Online Appendix to: Towards a Logical Reconstruction of a Theory for Locally Closed Databases

MARC DENECKER, ÁLVARO CORTÉS-CALABUIG,
and MAURICE BRUYNOOGHE
Katholieke Universiteit Leuven
and
OFER ARIELI
The Academic College of Tel-Aviv

A. SUPPLEMENTARY MATERIAL FOR SECTION 5

In the following, we provide a more detailed investigation on the accuracy of approximate query answering; the main results of this analysis are given in Section 5.

A.1 Squared Queries

In this section we consider in greater details the notion of squaredness, which is fundamental for the analysis in Section 5. Given is a satisfiable theory Γ over σ containing $\mathsf{DCA}(\sigma) \wedge \mathsf{UNA}(\sigma)$. Recall that a query $\mathcal{Q}(\bar{x})$ is *squared* in Γ if for every $\bar{d}$ in the Herbrand universe *HU* for σ, $\Gamma \models \mathcal{Q}(\bar{d})$ implies that $\mathcal{Q}(\bar{d})^{\mathcal{O}_\Gamma} = \mathbf{t}$.

Squaredness can be split up in two simple unrelated concepts, called "literal-based queries" (Definition A.1) and "Kleene-precise queries" (Definition A.2).

Notation A.1. We denote by $Lit(\Gamma)$ the first-order theory consisting of $\mathsf{DCA}(\sigma) \wedge \mathsf{UNA}(\sigma)$ and the set of all ground literals L entailed by Γ.

Definition A.1 (*Literal-Based Queries*). We say that a query $\mathcal{Q}(\bar{x})$ of arity n is *literal-based* in Γ if for every $\bar{d}HU^n$, $\Gamma \models \mathcal{Q}(\bar{d})$ implies $Lit(\Gamma) \models \mathcal{Q}(\bar{d})$.

If, vice versa, $Lit(\Gamma) \models \mathcal{Q}(\bar{d})$, then obviously, since $\Gamma \models Lit(\Gamma)$, we have $\Gamma \models \mathcal{Q}(\bar{d})$. Hence, a query is literal-based iff its certain answers under Γ and under $Lit(\Gamma)$ are the same.

Example A.1. For $\Gamma = \{\neg R(a) \vee \neg Q(a)\}$, we have that Γ entails no ground literals. The query $\neg R(a) \vee \neg Q(a)$ is not literal-based in Γ, since $\Gamma \models \neg R(a) \vee \neg Q(a)$ and $Lit(\Gamma) \not\models \neg R(a) \vee \neg Q(a)$. On the other hand, every other formula of the form $L \circ L'$ where $L \in \{R(a), \neg R(a)\}$, $\circ \in \{\wedge, \vee\}$ and $L' \in \{Q(a), \neg Q(a)\}$ is literal-based in Γ.

Clearly, the theory $Lit(\Gamma)$ and the optimal approximation $\mathcal{O}_\Gamma$ of Γ are equivalent concepts, in the sense that $P(\bar{a})^{\mathcal{O}_\Gamma} = \mathbf{t}$ iff $P(\bar{a}) \in Lit(\Gamma)$, and $P(\bar{a})^{\mathcal{O}_\Gamma} = \mathbf{f}$

DOI 10.1145/1806907.1806914 http://doi.acm.org/10.1145/1806907.1806914

iff $\neg P(\bar{a}) \in Lit(\Gamma)$. Therefore, the concept of a literal-based query can also be defined in terms of $\mathcal{O}_\Gamma$, as shown in the following proposition. Recall from Section 4.1 that $sv_{\mathcal{K}}(\varphi)$ denotes φ's supervaluation in $\mathcal{K}$.

PROPOSITION A.1. *$\mathcal{Q}(\bar{x})$ is literal-based in Γ iff $Cert_\Gamma(\mathcal{Q}(\bar{x})) = \{\bar{d} \in HU^n \mid sv_{\mathcal{O}_\Gamma}(\mathcal{Q}(\bar{d})) = \mathbf{t}\}$.*

In words, a query is literal-based in a theory iff its certain answers in that theory are its certain answers computed with supervaluation in the optimal approximation of that theory.

PROOF. Follows from the fact that $Lit(\Gamma) \models \mathcal{Q}(\bar{d}$ iff $sv_{\mathcal{O}_\Gamma}(\mathcal{Q}(\bar{d})) = \mathbf{t}$. The latter is a consequence of the fact that $M \models Lit(\Gamma)$ iff M is (isomorphic to) a Herbrand structure such that $\mathcal{O}_\Gamma \leqslant_{\mathrm{p}} M$. □

Note the similarity to Proposition 29, which uses Kleene's three-valued truth assignment while the previous proposition uses supervaluation. An interesting class of queries is when both truth assignments coincide.

Definition A.2 (*Kleene-Precise Queries*). We say that a query $\mathcal{Q}(\bar{x})$ of arity n is *Kleene-precise* in a three-valued structure $\mathcal{K}$ with domain *Dom* if for each $\bar{d} \in Dom^n$, $sv_{\mathcal{K}}(\mathcal{Q}(\bar{d}) = \mathbf{t}$ implies $\mathcal{Q}(\bar{d}^{\mathcal{K}} = \mathbf{t}$.

Since supervaluation is more precise than standard Kleene truth assignment, it follows that a query is Kleene-precise iff $Cert_{\mathcal{K}}(\mathcal{Q}(\bar{x})) = \{\bar{d} \in Dom^n \mid sv_{\mathcal{K}}(\mathcal{Q}(\bar{d})) = \mathbf{t}\}$, that is, its certain answers under standard Kleene truth assignment and supervaluation coincide.

Example A.2. The concepts of literal-based and Kleene-precise queries are unrelated. For example, $P(a) \vee \neg P(a)$ is literal-based in each theory but not Kleene-precise in $\mathcal{K}$ if $P(a)^{\mathcal{K}} = \mathbf{u}$. The query $P(a) \vee Q(a)$ is not literal-based in the theory $\{P(a) \vee Q(a)\}$ but $P(a) \vee Q(a)$ is Kleene-precise in every $\mathcal{K}$. The query $P(a) \wedge \neg P(a)$ is literal-based in every theory and Kleene-precise in every structure.

The following proposition connects the preceding concepts to each other.

