 1993] Proietti, M., and Pettorossi, A. 1993. The loop absorption and the generalization strategies for the development of logic programs and partial deduction. *J. Log. Programming* 16(1):123–161.

[Sirin et al. 2007] Sirin, E.; Parsia, B.; Cuenca Grau, B.; Kalyanpur, A.; and Katz, Y. 2007. Pellet: A practical OWL-DL reasoner. *J. Web Sem.* 5(2):51–53.

[Tamaki and Sato 1984] Tamaki, H., and Sato, T. 1984. Unfold/fold transformation of logic programs. In *ICLP*, 127–138.

[Wu et al. 2008] Wu, Z.; Eadon, G.; Das, S.; Chong, E. I.; Kolovski, V.; Annamalai, M.; and Srinivasan, J. 2008. Implementing an inference engine for RDFS/OWL constructs and user-defined rules in Oracle. In *ICDE*, 1239–1248.

[Zhou et al. 2013] Zhou, Y.; Cuenca Grau, B.; Horrocks, I.; Wu, Z.; and Banerjee, J. 2013. Making the most of your triple store: query answering in OWL 2 using an RL reasoner. In *WWW*, 1569–1580.

A Proofs for Section 3

Definition 16. Let $r = \bigwedge_{i=1}^{n} \beta_i \to \varphi$ be a rule and, for each $1 \leq i \leq n$, let ψ_i be a disjunction of facts $\psi_i = \chi_i \vee \alpha_i$ with α_i a single fact. Let σ be an MGU of each β_i, α_i. Then the following disjunction of facts φ' is a *hyperresolvent* of r and $\psi_1, \ldots, \psi_n$: $\varphi' = \varphi\sigma \vee \chi_1 \vee \cdots \vee \chi_n$.[6] $\diamond$

Definition 17. Let $\mathcal{P}$ be a program, let $\mathcal{D}$ be a dataset, and let φ be a disjunction of facts. A (hyperresolution) *derivation* of φ from $\mathcal{P} \cup \mathcal{D}$ is a pair $\rho = (T, \lambda)$ where T is a tree, λ a labeling function mapping each node in T to a disjunction of facts, and the following properties hold for each $v \in T$:

1. $\lambda(v) = \varphi$ if v is the root;
2. $\lambda(v) \in \mathcal{P} \cup \mathcal{D}$ if v is a leaf; and
3. if v has children $w_1, \ldots, w_n$, then $\lambda(v)$ is a hyperresolvent of a rule $r \in \mathcal{P}$ and $\lambda(w_1), \ldots, \lambda(w_n)$. $\diamond$

We write $\mathcal{P} \cup \mathcal{D} \vdash \varphi$ to denote that φ has a derivation from $\mathcal{P} \cup \mathcal{D}$. Hyperresolution is sound and complete in the following sense: If $\mathcal{P} \cup \mathcal{D}$ is unsatisfiable, then $\mathcal{P} \cup \mathcal{D} \vdash \bot$, and otherwise $\mathcal{P} \cup \mathcal{D} \vdash \alpha$ iff $\alpha \in \mathsf{Eval}(\mathcal{P}, \mathcal{D})$.[7]

Definition 18. Let $\mathcal{P}$ be a (disjunctive) program and $\mathcal{D}$ be a dataset. A $\top$-*stub* is a one-step derivation of a fact $\top(a)$ (for some a) from $\mathcal{D}$ using a rule in $\mathcal{P}_\top$. A derivation $\rho = (T, \lambda)$ from $\mathcal{P} \cup \mathcal{D}$ is *normal* if every node whose label involves $\top$ is the root of a $\top$-stub. $\diamond$

Proposition 19. *Let $\mathcal{P}$ be a disjunctive program, let $\mathcal{D}$ be a dataset, and let φ be a nonempty disjunction of facts. For every derivation of φ from $\mathcal{P} \cup \mathcal{D}$ there is a normal derivation of a nonempty subset of φ from $\mathcal{P} \cup \mathcal{D}$ or a normal derivation of $\bot$ from $\mathcal{P} \cup \mathcal{D}$.*

Proof. Let $\rho = (T, \lambda)$ be a derivation of φ from $\mathcal{P} \cup \mathcal{D}$ and let v be the root of T. We proceed by induction on the size of T. If $\top(a) \in \lambda(v)$ for some a, then a must occur in $\mathcal{P} \cup \mathcal{D}$, and hence we can derive $\top(a)$ in one step with a rule $P(x_1, \ldots, x_n) \to \top(x_i) \in \mathcal{P}_\top$ (if $P(a_1, \ldots, a_{i-1}, a, a_{i+1}, \ldots, a_n) \in \mathcal{D}$) or with the rule $(\to \top(a)) \in \mathcal{P}_\top$ (if a occurs in $\mathcal{P}$).

If $\top$ does not occur in $\lambda(v)$, we proceed as follows. Let $v_1, \ldots, v_n$ be the successors of v in T ($n = 0$ if v is a leaf in T), let $r \in \mathcal{P}$ be the rule used to derive $\lambda(v)$ from $\lambda(v_1), \ldots, \lambda(v_n)$, and let σ be the substitution used in the corresponding hyperresolution step. By the inductive hypothesis, for every $i \in [1, n]$ there is some ψ_i such that ψ_i is a nonempty subset of $\lambda(v_i)$ or $\psi_i = \bot$ and ψ_i has a normal derivation from $\mathcal{P} \cup \mathcal{D}$. W.l.o.g., let $\psi_i \neq \bot$ for every $i \in [1, n]$ (otherwise, the claim is immediate). We then distinguish two cases. If r applies to $\psi_1, \ldots, \psi_n$ with substitution σ, then the hyperresolvent φ' of r and $\psi_1, \ldots, \psi_n$ is a nonempty subset of φ (φ' is nonempty since the only rule in $\mathcal{P}$ with an empty head is $(\bot \to)$ but, by assumption, $\varphi_1 \neq \bot$). If r does not apply to $\psi_1, \ldots, \psi_n$ with substitution σ, the claim follows since, for some i, we have $\psi_i \subseteq \varphi$. $\square$

Proposition 19 allows us to consider only normal derivations. In the following, without loss of generality, we assume every derivation to be normal.

Proposition 20. *Let $\mathcal{P}$ be a disjunctive program, let $\mathcal{D}$ be a dataset, let $\rho = (T, \lambda)$ be a derivation from $\mathcal{P} \cup \mathcal{D}$, and let v be a node in T. If $\lambda(v)$ contains an EDB atom, then $\lambda(v)$ is a singleton.*

Proof. The claim follows since whenever $\lambda(v)$ contains an EDB atom, we either have that v is a leaf in T, and thus $\lambda(v) \in \mathcal{D}$, or that v is the root of a $\top$-stub (since ρ is implicitly assumed to be normal), and thus $\lambda(v) = \top(a)$ for some individual a. $\square$

Definition 21. A nonempty tree $T' = (V', E')$ is an *upper portion* of a tree $T = (V, E)$ if the following conditions hold:

- $V' \subseteq V$ and E' is the restriction of E to V'.
- T and T' have the same root.
- If v is an internal node in T', then every child of v in T is contained in V'.

