On the Relationship between Consistent Query Answering and Constraint Satisfaction Problems

Carsten Lutz[1] and Frank Wolter[2]

1 **Fachbereich Informatik, Universität Bremen, Germany**
clu@uni-bremen.de

2 **Department of Computer Science, University of Liverpool, UK**
wolter@liverpool.ac.uk

Abstract

Recently, Fontaine has pointed out a connection between consistent query answering (CQA) and constraint satisfaction problems (CSP) [23]. We investigate this connection more closely, identifying classes of CQA problems based on denial constraints and GAV constraints that correspond *exactly* to CSPs in the sense that a complexity classification of the CQA problems in each class is *equivalent (up to FO-reductions)* to classifying the complexity of all CSPs. We obtain these classes by admitting only monadic relations and only a single variable in denial constraints/GAVs and restricting queries to hypertree UCQs. We also observe that dropping the requirement of UCQs to be hypertrees corresponds to transitioning from CSP to its logical generalization MMSNP and identify a further relaxation that corresponds to transitioning from MMSNP to GMSNP (also know as MMSNP_2). Moreover, we use the CSP connection to carry over decidability of FO-rewritability and Datalog-rewritability to some of the identified classes of CQA problems.

1998 ACM Subject Classification H.2.4 [Systems]: Relational databases

Keywords and phrases Consistent Query Answering, Constraint Satisfaction, Data Complexity, Dichotomies, Rewritability

Digital Object Identifier 10.4230/LIPIcs.ICDT.2015.1

1 Introduction

In modern applications of database systems, it cannot always be guaranteed that the data is consistent with the relevant integrity constraints; for example, inconcistency occurs easily when the data is extracted from the web or integrated from multiple sources. A prominent approach to address this problem is consistent query answering (CQA) as introduced in [3] where one returns the certain answers over all *minimal repairs* of the inconsistent database, see also the surveys [7, 16, 42]. Since the data complexity of CQA can be coNP-complete or higher [2, 15, 17, 39], CQA is in general significantly harder than traditional query answering. This observation has resulted in a lot of research activity aiming to more precisely clarify the computational complexity of CQA, separating in particular the easy cases from the hard ones. Here, 'easy' might mean different things. The ideal result is that a CQA problem $\text{CQA}(C, q)$, defined by a set C of integrity constraints and a query q, is rewritable into a first-order logic (FO) query $\widehat{q}$ and thus answers can be computed by a classical RDBMS and in AC_0 data complexity [25, 40]. If FO-rewritability is not attainable, one might at least hope for Datalog-rewritability or PTime data complexity. Ultimate goals of this research programme would be to classify the exact complexity of *every* CQA problem and, closely related, to *decide* for a given CQA problem whether it is easy in some relevant sense, say whether it admits an FO-rewriting. Completely classifying the border between PTime and coNP

18th International Conference on Database Theory (ICDT'15).
Editors: Marcelo Arenas and Martin Ugarte; pp. 1–32
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

prominently involves solving dichotomy questions: for some important classes of integrity constraints and queries, it has been conjectured that CQA is in PTime or coNP-hard for every problem in the class [1, 41, 42]. In fact, with today's methods one cannot hope to completely classify the complexity of a class of CQA problems if that class does *not* have such a dichotomy.

Despite serious efforts, comprehensive results for dichotomies in CQA have so far been elusive. For several restricted cases which all require that the query to be answered is free of self-joins, dichotomies were obtained in [28, 41, 29]. An explanation of why general dichotomy results for CQA are difficult to obtain was recently given by Fontaine [23], who linked CQA with the area of constraint satisfaction problems (CSPs). CSPs constitute a subclass of NP that contains many relevant NP-complete problems such as 3SAT, 3COL, and integer programming over bounded domains, and it is widely believed that this class is computationally more well-behaved than NP itself. In particular, a long-standing conjecture due to Feder and Vardi states that CSPs enjoy a dichotomy between PTime and NP [21]. Massive research efforts in logic, complexity, and algebra have been directed towards proving this conjecture, and although steady progress has been made, the conjecture is still open. In contrast, strong results have been obtained on the FO-definability and Datalog-definability of CSPs, the counterpart of FO- and Datalog-rewritability in CQA: while both problems are undecidable for the entire class NP, they are decidable and NP-complete for CSPs [31, 5, 24].