PROPOSITION A.2. *A query $\mathcal{Q}(\bar{x})$ is squared in Γ iff $\mathcal{Q}(\bar{x})$ is literal-based in Γ and Kleene-precise in $\mathcal{O}_\Gamma$.*

PROOF. For every Γ and $\mathcal{Q}(\bar{x})$, the following inequalities hold.

$$Cert_\Gamma(\mathcal{Q}(\bar{x})) \supseteq \{\bar{d} \in HU^n \mid sv_{\mathcal{O}_\Gamma}(\mathcal{Q}(\bar{d})) = \mathbf{t}\} \supseteq Cert_{\mathcal{O}_\Gamma}(\mathcal{Q}(\bar{x}))$$

If $\mathcal{Q}(\bar{x})$ is literal-based in Γ and Kleene-precise in $\mathcal{O}_\Gamma$, the inequalities turn into equalities. Conversely, if $\mathcal{Q}(\bar{x})$ is squared, the three terms are equal, and it follows that the query is literal-based and Kleene-precise. □

We now present simple conditions for literal-based and Kleene-precise queries.

PROPOSITION A.3. *A query $\mathcal{Q}(\bar{x})$ is Kleene-precise in $\mathcal{O}_\Gamma$ and literal-based (and hence squared) in Γ when it is of the form $\forall \bar{y} : (C_1 \vee \cdots \vee C_n)$, where each C_i is a conjunction such that: (i) each nonliteral conjunct C_{i_k} of C_i is built from* *predicates that are two-valued in $\mathcal{O}_\Gamma$ and (ii) the set of conjunctions is mutually exclusive*[1] *with respect to $\mathcal{O}_\Gamma$.*

PROOF. Let us start by observing that since Γ contains $\mathsf{DCA}(\sigma) \wedge \mathsf{UNA}(\sigma)$, a formula $\forall \bar{y} : \varphi$ is literal-based in Γ if φ is literal-based in Γ. Likewise, it is Kleene-precise in $\mathcal{O}_\Gamma$ if φ is Kleene-precise in $\mathcal{O}_\Gamma$. Hence, it suffices to prove the proposition for the case that $\mathcal{Q}(\bar{x})$ is of the form $C_1 \vee \cdots \vee C_n$.

Nothing is to be proved for those $\bar{d}$ such that $\Gamma \not\models \mathcal{Q}(\bar{d})$. Since each pair $C_i[\bar{x}], C_j[\bar{x}]$ of $\mathcal{Q}(\bar{x})$ is mutually exclusive in $\mathcal{O}_\Gamma$, $\Gamma \models \mathcal{Q}(\bar{d})$ implies there is exactly one C_i such that $\Gamma \models C_i[\bar{d}]$. For each conjunct C_{i_k} of this C_i, we have also $\Gamma \models C_{i_k}[\bar{d}]$. When $C_{i_k}[\bar{d}]$ is a literal, it is true in $\mathcal{O}_\Gamma$; when it is not a literal, it is two-valued in $\mathcal{O}_\Gamma$, hence it is also true in $\mathcal{O}_\Gamma$. So, all conjuncts $C_{i_k}[\bar{d}]$ are true in $\mathcal{O}_\Gamma$ and hence also $C_i[\bar{d}]$ and $\mathcal{Q}(\bar{d})$ are true in $\mathcal{O}_\Gamma$. It follows that the query is squared. $\square$

Example A.3. Some trivial examples show that disjunctions or existentially closed formulas may not be literal-based, even if their component formulas are literal-based: the query $P(a) \vee P(b)$ is not literal-based in the theory $\{P(a) \vee P(b)\}$ and $\exists x : P(x)$ is not literal-based in the theory $\{\exists x : P(x)\}$.

Example A.4. Here is a query that illustrates that existential quantified formulas are not necessarily Kleene-precise, even if its components are.

$$\varphi := \exists x : (P(x) \wedge Q(a) \vee \neg P(x) \wedge \neg Q(a))$$

Consider the structure $\mathcal{K}$ with domain $\{a, b\}$ and such that $P(a)^{\mathcal{K}} = \mathbf{t}$, $P(b)^{\mathcal{K}} = \mathbf{f}$, and $Q^{\mathcal{K}}(a) = \mathbf{u}$. Since $P^{\mathcal{K}}$ is two-valued, the two disjuncts are mutually exclusive, hence the subformula $P(x) \wedge Q(a) \vee \neg P(x) \wedge \neg Q(a)$ satisfies the condition of Proposition A.3. Still, it holds that $sv_{\mathcal{K}}(\varphi) = \mathbf{t} \neq \varphi^{\mathcal{K}} = \mathbf{u}$.

PROPOSITION A.4. *Let $\mathcal{Q}(\bar{x})$ be a query such that each predicate P that has positive and negative occurrences in $\mathcal{Q}(\bar{x})$ is two-valued in $\mathcal{K}$. Then $\mathcal{Q}(\bar{x})$ is Kleene-precise in $\mathcal{K}$.*

PROOF. Let $\bar{d}$ be an arbitrary tuple of domain elements. We need to show that for $\varphi = \mathcal{Q}(\bar{d})$, $sv_{\mathcal{K}}(\varphi) = \varphi^{\mathcal{K}}$. Since $\varphi^{\mathcal{K}} \leqslant_{\mathrm{p}} sv_{\mathcal{K}}(\varphi)$, it suffices to show that $\varphi^{\mathcal{K}} = \mathbf{u}$ implies $sv_{\mathcal{K}}(\varphi) = \mathbf{u}$. Assume that $\varphi^{\mathcal{K}} = \mathbf{u}$. We construct two-valued structures $K, K' \geqslant_{\mathrm{p}} \mathcal{K}$ in the following way.

—K is the two-valued extension of $\mathcal{K}$ which maps each unknown atom $P(\bar{a})$ of $\mathcal{K}$ to $\mathbf{t}$ if P occurs positively in φ and to $\mathbf{f}$ otherwise.

—For K' the inverse strategy is followed and it maps unknown atoms $P(\bar{a})$ to $\mathbf{f}$ if P occurs positively, and to $\mathbf{t}$ otherwise.

To show that $sv_{\mathcal{K}}(\varphi) = \mathbf{u}$, we prove that $\varphi^{K'} = \mathbf{f}$ and $\varphi^{K} = \mathbf{t}$, by induction on the structure of φ. Recall that $\varphi^{\mathcal{K}} = \mathbf{u}$.