Let $\mathcal{P}$ be a (disjunctive) datalog program, $\mathcal{D}$ a dataset, and $\rho = (T, \lambda)$ a derivation of a fact $P(\vec{a})$ from $\mathcal{P} \cup \mathcal{D}$ where $P \neq \top$. An *upper portion* of ρ is a pair $\rho' = (T', \lambda')$ such that:

- T' is an upper portion of T;
- λ' is the restriction of λ to the nodes in T';
- If $\lambda'(v) = \top(b)$ for some $v \in T'$ and some individual b, then v is a leaf in T'. $\diamond$

Theorem 2. *If $\mathcal{P}$ is linear, then $\Xi(\mathcal{P})$ is a polynomial datalog rewriting of $\mathcal{P}$.*

Proof. Analogous to the proof of Theorem 10 in Appendix B (but simpler). $\square$

[6] We view disjunctions as sets of formulae.

[7] This implies that $\mathcal{P} \cup \mathcal{D} \vdash \bot$ iff $\mathsf{Eval}(\mathcal{P}, \mathcal{D}) = \{\bot\}$.

Lemma 22. *Let $\mathcal{P}$ be a datalog program, let $\mathcal{D}$ be a dataset, let P be an IDB predicate in $\mathcal{P}^e$, and let $\rho = (T, \lambda)$ be a derivation of a fact $P(\vec{a})$ from $\mathcal{P}^e \cup \mathcal{D}$. Given a tree T', let $\mathsf{leaves}(T')$ be the set of leaves of T' and, given a set of nodes S, let $\lambda(S) = \bigcup\{\, \lambda(t) \mid t \in S \,\}$. For every upper portion $\rho' = (T', \lambda')$ of ρ we have $\Psi(\mathcal{P}) \cup \mathcal{D} \models \bigvee_{Q(\vec{b}) \in \lambda(\mathsf{leaves}(T'))} Q^P(\vec{b}, \vec{a})$.*

Proof. Let $\rho' = (T', \lambda')$ be an upper portion of ρ and let v be the root of T'. In particular, we have $\lambda(v) = P(\vec{a})$. We proceed by induction on the size of T'. If v is the only node in T', the claim reduces to $\Psi(\mathcal{P}) \cup \mathcal{D} \models P^P(\vec{a}, \vec{a})$. This follows since $\varphi_\top \to P^P(\vec{x}, \vec{x}) \in \Psi(\mathcal{P})$ and $\Psi(\mathcal{P})_\top \cup \mathcal{D} \models \varphi_\top(\vec{a})$.

Now suppose T' contains more than one node and let $\mathsf{leaves}(T') = \{v_1, \ldots, v_n\}$. Since T' is a tree, it has a node w of height 1. W.l.o.g., let $v_1, \ldots, v_k$ ($1 \leq k \leq n$) be the children of w in T', let $r = \bigwedge_{i=1}^k Q_i(\vec{s}_i) \to R(\vec{t}) \in \mathcal{P}^e$ be the rule used to derive $\lambda(w)$ from $\lambda(s_1), \ldots, \lambda(s_k)$, and let σ be the substitution used in the corresponding hyperresolution step. Since w is an internal node in T', we have $R \neq \top$, and hence $r \notin \mathcal{P}^e_\top$. Then:

1. $S = \{w, v_{k+1}, \ldots, v_n\}$ is the set of leaves of an upper portion of ρ that is strictly smaller than ρ';
2. $\lambda(w) = R(\vec{t}\sigma)$ and $\lambda(v_i) = Q(\vec{s}_i\sigma)$ for every $i \in [1, k]$;
3. $(\lambda(S) \setminus \{R(\vec{t}\sigma)\}) \cup \{Q_1(\vec{s}_1\sigma), \ldots, Q_k(\vec{s}_k\sigma)\} \subseteq \lambda(\mathsf{leaves}(T'))$;
4. $R^P(\vec{t}, \vec{y}) \to \bigvee_{i=1}^k Q_i^P(\vec{s}_i, \vec{y}) \in \Psi(\mathcal{P})$.

By (1), (2), and the inductive hypothesis, $\Psi(\mathcal{P}) \cup \mathcal{D} \models R^P(\vec{t}\sigma, \vec{a}) \vee \bigvee_{Q(\vec{b}) \in \lambda(\{v_{k+1}, \ldots, v_n\})} Q^P(\vec{b}, \vec{a})$. Hence by (3), it suffices to show $\Psi(\mathcal{P}) \cup \mathcal{D} \models R^P(\vec{t}\sigma, \vec{a}) \to \bigvee_{i=1}^k Q_i^P(\vec{s}_i\sigma, \vec{a})$, which follows by (4). □

Lemma 23. *Let $\mathcal{P}$ be a datalog program. For every dataset $\mathcal{D}$ over the signature of $\mathcal{P}$ and every fact α such that $\mathcal{P}^e \cup \mathcal{D} \models \alpha$ we have $\Psi(\mathcal{P}) \cup \mathcal{D} \models \alpha$.*

Proof. Let $\mathcal{D}$ be a dataset and $P(\vec{a})$ a fact such that $\mathcal{P}^e \cup \mathcal{D} \models P(\vec{a})$. W.l.o.g., let P be IDB in $\mathcal{P}^e$. We show $\Psi(\mathcal{P}) \cup \mathcal{D} \models P(\vec{a})$. By completeness of hyperresolution, $P(\vec{a})$ has a derivation $\rho = (T, \lambda)$ from $\mathcal{P}^e \cup \mathcal{D}$. Let $\rho' = (T', \lambda')$ be the largest upper portion of ρ. By Lemma 22, $\Psi(\mathcal{P}) \cup \mathcal{D} \models \bigvee_{Q(\vec{b}) \in \lambda(\mathsf{leaves}(T'))} Q^P(\vec{b}, \vec{a})$. Therefore, it suffices to show that $\Psi(\mathcal{P}) \cup \mathcal{D} \cup \{Q^P(\vec{b}, \vec{a})\} \models P(\vec{a})$ for every $Q(\vec{b}) \in \lambda(\mathsf{leaves}(T'))$. Since ρ' is maximal, we distinguish the following three cases for $Q(\vec{b})$:

- $Q(\vec{b}) \in \mathcal{D}$. Then Q is EDB in $\mathcal{P}^e$ (since $\mathcal{D}$ only contains facts about predicates in $\mathcal{P}$ and every predicate in $\mathcal{P}$ is EDB in $\mathcal{P}^e$). Hence $Q(\vec{z}) \wedge Q^P(\vec{z}, \vec{y}) \to P(\vec{y}) \in \Psi(\mathcal{P})$. The claim follows.
- $Q(\vec{b})$ is ground and $(\to Q(\vec{b})) \in \mathcal{P}^e \setminus \mathcal{P}^e_\top$. Hence $Q^P(\vec{b}, \vec{y}) \to \bot \in \Psi(\mathcal{P})$, and consequently $\Psi(\mathcal{P}) \cup \mathcal{D} \cup \{Q^P(\vec{b}, \vec{a})\} \models \bot$. The claim follows.
- $Q(\vec{b}) = \top(b)$. Then the claim follows since $\top(z) \wedge \top^P(z, \vec{y}) \to P(\vec{y}) \in \Psi(\mathcal{P})$ and $\Psi(\mathcal{P}) \cup \mathcal{D} \models \top(b)$. □

Definition 24. Let $\mathcal{P}$ be a datalog program and let $Q(\vec{b})$ be a fact where Q is IDB in $\mathcal{P}$. A disjunction φ of facts is *focused on* $Q(\vec{b})$ w.r.t. $\mathcal{P}$ if every disjunct $\alpha \in \varphi$ has one of the following forms:

- $\alpha = P(\vec{a})$ where P is EDB in $\mathcal{P}$;
- $\alpha = \bot$;
- $\alpha = Q(\vec{b})$;
- $\alpha = P^Q(\vec{a}, \vec{b})$ for some P and $\vec{a}$.