Fontaine's main result in [23] links the Feder-Vardi conjecture to dichotomies in CQA under GAV constraints by showing that for every CSP problem CSP(A) defined by some template A, there is a CQA problem CQA(C, q) with C a set of GAV constraints and q a union of conjunctive queries (UCQ) such that CQA(C, q) and the complement coCSP(A) of CSP(A) are PTime-equivalent, that is, they have the same complexity up to PTime-reductions. Consequently, establishing a dichotomy between PTime and coNP for CQA with GAV constraints and UCQs implies the Feder-Vardi conjecture. The aim of this paper is to study the CSP-CQA connection in more detail, focussing on denial constraints and GAV constraints which have both received significant attention in CQA [18, 17, 8, 2, 38, 39, 23]. In particular, we aim to identify classes of CQA problems that *exactly* correspond to the class of CSPs in the sense that classifying the complexity of problems from both classes is equivalent in a strong sense.

Our first main observation is that CSPs correspond to CQA problems whose (denial or GAV) constraints involve *only monadic relation symbols* and *only a single variable* and in which the query to be answered is a *UCQ in which all CQs take the form of a hypertree* (all queries in this paper are Boolean). More specifically, let a *monadic disjointness constraint (MDiC)* be of the form $\forall x\, \neg(P_1(x) \wedge \cdots \wedge P_n(x))$ and a *monadic GAV constraint (MGAV)* be of the form $\forall x\, (P_1(x) \wedge \cdots \wedge P_n(x) \rightarrow Q(x))$. We use (MDiC, tUCQ) to denote the class of all CQA problems whose constraints are MDiCs and whose query is a UCQ in which every CQ is a hypertree in the sense that its incidence graph is a tree (without multi-edges), and likewise for (MGAV, tUCQ). We then show that (i) for every CSP(A) there is a problem CQA(C, q) in (MDiC, tUCQ) such that coCSP(A) and CQA(C, q) are FO-equivalent (that is, they have the same complexity up to FO-reductions), (ii) for every problem in (MDiC, tUCQ) there is one from (MGAV, tUCQ) that is FO-equivalent, and (iii) for every problem CQA(C, q) in (MGAV, tUCQ) there is a CSP(A) such that CQA(C, q) and coCSP(A) are FO-equivalent. This improves upon the main result of [23] in several ways. First, the CQA problems constructed in [23] involve constraints that contain non-monadic relations and the UCQs used there are not hypertrees since they include multi-edges; we thus identify much more restricted CQA classes whose complexity classification is already as hard as classifying CSPs; second, we also translate *from CQA to CSP* and thus show that a PTime vs. coNP dichotomy for (MDiC, tUCQ) and (MGAV, tUCQ) is *equivalent* to the Feder-Vardi conjecture (instead of only implying it); and third, we replace PTime-equivalence with FO-equivalence which (a) means that not only a PTime vs. (co)NP dichotomy carries over, but *any* complexity classification that involves complexity classes closed under FO-reductions (such as LogSpace), and (b) enables us to transfer results from CSP to the classes of CQA problems that we have identified.

Regarding (b), we show that for (MDiC, tUCQ) and (MGAV, tUCQ), rewritability into FO and into Datalog are decidable, referring to the version of Datalog which allows negation in front of body atoms that use a monadic EDB relation. Our approach yields NExpTime upper bounds for these problems and we demonstrate that the complexity is indeed high (despite the fact that we deal with rather restricted CQA problems) by establishing a PSpace lower bound for FO-rewritability, leaving the exact complexity open. We also transfer Bulatov's result that CSPs whose templates have at most three elements enjoy a dichotomy between PTime and NP [14] to CQA, identifying a (rather restricted) corresponding fragment of (MDiC, tUCQ) that has a dichotomy between PTime and coNP.