—Assume that $\varphi = P(\bar{a})$. Then, since P occurs positively, $P(\bar{a})^{K} = \mathbf{t}$ and $P(\bar{a})^{K'} = \mathbf{f}$.

[1] In the sense of Definition 20.

—If $\varphi = \phi \wedge \psi$ then one conjunct is unknown and the other is true or unknown. It follows from the induction hypothesis, that both conjuncts are true in K and at least one is false in K'. Hence, $(\phi \wedge \psi)^K = \mathbf{t}$ and $(\phi \wedge \psi)^{K'} = \mathbf{f}$.

—The cases of disjunction, and universal and existential quantifiers can be proved in a similar way.

—Assume that $\varphi = \neg\phi$. In that case, $\phi^{\mathcal{K}} = \mathbf{u}$ and ϕ is a formula in which the polarities of all predicates are switched. By switching the roles of K and K' and applying the induction hypothesis, we obtain that $\phi^K = \mathbf{f}$ and $\phi^{K'} = \mathbf{t}$. Consequently, $\varphi^K = \mathbf{t}$ and $\varphi^{K'} = \mathbf{f}$. □

Note A.1. Observe that this proposition implies that positive formulas as well as negative formulas are Kleene-precise. Also, the class of Kleene-precise formulas described in the proposition is closed under negation. This is not the case for Kleene-precise formulas in general. For example, the query $P(a) \wedge \neg P(a)$ is Kleene-precise in every three-valued structure in which $P(a)$ is unknown while its negation $\neg P(a) \vee P(a)$ is not.

A formula $\mathcal{Q}(\bar{x})$ which is Kleene-precise but does not satisfy the syntactical conditions of Proposition A.3 can still be guaranteed to be squared if Γ satisfies additional conditions expressed in the following definition.

Definition A.3 (*Atomical Theories*). A theory (or formula) Γ is atomical in σ iff for each two-valued Herbrand σ-structure I such that $\mathcal{O}_\Gamma \leqslant_p I$, $I \models \Gamma$.

When σ is clear from the context we say that Γ is atomical.

Proposition A.5. *Let Γ be a theory containing* $\mathsf{DCA}(\sigma) \wedge \mathsf{UNA}(\sigma)$. *$\Gamma$ is atomical in σ iff Γ and $Lit(\Gamma)$ are logically equivalent.*

It follows that in an atomical theory every query is literal-based. This leads to the following straightforward proposition.

Proposition A.6. *If Γ is atomical, every Kleene-precise query $\mathcal{Q}(\bar{x})$ in $\mathcal{O}_\Gamma$ is squared in Γ.*

The following proposition lists some common and simple cases of atomical formulas and theories.

Proposition A.7. *A literal is an atomical formula. A conjunction or a set of atomical formula's is atomical. The union of a set of atomical theories is atomical. If each ground instance $\varphi[\bar{d}]$ is atomical, then $\forall\bar{x} : \varphi[\bar{x}]$ is atomical. If Γ is an atomical formula that conveys CWI over $\varphi_1[\bar{x}], \ldots, \varphi_n[\bar{x}]$ and for each $\bar{d}$, $\psi[\bar{d}]$ is atomical in σ, then $\Gamma \cup \{\forall\bar{x} : (\psi \vee \varphi_1 \vee \cdots \vee \varphi_n)\}$ is atomical.*

Proof. All items of the proposition are straightforward with exception perhaps of the last one. The theory $\Gamma \cup \{\psi[\bar{d}] \vee \varphi_1[\bar{d}] \vee \cdots \vee \varphi_n[\bar{d}]\}$ is clearly equivalent with Γ if there exists an i such that $\Gamma \models \varphi_i[\bar{d}]$, and with $\Gamma \cup \{\psi[\bar{d}]\}$ otherwise, that is, if $\Gamma \models \neg\varphi_i[\bar{d}]$ for each i. Both theories are atomical. The union of all these theories for all $\bar{d}$ is also atomical. This theory is equivalent with $\Gamma \cup \{\forall\bar{x} : (\psi \vee \varphi_1 \vee \cdots \vee \varphi_n)\}$. □

A.2 Application to Locally Closed Databases

According to Theorem 6 of Section 5, a query is optimally answered by our approximate query answering techniques, provided that one can prove: (i) the squaredness of the query with respect to $\mathfrak{D}$, and (ii) the optimality of $\mathcal{C}_{\mathfrak{D}}$ on the (potentially small) set of negatively occurring predicates. We analyse each of these conditions in turn.

A.2.1 *Proving Squaredness of Queries.* Syntactic conditions for squared queries can be derived from Proposition A.3 and Proposition A.4. However, we will be able to improve the results of these propositions by exploiting some additional properties of locally closed databases. Recall Definition 21 where we introduced $\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))(:= \prec_{\cup}\{P|P \in^{-} \mathcal{Q}(\bar{x})\})$, the set of *negatively bound predicates* of a query $\mathcal{Q}(\bar{x})$, and its complement $\mathcal{P}os\mathcal{F}(\mathcal{Q}(\bar{x}))$, the set of *positive free predicates* of $\mathcal{Q}(\bar{x})$.

Definition A.4. Let $\mathfrak{D}_{\mathcal{Q}}$ be the extension of $\mathfrak{D} = (D, \mathcal{L})$ obtained by adding $\mathcal{LCWA}(P(\bar{x}), \mathbf{t})$ to $\mathcal{L}$, for each predicate P that is positive free in $\mathcal{Q}(\bar{x})$.