Let $\rho = (T, \lambda)$ be a derivation (not necessarily from $\mathcal{P}$). We call ρ *focused on* $Q(\vec{b})$ w.r.t. $\mathcal{P}$ if so is the label of every node in T. Given a node $v \in T$, we define $\lambda_{\text{base}}(v) := \{\, P(\vec{a}) \mid P^Q(\vec{a}, \vec{b}) \in \lambda(v) \,\}$. ◇

Lemma 25. *Let $\mathcal{P}$ be a datalog program and let $\mathcal{D}$ be a dataset over the signature of $\mathcal{P}$. Every derivation from $\Psi(\mathcal{P}) \cup \mathcal{D}$ is focused on some fact α w.r.t. $\mathcal{P}^e$.*

Proof. Let $\rho = (T, \lambda)$ be a derivation from $\Psi(\mathcal{P}) \cup \mathcal{D}$ and let v be the root of T. We show that ρ is focused on some α by induction on the size of T. If v is the only node in T, we distinguish the following cases:

- $\lambda(v) \in \mathcal{D}$. Then $\lambda(v) = P(\vec{a})$ where P is EDB in $\mathcal{P}^e$, and thus ρ is focused on every IDB predicate in $\mathcal{P}^e$.
- $\mathsf{arity}(Q) = 0$ and $\lambda(v)$ is obtained by a rule of the form $(\to \bigvee_{i=1}^n P_i^Q(\vec{s}_i)) \in \Psi(\mathcal{P})$ for some IDB predicate Q in $\mathcal{P}^e$. Then $\lambda(v) = \bigvee_{i=1}^n P_i^Q(\vec{s}_i)$, meaning ρ is focused on Q.

If v has successors $v_1, \ldots, v_m$, we distinguish the following cases depending on the shape of the rule $r \in \Psi(\mathcal{P})$ used to obtain $\lambda(v)$ from $\lambda(v_1), \ldots, \lambda(v_m)$.

- $r \in \Psi(\mathcal{P})_\top$. Then $m = 1$ and $\lambda(v) = \top(a)$ for some a. By the inductive hypothesis, $\lambda(v_1)$ is focused on a fact w.r.t. $\mathcal{P}^e$. The claim follows since $\top$ is EDB in $\mathcal{P}^e$.
- $r = \varphi_\top \wedge P^Q(\vec{t}, \vec{y}) \to \bigvee_{i=1}^n R^Q(\vec{s}_i, \vec{y})$. Let σ be the substitution used in the hyperresolution step. W.l.o.g., let $P^Q(\vec{t}\sigma, \vec{y}\sigma) \in \lambda(v_1)$. Then, by Proposition 20, $\lambda(v_j) \in \varphi_\top$ for $j \in [2, m]$. Hence, $\lambda(v) \subseteq \lambda(v_1) \cup \{\bot\}$. The claim follows since, by the inductive hypothesis, the subderivation rooted at v_1 is focused on $Q(\vec{b})$ (so, in particular, $\vec{y}\sigma = \vec{b}$), and $v_2, \ldots, v_m$ are focused on every IDB predicate in $\mathcal{P}^e$.
- $r = (\bot \to)$ or $r = \varphi_\top \to \bigvee_{i=1}^n R_i^Q(\vec{s}_i, \vec{b})$ for some R and $\vec{s}_1, \ldots, \vec{s}_n$. In both cases, the argument proceeds analogously to the preceding case (but simpler).
- $r = P(\vec{z}) \wedge P^Q(\vec{z}, \vec{y}) \to Q(\vec{y})$. Then $m = 2$. Let σ be the substitution used in the hyperresolution step and, w.l.o.g., let $P^Q(\vec{z}\sigma, \vec{y}\sigma) \in \lambda(v_1)$. Then, by Proposition 20, $\lambda(v_2) = P(\vec{z}\sigma)$. Hence, $\lambda(v) \subseteq \lambda(v_1) \cup \{Q(\vec{y}\sigma)\}$. The claim follows since, by the inductive hypothesis, the subderivation rooted at v_1 is focused on $Q(\vec{b})$ and $\lambda(v_2)$ is focused on every IDB predicate in $\mathcal{P}^e$. □

Since the root of a derivation of a fact α has to be labeled with α, Lemma 25 implies the following corollary.

Corollary 26. *Let $\mathcal{P}$ be a datalog program, let $\mathcal{D}$ be a dataset over the signature of $\mathcal{P}$, and let $Q(\vec{b})$ be a fact where Q is IDB in $\mathcal{P}^e$. Every derivation of $Q(\vec{b})$ from $\Psi(\mathcal{P}) \cup \mathcal{D}$ is focused on $Q(\vec{b})$ w.r.t. $\mathcal{P}^e$.*

Lemma 27. *Let $\mathcal{P}$ be a datalog program, let $\mathcal{D}$ be a dataset over the signature of $\mathcal{P}$, and let $\rho = (T, \lambda)$ be a derivation of a fact $Q(\vec{b})$ from $\Psi(\mathcal{P}) \cup \mathcal{D}$, where Q is IDB in $\mathcal{P}^e$. For every node v in T whose label contains an IDB atom in $\Psi(\mathcal{P})$, we have $\mathcal{P}^e \cup \mathcal{D} \models (\bigwedge_{\alpha \in \lambda_{\text{base}}(v)} \alpha) \to Q(\vec{b})$.*

Proof. We proceed by induction on the height of v in T. Let $v_1, \ldots, v_m$ be the successors of v in T (where $m = 0$ if v is a leaf in T). If $Q(\vec{b}) \in \mathcal{D}$, the claim is vacuous since $\mathcal{D}$ contains only facts about predicates in $\mathcal{P}$, every predicate in $\mathcal{P}$ is EDB in $\mathcal{P}^e$, and every EDB predicate in $\mathcal{P}^e$ is EDB in $\Psi(\mathcal{P})$. Otherwise, we distinguish the following cases depending on the shape of the rule $r \in \Psi(\mathcal{P}) \setminus \Psi(\mathcal{P})_\top$ used to obtain $\lambda(v)$ from $\lambda(v_1), \ldots, \lambda(v_m)$ (by Corollary 26, we only need to consider cases that can occur in a derivation focused on $Q(\vec{b})$ w.r.t. $\mathcal{P}^e$):