PROPOSITION A.8. *$\mathcal{O}_{\mathfrak{D}} \leqslant_{\mathrm{p}} \mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}$. More specifically, $P^{\mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}} = P^{D}$ if P is positive free in $\mathcal{Q}(\bar{x})$, and $P^{\mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}} = P^{\mathcal{O}_{\mathfrak{D}}}$ otherwise.*

PROOF. A positive free predicate P is a base predicate of $\mathfrak{D}_{\mathcal{Q}}$ and hence, $P^{\mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}} = P^{D}$. The databases $\mathfrak{D}_{\mathcal{Q}}|_{\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))}$ and $\mathfrak{D}|_{\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))}$ coincide and have the same optimal approximations. Since $\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))$ is $\prec$-downward closed, the extendibility Lemma 1 entails $P^{\mathcal{O}_{\mathfrak{D}}} = P^{\mathcal{O}_{\mathfrak{D}|_{\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))}}} = P^{\mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}|_{\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))}}} = P^{\mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}}$, for each $P \in \mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))$. □

PROPOSITION A.9. *If $\mathcal{Q}(\bar{x})$ is squared in $\mathfrak{D}_{\mathcal{Q}}$, then $\mathcal{Q}(\bar{x})$ is squared in $\mathfrak{D}$.*

PROOF. Assume that $\mathfrak{D} \models \mathcal{Q}(\bar{d})$. Since a model of $\mathfrak{D}_{\mathcal{Q}}$ is a model of $\mathfrak{D}$, we have $\mathfrak{D}_{\mathcal{Q}} \models \mathcal{Q}(\bar{d})$. The query $\mathcal{Q}(\bar{x})$ is squared in $\mathfrak{D}_{\mathcal{Q}}$, hence $\mathcal{Q}(\bar{d})^{\mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}} = \mathbf{t}$. Now, observe that $\mathcal{O}_{\mathfrak{D}}$ is an approximation of $\mathfrak{D}_{\mathcal{Q}}$ that satisfies the condition on $\mathcal{K}$ in Proposition 30: $P(\bar{a})^{\mathcal{O}_{\mathfrak{D}}} = \mathbf{t}$ iff $P(\bar{a})^{D} = \mathbf{t}$, and $P^{\mathcal{O}_{\mathfrak{D}}} = P^{\mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}}$ if P occurs negatively in $\mathcal{Q}(\bar{x})$. It follows that $Cert_{\mathcal{O}_{\mathfrak{D}}}(\mathcal{Q}(\bar{x})) = Cert_{\mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}}(\mathcal{Q}(\bar{x}))$. Consequently, $\mathcal{Q}(\bar{d})^{\mathcal{O}_{\mathfrak{D}}} = \mathbf{t}$. Thus $\mathcal{Q}(\bar{x})$ is squared in $\mathfrak{D}$. □

We now have all the basic material for proving the theorem about squaredness.

PROOF OF THEOREM 7. The theorem states:

A query $\mathcal{Q}(\bar{x})$ is squared in $\mathfrak{D}$ when it is of the form $\forall\bar{y} : (C_1 \vee \cdots \vee C_n)$, where each C_i is a conjunction such that: (i) each nonliteral conjunct C_{i_k} of the C_i consists of predicates which are either positive free in $\mathcal{Q}(\bar{x})$ or two-valued in $\mathcal{O}_{\mathfrak{D}}$, and (ii) the set of conjunctions is mutually exclusive with respect to $\mathcal{O}_{\mathfrak{D}}$.

To prove it, observe that if a set of conjunctions is mutually exclusive with respect to $\mathcal{O}_{\mathfrak{D}}$ then it is also mutually exclusive with respect to $\mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}$. This follows directly from $\mathcal{O}_{\mathfrak{D}} \leqslant_{\mathrm{p}} \mathcal{O}_{\mathfrak{D}_{\mathcal{Q}}}$. The theorem follows then from Proposition A.3 and Proposition A.9. □

Note A.2. Recall that since $\mathcal{C}_{\mathfrak{D}} \leqslant_{\mathrm{p}} \mathcal{O}_{\mathfrak{D}}$, we can use $\mathcal{C}_{\mathfrak{D}}$ to check two-valuedness of negatively bound predicates in $\mathcal{O}_{\mathfrak{D}}$ and mutual exclusiveness of the set of conjunctions in $\mathcal{O}_{\mathfrak{D}}$.

The class of queries covered by Theorem 7 comprises conjunctions of literals, positive formulas, and more in general, decision tree-like queries in which the test formulas are two-valued in $\mathcal{C}_{\mathfrak{D}}$ (e.g., contain only base predicates of $\mathfrak{D}$) and the leaves consist of conjunctions of database literals and formulas containing base predicates and (positive occurrences of) positive free predicates of $\mathcal{Q}(\bar{x})$.

A.2.2 *Proving Atomicality.* We now present conditions for the atomicality of $\mathfrak{D}$ and $\mathfrak{D}_{\mathcal{Q}}$ and derive a large class of squared queries in case of atomicality.

PROPOSITION A.10. *For every $\prec$-downward closed set of predicates $\mathcal{P}$, $\mathfrak{D}|_{\mathcal{P}}$ is atomical if $\prec$ is cycle-free in $\mathcal{P}$ and for each $P \in \mathcal{P}$, $\mathfrak{D}$ conveys CWI on $\Psi_P[\bar{x}]$.*

PROOF. $\mathfrak{D}|_{\mathcal{P}}$ is the union of all $\mathfrak{D}|_{\prec_\cup P}$, for all $P \in \mathcal{P}$. By Proposition A.7, $\mathfrak{D}|_{\mathcal{P}}$ is atomical if each $\mathfrak{D}|_{\prec_\cup P}$ is atomical. The proof is by induction on $\prec$. Assume that $\mathfrak{D}|_{\prec P}$ is atomical. Since $\mathfrak{D}$ conveys CWI on the window of expertise of P, it follows from Corollary 4 that so does $\mathfrak{D}|_{\prec P}$. $\mathfrak{D}|_{\prec_\cup P}$ consists of $\mathfrak{D}|_{\prec P}$, all atoms $P(\bar{a})$ in D, and

$$\forall \bar{x} : (\Psi[\bar{x}] \supset (P(\bar{x}) \supset P(\bar{x}) \in D)).$$

The latter formula is a disjunction consisting of two disjuncts on which there is CWI and a third atomical disjunct. By, again, Proposition A.7, $\mathfrak{D}|_{\prec_\cup P}$ is atomical. □

Although the atomicality condition is a strong condition, the previous proposition shows that it may easily arise in the context of locally closed databases. It suffices that the database has CWI on its own windows of expertise.