- $r = \varphi_\top \wedge P^Q(\vec{t}, \vec{y}) \to \bigvee_{i=1}^n R_i^Q(\vec{s}_i, \vec{y})$ such that $r' = \bigwedge_{i=1}^n R_i(\vec{s}_i) \to P(\vec{t}) \in \mathcal{P}^e$. Let σ be the substitution used in the corresponding hyperresolution step and let, w.l.o.g., $P^Q(\vec{t}\sigma, \vec{y}\sigma) \in \lambda(v_1)$. Then, by the inductive hypothesis, $\mathcal{P}^e \cup \mathcal{D} \models (\bigwedge_{\alpha \in \lambda_{\text{base}}(v_1)} \alpha) \to Q(\vec{b})$. Moreover, $(\lambda(v_1) \setminus \{P^Q(\vec{t}\sigma, \vec{y}\sigma)\}) \cup \bigcup_{i=1}^n R_i^Q(\vec{s}_i\sigma, \vec{y}\sigma) \subseteq \lambda(v)$. The claim follows since, by r', $\mathcal{P}^e \models (\bigwedge_{\alpha \in \lambda_{\text{base}}(v)} \alpha) \to P(\vec{t}\sigma)$.
- $r = \varphi_\top \to \bigvee_{i=1}^n R_i^Q(\vec{s}_i, \vec{y})$ such that $r' = \bigwedge_{i=1}^n R_i(\vec{s}_i) \to \bot \in \mathcal{P}^e$. Let σ be the substitution used in the corresponding hyperresolution step. Then $\bigcup_{i=1}^n R_i^Q(\vec{s}_i\sigma, \vec{y}\sigma) \subseteq \lambda(v)$, and hence, by r', $\mathcal{P}^e \models (\bigwedge_{\alpha \in \lambda_{\text{base}}(v)} \alpha) \to \bot$. The claim follows.
- $r = \varphi_\top \to Q^Q(\vec{y}, \vec{y})$. Since ρ is focused on $Q(\vec{b})$, we have $\lambda(v) = Q^Q(\vec{b}, \vec{b})$, and the claim ($\mathcal{P}^e \cup \mathcal{D} \models Q(\vec{b}) \to Q(\vec{b})$) is immediate.
- $r = P(\vec{z}) \wedge P^Q(\vec{z}, \vec{y}) \to Q(\vec{y})$ for some EDB predicate P in $\mathcal{P}^e$. Let σ be the substitution used in the corresponding hyperresolution step (in particular, $\vec{y}\sigma = \vec{b}$). Let, w.l.o.g., $P^Q(\vec{z}\sigma, \vec{b}) \in \lambda(v_1)$ and $\lambda(v_2) = P(\vec{z}\sigma)$ (Proposition 20). Since P is EDB in $\mathcal{P}^e$ and hence in $\Psi(\mathcal{P})$, we have $P(\vec{z}\sigma) \in \mathcal{D}$. By the inductive hypothesis, $\mathcal{P}^e \cup \mathcal{D} \models (\bigwedge_{\alpha \in \lambda_{\text{base}}(v_1)} \alpha) \to Q(\vec{b})$, and therefore $\mathcal{P}^e \cup \mathcal{D} \models (\bigwedge_{\alpha \in \lambda_{\text{base}}(v_1) \setminus \{P(\vec{z}\sigma)\}} \alpha) \to Q(\vec{b})$. The claim follows since $\lambda(v_1) \setminus \{P^Q(\vec{z}\sigma, \vec{b})\} \subseteq \lambda(v)$. □

By completeness of hyperresolution and the observation that $\lambda_{\text{base}}(v) = \emptyset$ whenever $\lambda(v)$ contains no facts of the form $P^Q(\vec{a})$, we obtain the following corollary.

Corollary 28. *Let $\mathcal{P}$ be a datalog program. For every dataset $\mathcal{D}$ over the signature of $\mathcal{P}$ and every atom α over the signature of $\mathcal{P}^e$ such that $\Psi(\mathcal{P}) \cup \mathcal{D} \models \alpha$ we have $\mathcal{P}^e \cup \mathcal{D} \models \alpha$.*

Theorem 5. *If $\mathcal{P}$ is datalog, then $\Psi(\mathcal{P})$ is a polynomial rewriting of $\mathcal{P}$ into a linear disjunctive program.*

Proof. By construction, $\Psi(\mathcal{P})$ is a linear disjunctive program of size quadratic in the size of $\mathcal{P}$. Since $\mathcal{P}^e$ is a rewriting of $\mathcal{P}$, to prove that $\Psi(\mathcal{P})$ is a rewriting of $\mathcal{P}$ it suffices to show that $\mathsf{Eval}(\mathcal{P}^e, \mathcal{D})|_S = \mathsf{Eval}(\Psi(\mathcal{P}), \mathcal{D})|_S$ for every dataset $\mathcal{D}$ over the signature of $\mathcal{P}$ and every set S of predicates in $\mathcal{P}^e$. This follows by Lemma 23 and Corollary 28. □

B Proofs for Section 4

We begin by generalising Proposition 20 as follows.

Proposition 29. *Let $\mathcal{P}$ be a disjunctive program, let $\mathcal{D}$ be a dataset, let $\rho = (T, \lambda)$ be a derivation from $\mathcal{P} \cup \mathcal{D}$, and let v be a node in T. If $\lambda(v)$ contains a datalog atom, then $\lambda(v)$ is a singleton.*

Proof. Straightforward induction on the height of v in ρ. The case where $\lambda(v)$ contains an atom of the form $\top(a)$ follows by the implicit assumption that ρ is normal. □

Proposition 30. *Let $\mathcal{P}$ be a disjunctive program, let Q be a datalog predicate in $\mathcal{P}$, and let $\mathcal{P}_Q$ be the set of all rules in $\mathcal{P}$ on which Q depends. For every dataset $\mathcal{D}$ and vector of individuals $\vec{a} = a_1 \dots a_{\mathsf{arity}(Q)}$: $\mathcal{P} \cup \mathcal{D} \vdash Q(\vec{a})$ iff $\mathcal{P}_Q \cup \mathcal{D} \vdash Q(\vec{a})$.*

Proof. The inclusion from right to left is immediate. The inclusion from left to right follows by a straightforward induction on the derivation of a fact $Q(\vec{a})$ from $\mathcal{P} \cup \mathcal{D}$ exploiting the implicit normality assumption. □

Since datalog predicates only depend on rules that contain no disjunctive predicates, and a WL program $\mathcal{P}$ coincides with $\Xi'(\mathcal{P})$ on rules that contain no disjunctive predicates, we obtain (by correctness of hyperresolution):

Corollary 31. *Let $\mathcal{P}$ be a WL program and let Q be a datalog predicate in $\mathcal{P}$. For every dataset $\mathcal{D}$ and vector of individuals $\vec{a} = a_1 \dots a_{\mathsf{arity}(Q)}$: $\mathcal{P} \cup \mathcal{D} \models Q(\vec{a})$ iff $\Xi'(\mathcal{P}) \cup \mathcal{D} \models Q(\vec{a})$.*

Lemma 32. *Let $\mathcal{P}$ be a WL program, let $\mathcal{D}$ be a dataset, let P be a disjunctive predicate in $\mathcal{P}$, and let $\rho = (T, \lambda)$ be a derivation of a fact $P(\vec{a})$ from $\mathcal{P} \cup \mathcal{D}$. Then for every node $v \in T$ in an upper portion of ρ and every disjunct $Q(\vec{b}) \in \lambda(v)$ where Q is disjunctive in $\mathcal{P}$, we have $\Xi'(\mathcal{P}) \cup \mathcal{D} \models Q^P(\vec{b}, \vec{a})$.*

Proof. Let $v \in T$ and $Q(\vec{b}) \in \lambda(v)$ be as required. We show the claim by induction on the distance of v from the root of T. If v is the root of T, then $Q(\vec{b}) = P(\vec{a})$ and the claim ($\Xi'(\mathcal{P}) \cup \mathcal{D} \models P^P(\vec{a}, \vec{a})$) follows since $\top(y_1) \wedge \dots \wedge \top(y_{\mathsf{arity}(P)}) \to P^P(\vec{y}, \vec{y}) \in \Xi'(\mathcal{P})$ and $\Xi'(\mathcal{P})_\top \cup \mathcal{D} \models \top(a_i)$ for every $a_i \in \vec{a}$.

If v is not the root of T, then it must have a predecessor w and siblings $v_1, \dots, v_n$ ($n \geq 0$) in T such that either (a) $Q(\vec{b}) \in \lambda(w)$ or (b) $\lambda(w)$ is a hyperresolvent of $\lambda(v), \lambda(v_1), \dots, \lambda(v_n)$ and some rule $Q(\vec{s}) \wedge \bigwedge_{i=1}^n R_i(\vec{s}_i) \to \bigvee_{j=1}^m S_j(\vec{t}_j) \in \mathcal{P} \setminus \mathcal{P}_\top$, where the atom $Q(\vec{s})$ is resolved with $\lambda(v)$ and the atoms $R_i(\vec{s}_i)$ are resolved with $\lambda(v_i)$. If $Q(\vec{b}) \in \lambda(w)$, the claim follows by the inductive hypothesis so, w.l.o.g., suppose we are in Case (b). Since, by assumption, Q is disjunctive and $\mathcal{P}$ is WL, all R_i are datalog. Hence, $\Xi'(\mathcal{P})$ contains a rule $r = \varphi_\top \wedge (\bigwedge_{j=1}^m S_j^P(\vec{t}_j, \vec{y})) \wedge \bigwedge_{i=1}^n R_i(\vec{s}_i) \to Q^P(\vec{s}, \vec{y})$. Let σ be the substitution used in the hyperresolution step deriving $\lambda(w)$. Then $\vec{s}\sigma = \vec{b}$, $\lambda(v_i) = R_i(\vec{s}_i\sigma)$ for every $i \in [1, n]$ (by Proposition 29), and $\bigvee_{j=1}^m S_j(\vec{t}_j\sigma) \subseteq \lambda(w)$. By the inductive hypothesis, we then have $\Xi'(\mathcal{P}) \cup \mathcal{D} \models S_j^P(\vec{t}_j\sigma, \vec{a})$ for every $j \in [1, m]$. Moreover, by Corollary 31, $\Xi'(\mathcal{P}) \cup \mathcal{D} \models R_i(\vec{s}_i\sigma)$ for every $i \in [1, n]$. Finally, we have $\Xi'(\mathcal{P})_\top \cup \mathcal{D} \models \varphi_\top\sigma$. The claim follows with r. □

Lemma 33. *Let $\mathcal{P}$ be a WL program. For every dataset $\mathcal{D}$ and every fact α such that $\mathcal{P} \cup \mathcal{D} \models \alpha$ we have $\Xi'(\mathcal{P}) \cup \mathcal{D} \models \alpha$.*

Proof. Let $\mathcal{P} \cup \mathcal{D} \models P(\vec{a})$. We show that $\Xi'(\mathcal{P}) \cup \mathcal{D} \models P(\vec{a})$. W.l.o.g., $P(\vec{a}) \notin \mathcal{D}$ (otherwise, the claim is trivial) and P is disjunctive (otherwise, the claim follows by Corollary 31). By completeness of hyperresolution, there is a derivation $\rho = (T, \lambda)$ of $P(\vec{a})$ from $\mathcal{P} \cup \mathcal{D}$. Since $P(\vec{a}) \notin \mathcal{D}$ and P is disjunctive, there is an upper portion ρ' of ρ and a node v in ρ' such that:

1. $\lambda(v)$ contains a disjunctive predicate;
2. v has no successor w in T such that $\lambda(w)$ contains a disjunctive predicate.

We distinguish two cases. If $\lambda(v) \in \mathcal{D}$, then $\lambda(v) = Q(\vec{b})$ for some Q and $\vec{b}$. By Lemma 32, we have $\Xi'(\mathcal{P}) \cup \mathcal{D} \models Q^P(\vec{b}, \vec{a})$. The claim follows since $Q(\vec{z}) \wedge Q^P(\vec{z}, \vec{y}) \to P(\vec{y}) \in \Xi'(\mathcal{P})$.

If $\lambda(v) \notin \mathcal{D}$, then v has successors $v_1, \dots, v_n$ ($n \geq 0$) in T such that $\lambda(v)$ is a hyperresolvent of $\lambda(v_1), \dots, \lambda(v_n)$ and a rule in $\mathcal{P} \setminus \mathcal{P}_\top$ of the form $\bigwedge_{i=1}^n R_i(\vec{s}_i) \to \bigvee_{j=1}^m S_j(\vec{t}_j)$, where the atoms $R_i(\vec{s}_i)$ are resolved with $\lambda(v_i)$. Since, by assumption, all R_i are datalog, $\Xi'(\mathcal{P})$ contains a rule $r = (\bigwedge_{j=1}^m S_j^P(\vec{t}_j, \vec{y})) \wedge \bigwedge_{i=1}^n R_i(\vec{s}_i) \to P(\vec{y})$. Let σ be the substitution used in the hyperresolution step deriving $\lambda(v)$. By Lemma 32, we then have $\Xi'(\mathcal{P}) \cup \mathcal{D} \models S_j^P(\vec{t}_j\sigma, \vec{a})$ for every $j \in [1, m]$. By Proposition 29, we have $\lambda(v_i) = R_i(\vec{s}_i\sigma)$, and hence, by Corollary 31, $\Xi'(\mathcal{P}) \cup \mathcal{D} \models R_i(\vec{s}_i\sigma)$ for every $i \in [1, n]$. The claim follows with r. □

Lemma 34. *Let $\mathcal{P}$ be a WL program, let $\mathcal{D}$ be a dataset over the signature of $\mathcal{P}$, and let $\rho = (T, \lambda)$ be a derivation of a fact α from $\Xi'(\mathcal{P}) \cup \mathcal{D}$ where α is not of the form $\top(a)$. Then:*

1. *For every $v \in T$, $\lambda(v) = P(\vec{a})$ where P occurs in $\mathcal{P}$, or $\lambda(v) = P^Q(\vec{a}, \vec{b})$ where P, Q are disjunctive in $\mathcal{P}$.*
2. *If $\alpha = P(\vec{a})$ where P occurs in $\mathcal{P}$, then $\mathcal{P} \cup \mathcal{D} \models P(\vec{a})$.*

3. If $\alpha = P^Q(\vec{a}, \vec{b})$, then $\mathcal{P} \cup \mathcal{D} \models P(\vec{a}) \to Q(\vec{b})$.

Proof. We begin by showing (1). Since $\Xi'(\mathcal{P})$ is datalog, $\lambda(v)$ contains only one atom for every $v \in T$. The claim follows since $\mathcal{D}$ contains only predicates in $\mathcal{P}$ and the rules of $\Xi'(\mathcal{P})$ can only infer facts of the form $P(\vec{a})$ where P occurs in $\mathcal{P}$ or $P^Q(\vec{a}, \vec{b})$ where P, Q are disjunctive in $\mathcal{P}$.

We now show (2) and (3) by simultaneous induction on the height n of T. If $n = 0$, we distinguish three cases:

- $\alpha \in \mathcal{D}$. Then $\mathcal{D} \models \alpha$ and the claim is immediate.
- $\alpha = P(\vec{a})$ where P is datalog in $\mathcal{P}$ and $r = (\to P(\vec{a})) \in \Xi'(\mathcal{P})$. Then $r \in \mathcal{P}$ and the claim is immediate.
- $\alpha = \bot^Q$, $\mathsf{arity}(Q) = 0$, and $(\to \bot^Q) \in \Xi'(\mathcal{P})$. Then, since $(\bot \to) \in \mathcal{P}$, we have $\mathcal{P} \models \bot \to Q$ for every predicate Q.
- $\alpha = P^P$ where P is disjunctive in $\mathcal{P}$ and $\mathsf{arity}(P) = 0$. The claim ($\mathcal{P} \cup \mathcal{D} \models P \to P$) is immediate.