Example A.5. The following example is based on Example 11. Consider the local closed world assumptions.

$$\mathcal{L}_1 = \left\{ \begin{array}{l} \mathcal{LCWA}(Loc(p,l),(l = Bx \wedge \exists m, id : CarO(p,m,id))), \\ \mathcal{LCWA}(CarO(p,m,id), Loc(p,Bx)) \end{array} \right\}$$

Take a locally closed database $\mathfrak{D}_1$ consisting of $\mathcal{L}_1$, the empty relation interpreting *Loc* and *CarO*, and a domain consisting of the locations Q, Bx, the person LD, car V, and car id $V40$. This database conveys no CWI on its windows of expertise. For instance, Lien Desmet (LD) could live in the Bronx ($Loc(LD, Bx)$) or not; she could have a car ($CarO(LD, V, V40)$) or not. But if she lives in the Bronx, then she cannot have the car and vice versa. This is to say that $\mathfrak{D}_1 \models \neg Loc(LD, Bx) \vee \neg CarO(LD, V, V40)$ while $Lit(\mathfrak{D}_1)$ is empty and does not entail this formula. This shows that this database is not atomical.

On the other hand, consider the following LCWA's.

$$\mathcal{L}_2 = \{\mathcal{LCWA}(Loc(n,l), \mathbf{t}), \mathcal{LCWA}(CarO(n,m,id), Loc(n,Bx))\}$$

It is easy to see that each database containing $\mathcal{L}_2$ conveys CWI on its windows of expertise and hence is atomical. For instance, the database $\mathfrak{D}_2$ obtained from $\mathfrak{D}_1$ by substituting $\mathcal{L}_2$ for $\mathcal{L}_1$ is atomical. $Lit(\mathfrak{D}_2)$ contains $\neg Loc(LD, Bx)$ and entails $\neg Loc(LD, Bx) \vee \neg CarO(LD, V, V40)$.

The interest of the atomicality of $\mathfrak{D}$ lies in the fact that the large class of Kleene-precise queries are squared in it. A sufficient reason for $\mathfrak{D}$ to be atomical is CWI on all windows of expertise. However, for a Kleene-precise query to be squared, it suffices that there is CWI on a potentially very small set of windows of expertise. Next we investigate this.

THEOREM A.1 (SQUAREDNESS IN ATOMICAL DATABASES). *A query $\mathcal{Q}(\bar{x})$ is squared in $\mathfrak{D}$ if: (i) $\prec$ is acyclic in $\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))$ and $\mathfrak{D}$ conveys CWI on the windows of expertise on each $P \in \mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))$ and (ii) all predicates with positive and negative occurrences in $\mathcal{Q}(\bar{x})$ are base predicates of $\mathfrak{D}$.*

PROOF. We first show that $\mathfrak{D}_Q$ is atomical. The full dependency graph of $\mathfrak{D}_Q$ is that of $\mathfrak{D}$ restricted to $\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))$ and is acyclic. The window of expertise of each positive free predicate in $\mathfrak{D}_Q$ is the base predicate $\mathbf{t}$. Since $\mathfrak{D}$ conveys CWI on the windows of expertise of negatively bound predicates, so does the stronger theory $\mathfrak{D}_Q$. If follows from Proposition A.10 that $\mathfrak{D}_Q$ is atomical.

Since $\mathcal{Q}(\bar{x})$ is Kleene-precise in $\mathcal{O}_{\mathfrak{D}_Q}$ (Proposition A.4), it is squared in the atomical database $\mathfrak{D}_Q$. By Proposition A.9, $\mathcal{Q}(\bar{x})$ is squared in $\mathfrak{D}$. □

Example A.6. By Theorem A.1, the query $\neg(Loc(LD, Bx) \wedge CarO\ (LD, V, B1))$ is squared in the atomical database $\mathfrak{D}_2$ of Example A.5. This query is not squared in the database $\mathfrak{D}_1$ of that example and cannot be accurately processed in it by the approximate methods. Indeed, it is a Kleene-precise query and is certainly entailed by $\mathfrak{D}_1$, but this formula in unknown in $\mathcal{O}_{\mathfrak{D}_1}$ and $\mathcal{C}_{\mathfrak{D}_1}$. Note that it is true in the optimal approximation of the atomical database $\mathfrak{D}_2$.

The class of queries covered by Theorem A.1 allows arbitrary quantification. Since $\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))$ is empty if $\mathcal{Q}(\bar{x})$ is a positive query, positive queries are covered by this theorem (they were covered also by Theorem 7).

Example A.7. In Example 11 as in $\mathfrak{D}_1$ of Example A.5, all predicates are positive free for $CarO(n, m, id)$, and none for $\neg CarO(n, m, id)$ since $CarO$ depends on Loc.

A.2.3 *Proving Partial Optimality of $\mathcal{C}_{\mathfrak{D}}$.* We say that $\mathcal{C}_{\mathfrak{D}}$ is optimal in a predicate P if $P^{\mathcal{C}_{\mathfrak{D}}} = P^{\mathcal{O}_{\mathfrak{D}}}$. For an accurate answer to a query $\mathcal{Q}(\bar{x})$ using $\mathcal{C}_{\mathfrak{D}}$, it does not suffice that the query is squared. It should also hold that $P^{\mathcal{C}_{\mathfrak{D}}}$ is optimal for each $P \in^- \mathcal{Q}(\bar{x})$. We now present conditions for the optimality of $\mathcal{C}_{\mathfrak{D}}$ in some set of predicates. We start with a construction proposition showing how optimality of $\mathcal{C}_{\mathfrak{D}}$ may propagate from predicates in lower levels of $\prec^-$ to higher levels.

PROPOSITION A.11. *For database predicate P, it holds that $P^{\mathcal{C}_{\mathfrak{D}}} = P^{\mathcal{O}_{\mathfrak{D}}}$ if the following conditions are satisfied:*

—*$\Psi_P[\bar{x}]$ is squared in $\mathfrak{D}$,*
—*for each atom $P(\bar{a}) \notin D$, if $\mathfrak{D} \models \neg P(\bar{a})$ then $\mathfrak{D} \models \Psi_P[\bar{d}]$,*
—*for every predicate Q that occurs negatively in Ψ_P, $Q^{\mathcal{C}_{\mathfrak{D}}} = Q^{\mathcal{O}_{\mathfrak{D}}}$.*

Proof. By the soundness of $\mathcal{C}_{\mathfrak{D}}$, it holds for every $\bar{a} \in Dom^n$ that if $P(\bar{a})^{\mathcal{C}_{\mathfrak{D}}} \neq \mathbf{u}$, then $P(\bar{a})^{\mathcal{C}_{\mathfrak{D}}} = P(\bar{a})^{\mathcal{O}_{\mathfrak{D}}}$. So, let us assume that $P(\bar{a})^{\mathcal{C}_{\mathfrak{D}}} = \mathbf{u}$. Observe that, in this case, $P(\bar{a}) \notin D$.