If $n > 0$, the root v of T has children $v_1, \dots, v_n$ and α is a hyperresolvent of $\lambda(v_1), \dots, \lambda(v_n)$ and a rule $r \in \Xi'(\mathcal{P}) \setminus \Xi'(\mathcal{P})_\top$. We distinguish five cases:

- r contains no disjunctive predicates. Then α is a datalog atom and the claim follows by Corollary 31.
- $r = \varphi_\top \to P^P(\vec{y}, \vec{y})$ where P is disjunctive in $\mathcal{P}$. Then $\alpha = P^P(\vec{a}, \vec{a})$ for some $\vec{a}$, and the claim ($\mathcal{P} \cup \mathcal{D} \models P(\vec{a}) \to P(\vec{a})$) is immediate.
- $r = Q(\vec{z}) \wedge Q^P(\vec{z}, \vec{y}) \to P(\vec{y})$. Then $\alpha = P(\vec{a})$ for some $\vec{a}$. By the Corollary 31, we have $\mathcal{P} \cup \mathcal{D} \models Q(\vec{b})$, and by the inductive hypothesis, $\mathcal{P} \cup \mathcal{D} \models Q(\vec{b}) \to P(\vec{a})$ for some $\vec{b}$. Hence $\mathcal{P} \cup \mathcal{D} \models P(\vec{a})$.
- $r = \varphi_\top \wedge \varphi \wedge \bigwedge_{i=1}^n R_i^Q(\vec{s}_i, \vec{y}) \to P^Q(\vec{t}, \vec{y})$ where φ is the conjunction of all datalog atoms in r and $r' = \varphi \wedge P(\vec{t}) \to \bigvee_{i=1}^n R_i(\vec{s}_i) \in \mathcal{P}$. Then $\alpha = P^Q(\vec{a}, \vec{b})$ for some $\vec{a}$ and $\vec{b}$. By Corollary 31, for every $i \in [1, n]$ there is some $\vec{c}_i$ such that $\mathcal{P} \cup \mathcal{D} \models \varphi|^{\vec{a}\vec{b}\vec{c}_1 \dots \vec{c}_n}_{\vec{t}\vec{y}\vec{s}_1 \dots \vec{s}_n}$, and by the inductive hypothesis, $\mathcal{P} \cup \mathcal{D} \models R_i(\vec{c}_i) \to Q(\vec{b})$. With r', we obtain $\mathcal{P} \cup \mathcal{D} \models P(\vec{a}) \to Q(\vec{b})$.
- $r = \varphi \wedge \bigwedge_{i=1}^n R_i^P(\vec{s}_i, \vec{y}) \to P(\vec{y})$ where φ is the conjunction of all datalog atoms in r and $r' = \varphi \to \bigvee_{i=1}^n R_i(\vec{s}_i) \in \mathcal{P}$. Then $\alpha = P(\vec{a})$ for some $\vec{a}$. By Corollary 31, for every $i \in [1, n]$ there is some $\vec{b}_i$ such that $\mathcal{P} \cup \mathcal{D} \models \varphi|^{\vec{a}\vec{b}_1 \dots \vec{b}_n}_{\vec{y}\vec{s}_1 \dots \vec{s}_n}$, and by the inductive hypothesis, $\mathcal{P} \cup \mathcal{D} \models R_i(\vec{b}_i) \to P(\vec{a})$. With r', we obtain $\mathcal{P} \cup \mathcal{D} \models P(\vec{a})$. □

By completeness of hyperresolution (and Corollary 31 for facts of the form $\top(a)$), Lemma 34(2) implies:

Corollary 35. *Let $\mathcal{P}$ be a WL program. For every dataset $\mathcal{D}$ and atom α over the signature of $\mathcal{P}$ such that $\Xi'(\mathcal{P}) \cup \mathcal{D} \models \alpha$ we have $\mathcal{P} \cup \mathcal{D} \models \alpha$.*

Theorem 10. *If $\mathcal{P}$ is WL, then $\Xi'(\mathcal{P})$ is a polynomial datalog rewriting of $\mathcal{P}$.*

Proof. By construction, $\Xi'(\mathcal{P})$ is a datalog program of size quadratic in the size of $\mathcal{P}$. Correctness of the transformation (i.e., $\Xi'(\mathcal{P})$ being a rewriting of $\mathcal{P}$) follows with Lemma 33 and Corollary 35. □

Theorem 11. *Let $\mathcal{P}$ be WL, S a set of predicates in $\mathcal{P}$, and $\mathcal{P}'$ obtained from $\Xi'(\mathcal{P})$ by removing all rules with a predicate X^R for $R \notin S$. Then $\mathcal{P}'$ is a rewriting of $\mathcal{P}$ w.r.t. S.*

Proof. Follows analogously to Theorem 10 with minor adaptations of the relevant lemmas. □

C Proofs for Section 5

Definition 36. Let $r = \alpha \wedge \varphi_r \to \psi_r$ and $s = \varphi_s \to \beta \vee \psi_s$ be rules such that atom α is unifiable with β with MGU θ. The *elementary unfolding* $\mathsf{ElemUnfold}(r, \alpha, s, \beta)$ of r at α using s at β is the pair $((\varphi_r \wedge \varphi_s \to \psi_r \vee \psi_s)\theta, \theta)$. ◇

An elementary unfolding step thus amounts to resolving the relevant rules over the given predicates. *Unfolding* is then a transformation that allows us to replace a rule in a program with a sequence of elementary unfoldings in such a way that equivalence is preserved.

Definition 37. Let $\mathcal{P}$ be a disjunctive program, let $r \in \mathcal{P}$ and let α be a body atom in r; then, the *unfolding of r at α in $\mathcal{P}$*, denoted $\mathsf{Unfold}(\mathcal{P}, r, \alpha)$, is the result of applying Procedure 2 to $\mathcal{P}$, r, and α. ◇

Proposition 38. *Let $\mathcal{P}$ be a disjunctive program, r a rule in $\mathcal{P}$, and α a body atom of r. Then $\mathcal{P} \models \mathsf{Unfold}(\mathcal{P}, r, \alpha)$.*

Proof. The claim follows by soundness of resolution since every clause in $\mathsf{Unfold}(\mathcal{P}, r, \alpha) \setminus \mathcal{P}$ is obtained by resolution from clauses in $\mathcal{P}$. □

Lemma 39. *Let $\mathcal{P}$ be a disjunctive program, let $r = \bigwedge_{i=1}^n \alpha_i \to \psi$ ($n \geq 1$) be a rule in $\mathcal{P}$ where α_1 is IDB in $\mathcal{P}$, let $\mathcal{D}$ be a dataset containing no occurrences of IDB predicates in $\mathcal{P}$, and let σ be a ground substitution. If $\mathsf{Unfold}(\mathcal{P}, r, \alpha_1) \cup \mathcal{D} \models \alpha_i\sigma \vee \chi_{\alpha_i}$ for every $i \in [1, n]$ (where each χ_{α_i} is a ground disjunction of facts), then $\mathsf{Unfold}(\mathcal{P}, r, \alpha_1) \cup \mathcal{D} \models \psi\sigma \vee \chi_{\alpha_1} \vee \dots \vee \chi_{\alpha_n}$.*