To show that $P(\bar{a})^{\mathcal{O}_{\mathfrak{D}}} = \mathbf{u}$, we need to construct two models M, M' of $\mathfrak{D}$ such that $M \models P(\bar{a})$ and $M' \models \neg P(\bar{a})$. Since $P(\bar{a}) \notin D$, we can take $M' = D$ which is indeed a model of $\mathfrak{D}$. Let us now construct M. By construction of $\mathcal{C}_{\mathfrak{D}}$, it holds that $P(\bar{a}) \notin D$ and $\Psi_P[\bar{d}]^{\mathcal{C}_{\mathfrak{D}}} \neq \mathbf{t}$. The sentence $\Psi_P[\bar{d}]$ is squared in $\mathfrak{D}$ and by assumption, it holds that $Q^{\mathcal{C}_{\mathfrak{D}}} = Q^{\mathcal{O}_{\mathfrak{D}}}$ for each negatively occurring predicate Q in this formula. Therefore, the conditions of Theorem 6 are satisfied, so there is a model N of $\mathfrak{D}$ such that $N \models \neg\Psi_P[\bar{d}]$ and this means that $\mathfrak{D} \not\models \Psi_P[\bar{d}]$. By contraposition of the second assumption, it follows that $\mathfrak{D} \not\models \neg P(\bar{a})$, hence $\mathfrak{D}$ has a model M in which $P(\bar{a})$ is true. □

The preceding proposition reduces the problem of the optimality of P in $\mathcal{C}_{\mathfrak{D}}$ to the problem of the optimality of the negatively occurring predicates in P's window of expertise. Results from earlier propositions can be used to determine whether P's window of expertise is squared. For example, if Ψ_P is a positive formula, or a formula in which only base predicates occur negatively, then Ψ_P is squared in $\mathfrak{D}$.

The other condition is that $\mathfrak{D} \models \neg P(\bar{a})$ should imply $\mathfrak{D} \models \Psi_P[\bar{d}]$. Equivalently, if $\neg\Psi_P[\bar{d}]$ is satisfiable in $\mathfrak{D}$, then so should be $P(\bar{a})$. This is a condition that one would expect to be satisfied, given that the "main" axiom of $\mathfrak{D}$ about $P(\bar{a})$ is as follows. We have

$$\Psi_P[\bar{d}] \supset (P(\bar{a}) \supset P(\bar{a}) \in D)$$

and this formula is satisfied in any model M in which $\Psi_P[\bar{d}]$ is false, independent of the truth value of $P(\bar{a})$ in M. Yet, it is possible to engineer examples in which the condition is not satisfied. If we start from a model M of such databases in which $\Psi_P[\bar{d}]$ is false and we try to update M to make $P(\bar{a})$ true, then we cause a chain of forced updates to M with the effect that ultimately $\Psi_P[\bar{d}]$ is made true as well. In such a case, accuracy may indeed be lost, as illustrated next.

Example A.8. In the database $\mathfrak{D} = (\{\}, \{\mathcal{LCWA}(P(x), x = a \wedge P(a))\})$ of Example 17, precision is lost on $P(a)$. We have $\mathfrak{D} \models \neg P(a)$ but $\mathfrak{D} \not\models \Psi_P[a]$.

Example A.9. The database $\mathfrak{D} = (\{\}, \{\mathcal{LCWA}(P(x), \neg Q(x)), \mathcal{LCWA}(Q(x), P(x))\})$ of Example 18 is equivalent to $\forall x : \neg P(x)$, yet $P(d)^{\mathcal{C}_{\mathfrak{D}}} = \mathbf{u}$, for every $d \in HU$.

In the two preceding examples, all conditions of Proposition A.11 are satisfied for P except for the second condition. In both cases, precision is lost.

Next, we present a simple syntactic condition that guarantees the second condition of Proposition A.11.

Proposition A.12. *If $P \not\in^+ \Psi_P$ (i.e., P has no positive occurrence in its own window of expertise) and for all $Q \prec^- P$ we have $P \not\in^+ \Psi_Q$, then for each atom $P(\bar{a})$ such that $\mathfrak{D} \models \neg P(\bar{a})$ we have $\mathfrak{D} \models \Psi_P[\bar{d}]$.*

Notice that the first condition is violated in Example 17, and the second condition in Example 18.

PROOF. We will prove the contrapositive, that is, if $\mathfrak{D} \not\models \Psi_P[\bar{d}]$ then $\mathfrak{D} \not\models \neg P(\bar{a})$. Let M be a model of $\mathfrak{D}$ such that $M \models \neg\Psi_P[\bar{d}]$. If $P(\bar{a})$ is true in M, there is nothing to prove. So assume that $P(\bar{a})$ is false in M. We will modify M in N such that N is still a model of $\mathfrak{D}$ and $P(\bar{a})$ is true in N.

Consider the set $\mathcal{S}_P = \{Q \in \Sigma \mid P \in^+ \Psi_Q\} \cup \{R \in \Sigma \mid \exists Q \in \Sigma : P \in^+ \Psi_Q$ and $Q \prec^- R\}$. It consists of all predicates that have a positive occurrence of P in their window of expertise and all predicates that negatively depend on such predicates. It follows from the condition of the proposition that $P \notin \mathcal{S}_P$. In the windows of expertise of predicates Q in the complement of $\mathcal{S}_P$, P has only negative occurrences and predicates of $\mathcal{S}_P$ have only positive occurrences.

Define N as the structure obtained by modifying M as follows:

— $P^N = P^M \cup \{\bar{d}\}$, that is, $P(\bar{a})$ is made true;
— $Q^N = Q^D$ for $Q \in \mathcal{S}_P$.

This modification increases P and decreases all predicates of $\mathcal{S}_P$; that is, $P^M \leq P^N$, $Q^N \leq Q^M$ for $Q \in \mathcal{S}_P$, and $Q^N = Q^M$ otherwise. Thus, formulas with only positive occurrences of P and only negative occurrences of predicates $Q \in \mathcal{S}_P$ have a larger truth value in N than in M.

To verify that N is a model of $\mathfrak{D}$, it suffices to check that N satisfies all local closed world assumptions. Consider any instance of a local closed world assumption.

$$\varphi \equiv \neg\Psi_Q[\bar{d}] \vee \neg Q(\bar{d}) \vee (Q(\bar{d}) \in D)$$

Each of these formulas is satisfied in M. Let us verify that it is satisfied in N as well. There are four cases:

— $Q = P$ and $\bar{d} = \bar{d}$ (i.e., Ψ_Q is P's window of expertise): in this case, $M \models \neg\Psi_P[\bar{d}]$. The formula $\neg\Psi_P[\bar{d}]$ contains only positive occurrences of P and only negative occurrences of predicates $Q \in \mathcal{S}_P$, hence $\mathbf{t} = (\neg\Psi_P[\bar{d}])^M \leq (\neg\Psi_P[\bar{d}])^N$.
— $Q = P$ and $\bar{d} \neq \bar{d}$: we have $(\neg P(\bar{d}) \vee (P(\bar{d}) \in D))^M = (\neg P(\bar{d}) \vee (P(\bar{d}) \in D))^N$ and $\neg\Psi_P[\bar{d}]$ contains only positive occurrences of P and only negative occurrences of predicates $Q \in \mathcal{S}_P$, hence $\neg\Psi_P[\bar{d}]^M \leq \neg\Psi_P[\bar{d}]^N$ and it follows that $\mathbf{t} = \varphi^M \leq \varphi^N$.
— $Q \in \mathcal{S}_P$: N satisfies $\neg Q(\bar{d}) \vee (Q(\bar{d}) \in D)$, hence $\varphi^N = \mathbf{t} = \varphi^M$.
— $Q \notin \mathcal{S}_P$ and $Q \neq P$: φ contains only positive occurrences of P and only negative occurrences of predicates of $\mathcal{S}_P$; hence $\mathbf{t} = \varphi^M \leq \varphi^N$. □

Thus, P should not occur positively in its own window of expertise or of any predicate on which P negatively depends. This condition is satisfied if Ψ_P is a positive formula not containing P, or if $P \not\prec P$, that is, the full dependency graph has no cycle in P.

The following corollary combines the preceding propositions.

COROLLARY A.1. *For predicate P, it holds that $P^{\mathcal{C}_{\mathfrak{D}}} = P^{\mathcal{O}_{\mathfrak{D}}}$ if the following conditions are satisfied:*

—$P \not\in^+ \Psi_P$,
—*for all* $Q \prec^- P$, $P \not\in^+ \Psi_Q$,
—$\Psi_P[\bar{x}]$ *is squared in* $\mathfrak{D}$, *and*
—*for every predicate* $Q \in^- \Psi_P$, $Q^{\mathcal{C}_{\mathfrak{D}}} = Q^{\mathcal{O}_{\mathfrak{D}}}$.

This corollary reduces the problem of proving optimality of $\mathcal{C}_{\mathfrak{D}}$ in a predicate P to the problem of proving optimality of $\mathcal{C}_{\mathfrak{D}}$ in predicates $Q \in^- \Psi_P$. It allows us to prove the optimality theorem.

PROOF OF THEOREM 8. The theorem states: Let $\mathfrak{D} = (D, \mathcal{L})$ be a locally closed database. It holds that $P^{\mathcal{C}_{\mathfrak{D}}} = P^{\mathcal{O}_{\mathfrak{D}}}$ if $\prec^-$ is acyclic in $\prec^-_{\cup} P$[2] and for each $Q \in \prec^-_{\cup} P$:

—$Q \not\in^+ \Psi_Q$,
—for all $R \prec^- Q$, $Q \not\in^+ \Psi_R$,
—$\Psi_Q[\bar{x}]$ is squared in $\mathfrak{D}$.

To prove it, observe that $\prec^-$ is a strict (well-founded) order on the set $\prec^-_{\cup} P = \{P\} \cup \{Q \mid Q \prec^- P\}$. It is easy to see that for each predicate in this set, the first three conditions of Corollary A.1 hold. By iterated application of this corollary along $\prec^-$, we obtain that $Q^{\mathcal{C}_{\mathfrak{D}}}$ is optimal, for all Q in this set. The theorem follows. □

Example A.10. The windows of expertise in the database of Example 11 and also of $\mathfrak{D}_1$ in Example A.5 are positive formulas. Hence, although $\prec$ is cyclic, the predicates *Loc* and *CarO* have optimal interpretation in $\mathcal{C}_{\mathfrak{D}}$. Nevertheless, as shown in Example A.6, the Kleene-precise query $\neg(Loc(LD, Bx) \wedge CarO(LD, V, B1))$ cannot be answered accurately by the approximate methods because $\mathfrak{D}_1$ is not atomical.

By a combination of Theorem 8 and Theorem A.1, we also have the following result.

THEOREM A.2 (COMPLETENESS IN PARTIALLY ATOMIC DATABASES). *It holds that* $Cert_{\mathcal{C}_{\mathfrak{D}}}(\mathcal{Q}(\bar{x})) = Cert_{\mathfrak{D}}(\mathcal{Q}(\bar{x}))$ *if: (i)* $\prec$ *is acyclic in* $\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))$ *and* $\mathfrak{D}$ *conveys CWI on* Ψ_P *for each* $P \in \mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))$ *and (ii) only base predicates of* $\mathfrak{D}$ *occur both positively and negatively in* $\mathcal{Q}(\bar{x})$ *and in* Ψ_P, *for each* $P \prec^- \mathcal{Q}(\bar{x})$.

PROOF. The conditions of this theorem entail those of Theorem A.1. As a consequence, $\mathfrak{D}_{\mathcal{Q}}$ is atomical and $\mathcal{Q}(\bar{x})$ is squared in $\mathfrak{D}$. Since $\prec$ is acyclic in $\mathcal{N}eg\mathcal{B}(\mathcal{Q}(\bar{x}))$, the acyclicity conditions of Theorem 8 and Theorem 9 are trivially satisfied, and by the atomicality of $\mathfrak{D}_{\mathcal{Q}}$, for each database predicate $P \prec^- \mathcal{Q}(\bar{x})$, Ψ_P is squared in $\mathfrak{D}$. It follows from Theorem 8 that $P^{\mathcal{C}_{\mathfrak{D}}}$ is optimal for each $P \prec^- \mathcal{Q}(\bar{x})$, and by Theorem A.1 that $Cert_{\mathcal{C}_{\mathfrak{D}}}(\mathcal{Q}(\bar{x})) = Cert_{\mathfrak{D}}(\mathcal{Q}(\bar{x}))$. □

[2]Recall that this set is defined as $\{P\} \cup \{Q | Q \prec^- P\}$.