Procedure 2 Unfold

Input: $\mathcal{P}$: a disjunctive program; r: a rule; α: a body atom of r
Output: the unfolding of r at α by $\mathcal{P}$

$S_0 := \{ (s, \beta) \mid s \in \mathcal{P}, \beta \text{ a head atom in } s \text{ unifiable with } \alpha \}$
$i := 0$
repeat
  $S_{i+1} := \emptyset$
  for each $(s, \beta) \in S_i$ **do**
    $(s', \theta) := \mathsf{ElemUnfold}(r, \alpha, s, \beta)$
    $S_{i+1} := S_{i+1} \cup \{ (s', \beta'\theta) \mid (s, \beta') \in S_i, \beta \neq \beta' \}$
  $i := i + 1$
until $S_i \neq \emptyset$
return $(\mathcal{P} \setminus \{r\}) \cup \{ s \mid (s, \beta) \in S_j, \text{ for } 1 \leq j < i \}$

Proof. For every $i \in [1, n]$, let $\mathsf{Unfold}(\mathcal{P}, r, \alpha_1) \cup \mathcal{D} \models \alpha_i\sigma \vee \chi_{\alpha_i}$. Let $\rho = (T, \lambda)$ be a derivation of $\alpha_1\sigma \vee \chi'_{\alpha_1}$ from $\mathsf{Unfold}(\mathcal{P}, r, \alpha_1) \cup \mathcal{D}$ for some $\chi'_{\alpha_1} \subseteq \chi_{\alpha_1}$ (existence of ρ follows by completeness of hyperresolution). Let s be the rule used to derive the label of the root v of ρ (i.e., $\alpha_1\sigma \vee \chi'_{\alpha_1}$) from the labels of its children $v_1, \dots, v_m$ (s must exist since α_1 is an IDB predicate and hence, by assumption, $\alpha_1 \notin \mathcal{D}$), and let τ be substitution used in the corresponding hyperresolution step. Then $s = \bigwedge_{j=1}^{m} \beta_j \to \alpha'_1 \vee \dots \vee \alpha'_l \vee \psi_{\alpha'}$ such that $\lambda(v_j) = \beta_j\tau \vee \chi_{\beta_j}$ for every $j \in [1, m]$, $\alpha_1\sigma = \alpha'_1\tau = \dots = \alpha'_l\tau$ and $\chi'_{\alpha_1} = \psi_{\alpha'}\tau \vee \chi_{\beta_1} \vee \dots \vee \chi_{\beta_m}$. Let r_1 be the rule obtained by elementary unfolding of r at α_1 using s at α'_1, and let r_k ($2 \leq k \leq l$) be the rule obtained by elementary unfolding of r at α_1 using r_{k-1} at α'_k. Then $(\bigwedge_{j=1}^{m} \beta_j\tau) \wedge (\bigwedge_{i=2}^{n} \alpha_i\sigma)$ is a substitution instance of the body of r_l and $\psi_{\alpha'}\tau \vee \psi\sigma$ is the corresponding instance of the head of r_l. Hence, the claim follows from the assumption ($\mathsf{Unfold}(\mathcal{P}, r, \alpha_1) \cup \mathcal{D} \models \alpha_i\sigma \vee \chi_{\alpha_i}$ for every $i \in [2, n]$) and soundness of hyperresolution (which implies $\mathsf{Unfold}(\mathcal{P}, r, \alpha_1) \cup \mathcal{D} \models \beta_j\tau \vee \chi_{\beta_j}$ for every $j \in [1, m]$) with r_l: we obtain $\mathsf{Unfold}(\mathcal{P}, r, \alpha_1) \cup \mathcal{D} \models \psi_{\alpha'}\tau \vee \psi\sigma \vee \chi_{\beta_1} \vee \dots \vee \chi_{\beta_m} \vee \chi_{\alpha_2} \vee \dots \vee \chi_{\alpha_n} = \psi\sigma \vee \chi'_{\alpha_1} \vee \chi_{\alpha_2} \vee \dots \vee \chi_{\alpha_n} \subseteq \psi\sigma \vee \chi_{\alpha_1} \vee \dots \vee \chi_{\alpha_n}$. □

Lemma 40. *Let $\mathcal{P}$ be a disjunctive program, let $\mathcal{D}$ be a dataset containing no occurrences of IDB predicates in $\mathcal{P}$, let r be a rule in $\mathcal{P}$, and let α be an IDB body atom of r. For every disjunction of facts φ such that $\mathcal{P} \cup \mathcal{D} \vdash \varphi$ we have $\mathsf{Unfold}(\mathcal{P}, r, \alpha) \cup \mathcal{D} \models \varphi$.*

Proof. Let $\rho = (T, \lambda)$ be a derivation of φ from $\mathcal{P} \cup \mathcal{D}$ and let v be the root of T. We proceed by induction on the size of T. If $\lambda(v) \in \mathcal{D}$, the claim is immediate. Otherwise, let $v_1, \dots, v_n$ be the successors of v in T ($n = 0$ if v is a leaf in T). By the inductive hypothesis, we have $\mathsf{Unfold}(\mathcal{P}, r, \alpha) \cup \mathcal{D} \models \lambda(v_i)$ for every $i \in [1, n]$. We distinguish two cases, depending on the rule $s \in \mathcal{P}$ used to derive $\lambda(v)$ from $\lambda(v_1), \dots, \lambda(v_n)$. If $s \neq r$, we have $s \in \mathsf{Unfold}(\mathcal{P}, r, \alpha)$, and the claim follows with s. If $s = r$, the claim follows by Lemma 39. □

Theorem 12. *Let $\mathcal{P}_0$ be a disjunctive program and let $\mathcal{P}$ be a rewriting of $\mathcal{P}_0$ such that no IDB predicate in $\mathcal{P}$ occurs in $\mathcal{P}_0$. Let r be a rule in $\mathcal{P}$ and α be an IDB body atom of r. Then $\mathsf{Unfold}(\mathcal{P}, r, \alpha)$ is a rewriting of $\mathcal{P}_0$. Moreover, no IDB predicate in $\mathsf{Unfold}(\mathcal{P}, r, \alpha)$ occurs in $\mathcal{P}_0$.*

Proof. Since $\mathcal{P}$ is a rewriting of $\mathcal{P}_0$, for the first claim it suffices to show $\mathsf{Eval}(\mathcal{P}, \mathcal{D}) = \mathsf{Eval}(\mathsf{Unfold}(\mathcal{P}, r, \alpha), \mathcal{D})$ for datasets $\mathcal{D}$ over the signature of $\mathcal{P}_0$. This follows by Proposition 38 and Lemma 40 since $\mathcal{P}_0$ contains no occurrences of IDB predicates in $\mathcal{P}$, and hence neither do datasets over the signature of $\mathcal{P}_0$. The second claim is immediate since all rules in $\mathsf{Unfold}(\mathcal{P}, r, \alpha) \setminus \mathcal{P}$ are obtained by resolution from those in $\mathcal{P}$, and the set of IDB predicates in a program is closed under resolution. □

D Proofs for Section 6

Proposition 41. *Checking $\mathcal{O} \cup \mathcal{D} \models \alpha$ for $\mathcal{O}$ an RL$^{\sqcup}$ ontology, $\mathcal{D}$ a dataset, and α a fact is co-*NP*-complete.*