B. A CIRCUMSCRIPTIVE APPROACH TO THE LCWA

In this appendix, we consider an alternative approach, based on circumscription [McCarthy 1980; Lifschitz 1994], for representing the closed world assumption by second-order formulas. As shown shortly, this approach is equivalent to the LCWA representation in Section 2.

Consider, for instance, the expression $\mathcal{LCWA}(CarO(x, y, z), x = MC)$ in the context of Example 2. The meaning of this local closed world assumption may be rephrased by stating that the restriction of the predicate $CarO(x, y, z)$ to the tuples with $x = MC$ should be *minimally* satisfying the database, that is, no tuples can be deleted from it without falsifying tuples from D. This minimization can be expressed through a circumscription-like second-order axiom, as given in the next definition. It is not surprising, therefore, that a variant of the notion of local closed world assumption presented here has already been expressed in terms of circumscription (see the discussion on the work of Doherty et al. [2000] in Section 6.)

Definition B.1 (*Pseudocircumscriptive Form of a LCWA*). Let $\theta = \mathcal{LCWA}$ $(P(\bar{x}), \Psi[\bar{x}])$ be a local closed world assumption for a database instance D. The *pseudocircumscriptive form* of θ is the following (second-order) formula, denoted $C_D(\theta)$:

$$\forall X : \left(\bigwedge_{P(\bar{d}) \in D} X(\bar{d}) \supset (\forall \bar{x} : (\Psi(\bar{x}) \supset (X(\bar{x}) \supset P(\bar{x}))) \supset \right.$$

$$\left. \forall \bar{x} : (\Psi(\bar{x}) \supset (P(\bar{x}) \supset X(\bar{x})))) \right),$$

where X is a predicate variable with the same arity as P.

$C_D(\theta)$ is called pseudocircumscriptive since it differs from a pure circumscription schema by introducing the first-order formula Ψ into the representation. Just as in Definition 4, Ψ represents the context in which P should be minimal.

The axiom states that for each relation X that contains all tuples in P^D, if X is smaller than P within the window of expertise of P, then P is smaller than (and hence, identical to) X within this window. Stated in a contrapositive way, it is impossible to delete from P a nonempty set of tuples in $P's$ window of expertise without violating the database.

Definition B.2. Let $\mathfrak{D} = (D, \mathcal{L})$ be a locally closed database. Denote $\mathfrak{C}(\mathfrak{D})$ the following set of sentences:

Soundness:	$\bigwedge_{A \in D} A$
Local Completeness:	$\bigwedge_{\theta \in \mathcal{L}} \mathcal{C}_D(\theta)$
Domain Closure Axiom (DCA($Dom^{\mathfrak{D}}$)):	$\forall x : (\bigvee_{C \in Dom^{\mathfrak{D}}} x = C)$
Unique Name Axiom (UNA($Dom^{\mathfrak{D}}$)):	$\bigwedge_{C \neq C' \in Dom^{\mathfrak{D}}} C \neq C'$

THEOREM B.1. *For every database $\mathfrak{D}$, $\mathcal{M}(\mathfrak{D})$ is equivalent to $\mathfrak{C}(\mathfrak{D})$.*

PROOF. The two theories are identical with exception of the local completeness axioms. Let M be any model satisfying the soundness axiom $\bigwedge_{A \in D} A$. We will show that M satisfies

$$\forall \bar{x} : (\Psi(\bar{x}) \supset (P(\bar{x}) \supset (P(\bar{x}) \in D))) \tag{1}$$

iff it satisfies

$$\textit{for all } X : \left(\underbrace{\bigwedge_{P(\bar{d}) \in D} X(\bar{d})}_{(a)} \supset \left(\underbrace{\forall \bar{x} : (\Psi(\bar{x}) \supset (X(\bar{x}) \supset P(\bar{x}))}_{(b)} \supset \underbrace{\forall \bar{x} : (\Psi(\bar{x}) \supset (P(\bar{x}) \supset X(\bar{x})))}_{(c)} \right) \right). \tag{2}$$

Indeed, suppose first that M satisfies the formula in (1) and that M' is an extension of M with an interpretation for X that satisfies the subformula (a) of (2). To show that M' satisfies (b) $\supset$ (c), we show the stronger proposition that M' satisfies (c). Let M'' extend M' with an arbitrary interpretation for $\bar{x}$, and assume that it satisfies $\Psi[\bar{x}]$ and $P(\bar{x})$. Then by the formula in (1), M'' satisfies $P(\bar{x}) \in D$, and by (a), it also satisfies $X(\bar{x})$. Hence M' satisfies (c).

For the converse, assume that M satisfies the formula in (2). Extend M by interpreting X by the relation P^D. Then the formula (a) is true in $M[X : P^D]$. Moreover, by the soundness axiom, $M[X : P^D]$ satisfies (b), so (c) is true as well. Now, given the value of X in $M[X : P^D]$, the formula (c) is equivalent to the formula in (1), hence the latter is satisfied in M. □

An interesting aspect of the pseudocircumscriptive formula of Definition B.1 is that it allows to extend the concept of LCWA to knowledge bases consisting of first-order axioms rather than atoms. Indeed, assume that the database instance D is an arbitrary first-order theory. Using $D[P/X]$ to denote the substitution of all occurrences in D of atoms $P(\bar{t})$ by $X(\bar{t})$, we could then express the LCWA $\mathcal{LCWA}(P(\bar{x}), \Psi[\bar{x}])$ by the axiom

$$\forall X : \left(D[P/X] \supset \left(\forall \bar{x} : \left(\Psi(\bar{x}) \supset (X(\bar{x}) \supset P(\bar{x})) \right) \supset \forall \bar{x} : \left(\Psi(\bar{x}) \supset (P(\bar{x}) \supset X(\bar{x})) \right) \right) \right).$$

For instance, consider a disjunctive database with $D = \{P(A, C) \vee P(B, C)\}$, and the assumption $\mathcal{LCWA}(P(x, y), y = C)$. Intuitively, this database expresses that $P(A, C)$ or $P(B, C)$ is true and that the set of tuples (x, C) in P is minimal. This means that the only value x for which $P(x, C)$ is true is either A or B *but not both*.