Proof. Membership in co-NP follows from the fact that both the rules in Table 1 and the rules axiomatising equality and $\top$ contain a bounded number of variables and atoms; hence, the corresponding programs can be grounded in polynomial time and entailment in the resulting propositional program can be checked in co-NP. For hardness, it suffices to provide a straightforward encoding of non-3-colorability. The following DL ontology $\mathcal{O}$ can be normalised into an RL$^{\sqcup}$ ontology

$$V \sqsubseteq R \sqcup G \sqcup B \qquad B \sqcap \exists \mathsf{edge}.B \sqsubseteq \bot \qquad B \sqcap G \sqsubseteq \bot$$
$$G \sqcap \exists \mathsf{edge}.G \sqsubseteq \bot \qquad G \sqcap R \sqsubseteq \bot$$
$$R \sqcap \exists \mathsf{edge}.R \sqsubseteq \bot \qquad B \sqcap R \sqsubseteq \bot$$

Given an undirected graph $G = (V, E)$, the dataset $\mathcal{D}_G$ contains a fact $V(a)$ for each node $a \in V$ and facts $\mathsf{edge}(a, b)$ and $\mathsf{edge}(b, a)$ for each edge connecting a and b in E. Then, G is non-3-colorable iff $\mathcal{O} \cup \mathcal{D}_G$ is unsatisfiable. □

Theorem 15. *Checking $\mathcal{O} \cup \mathcal{D} \models \alpha$, for $\mathcal{O}$ an RL$^{\sqcup}$ ontology that corresponds to a WL program, is* PTIME*-complete.*

Proof. Hardness follows directly from the fact that the problem is already PTIME-hard if $\mathcal{O}$ is an OWL 2 RL ontology; thus, we focus on proving membership in PTIME. By Theorem 2 we have that $\mathcal{O} \cup \mathcal{D} \models \alpha$ iff $\Xi(\mathcal{O}) \cup \mathcal{D} \models \alpha\theta$ for some injective predicate renaming θ. Thus, it suffices to show that the evaluation of $\Xi(\mathcal{O})$ over $\mathcal{D}$ can be computed in polynomial time in the size of $\mathcal{O}$ and $\mathcal{D}$. First, $\Xi(\mathcal{O})$ is of size at most quadratic in the size of $\mathcal{O}$, and the arity of a predicate in $\Xi(\mathcal{O})$ is at most double the arity of a predicate in $\mathcal{O}$. As we can see in Table 1, the rules corresponding to Axioms 1-11 contain a bounded number of variables and atoms in the body, and hence the number of variables in the body of each rule and the arity of predicates in $\Xi(\mathcal{O})$ is bounded as well, as required. □

Let us fix an arbitrary $\mathcal{SHIQ}$ ontology $\mathcal{O}$, and let $\Omega_{\mathcal{O}}$ be obtained from $\mathcal{O}$ by first removing all axioms of the form 5 and then adding the relevant axioms to preserve fact entailment as described in (Cuenca Grau et al. 2013). Furthermore, let us denote with $\mathcal{R}_{\mathcal{O}}$ the subset of all axioms in $\mathcal{O}$ of the form 5, 4, and 8. Finally, let $\mathsf{DD}(\Omega_{\mathcal{O}})$ be the result of applying the algorithm in (Hustadt, Motik, and Sattler 2007) to $\Omega_{\mathcal{O}}$. The following lemma summarises the results in (Hustadt, Motik, and Sattler 2007; Cuenca Grau et al. 2013) that are relevant to us.

Lemma 42. *The following properties hold:*

1. $\Omega_{\mathcal{O}}$ *is a model conservative extension of* $\mathcal{O}$.
2. $\Omega_{\mathcal{O}} \models \mathsf{DD}(\Omega_{\mathcal{O}})$.
3. *For each dataset* $\mathcal{D}$ *and each fact* α *the following holds:*

$$\mathcal{O} \cup \mathcal{D} \models \alpha \quad \textit{iff} \quad \Omega_{\mathcal{O}} \cup \mathsf{Eval}(\mathcal{R}_{\mathcal{O}}, \mathcal{D}) \models \alpha \tag{21}$$

$$\Omega_{\mathcal{O}} \cup \mathcal{D} \models \alpha \quad \textit{iff} \quad \mathsf{DD}(\Omega_{\mathcal{O}}) \cup \mathcal{D} \models \alpha \tag{22}$$

Then, the following theorem states that the program obtained from $\Omega_{\mathcal{O}}$ by applying the algorithm in (Hustadt, Motik, and Sattler 2007) and then adding the rules in $\mathcal{R}_{\mathcal{O}}$ entails the same facts as $\mathcal{O}$ w.r.t. all datasets.

Theorem 43. $\mathcal{O} \cup \mathcal{D} \models \alpha$ *iff* $\mathsf{DD}(\Omega_{\mathcal{O}}) \cup \mathcal{R}_{\mathcal{O}} \cup \mathcal{D} \models \alpha$*, for every dataset* $\mathcal{D}$ *and fact* α *about individuals in* $\mathcal{D}$.

Proof. Assume that $\mathcal{O} \cup \mathcal{D} \models \alpha$. By Lemma 42, Condition (21) we have $\Omega_{\mathcal{O}} \cup \mathsf{Eval}(\mathcal{R}_{\mathcal{O}}, \mathcal{D}) \models \alpha$. Since $\mathsf{Eval}(\mathcal{R}_{\mathcal{O}}, \mathcal{D})$ is a dataset, by Lemma 42, Condition (22) we also have $\mathsf{DD}(\Omega_{\mathcal{O}}) \cup \mathsf{Eval}(\mathcal{R}_{\mathcal{O}}, \mathcal{D}) \models \alpha$, which then implies $\mathsf{DD}(\Omega_{\mathcal{O}}) \cup \mathcal{R}_{\mathcal{O}} \cup \mathcal{D} \models \alpha$, as required.

Assume that $\mathcal{O} \cup \mathcal{D} \not\models \alpha$. Since, by Lemma 42, $\Omega_{\mathcal{O}}$ is a conservative extension of $\mathcal{O}$ and $\mathcal{R}_{\mathcal{O}} \subseteq \mathcal{O}$ we have that $\Omega_{\mathcal{O}} \cup \mathcal{R}_{\mathcal{O}} \cup \mathcal{D} \not\models \alpha$. Again, by Lemma 42, we have that $\Omega_{\mathcal{O}} \models \mathsf{DD}(\Omega_{\mathcal{O}})$ and hence $\mathsf{DD}(\Omega_{\mathcal{O}}) \cup \mathcal{R}_{\mathcal{O}} \cup \mathcal{D} \not\models \alpha$. □

We define a program $\mathcal{P}$ to be a *rewriting of an ontology* $\mathcal{O}$ if $\mathcal{P}$ is a rewriting of $\mathsf{DD}(\Omega_{\mathcal{O}}) \cup \mathcal{R}_{\mathcal{O}}$. By Theorems 43 and 13, we then obtain the following.

Theorem 44. *Let* $\mathcal{O}$ *be an ontology. If* Rewrite *terminates on* $\mathsf{DD}(\Omega_{\mathcal{O}}) \cup \mathcal{R}_{\mathcal{O}}$ *with a datalog program* $\mathcal{P}$*, then* $\mathcal{P}$ *is a rewriting of* $\mathcal{O}$.

Proof. Immediate by Theorems 43 and 13. □