We then investigate the effect of dropping the requirement that queries are hypertrees while maintaining all restrictions on integrity constraints, which gives rise to the classes of CQA problems (MDiC, UCQ) and (MGAV, UCQ). We show that this corresponds to transitioning from CSP to its logical generalization MMSNP, which was introduced by Feder and Vardi when studying the descriptive complexity of CSP [21] and has received considerable interest, see e.g. [34, 11]. More precisely, we establish results that exactly parallel (i) to (iii) above, replacing tUCQs with UCQs and CSP with MMSNP. Again, all reductions are FO-reductions. The known results that CSP $\subseteq$ MMSNP and that for every MMSNP problem there is a CSP problem that is PTime-equivalent [21, 30] then also yields that (MDiC, UCQ) has a dichotomy between PTime and coNP if and only if this is the case for (MDiC, tUCQ), and likewise for the corresponding CQA classes based on MGAVs. This does not imply, though, that a *full* complexity classification of these classes is equivalent since the relation between CSP and MMSNP in terms of FO-reductions is open. These results also shed some light on the decidability of FO- and Datalog-rewritability in (MDiC, UCQ) and (MGAV, UCQ), which is equivalent to the decidability of FO-definability of MMSNP problems, an open problem. Finally, we generalize (MDiC, UCQ) by giving up the restriction that integrity constraints are monadic and comprise only a single variable, instead requiring that every atom in the integrity constraint comprises the same variables in the same order. We then show that this corresponds to the transition from MMSNP to GMSNP [10] (also known as MMSNP_2 [33]), which means to replace monadicity as stipulated in MMSNP for certain relations with a guardedness condition.

Some proof details are deferred to the appendix of the long version of this paper available at http://www.informatik.uni-bremen.de/tdki/research/papers.html.

2 Preliminaries

A *schema* is a finite collection $\mathbf{S} = (S_1, \dots, S_k)$ of relation symbols with associated non-zero arity. A *fact* over $\mathbf{S}$ is an expression of the form $S(a_1, \dots, a_n)$ where $S \in \mathbf{S}$ is an n-ary relation symbol, and $a_1, \dots, a_n$ are elements of some fixed, countably infinite set const of *constants*. An *instance* I over $\mathbf{S}$ is a finite set of facts over $\mathbf{S}$. The *active domain* $\mathsf{adom}(I)$ of I is the set of all constants that occur in the facts of I. We will frequently use boldface notation for tuples, such as in $\mathbf{a} = a_1 \cdots a_n$.

A *conjunctive query (CQ)* takes the form $q = \exists \mathbf{x}\, \varphi(\mathbf{x})$ where φ is a conjunction of relational atoms, neither constants nor equality allowed. A *union of conjunctive queries (UCQ)* is a disjunction of CQs. Note that we consider only Boolean queries for simplicity, see the conclusion for some further remarks on this issue.

A *denial constraint (DC)* has the form $\forall \mathbf{x}\, \neg\varphi(\mathbf{x})$, where φ is a conjunction of relational atoms. A *global as view constraint (GAV)* takes the form $\forall \mathbf{x}\, \varphi(\mathbf{x}) \rightarrow S(\mathbf{x})$ where φ is a conjunction of relational atoms. Let I be an instance and C a set of constraints. An instance J is a *minimal repair of I w.r.t. C* if J satisfies all constraints in C and there is no instance J' such that J' satisfies all constraints in C and $I \,\Delta\, J' \subsetneq I \,\Delta\, J$, where '$\Delta$' denotes symmetric difference. We generally omit 'w.r.t. C' when C is clear from the context. For a query q, we write $I \models_C q$ if every minimal repair J of I satisfies $J \models q$.

A *consistent query answering (CQA) problem*, denoted $\text{CQA}(C, q)$, is defined by a set of constraints C and a query q. As input, an $\mathbf{S}$-instance I is given where $\mathbf{S}$ is the set of relation symbols used in C or q. The question is whether $I \models_C q$. We use (DC, UCQ) to denote the set of problems $\text{CQA}(C, q)$ where C is a set of DCs and q a UCQ, and likewise for (GAV, UCQ) and other combinations of constraint language and query language.

In this paper, all considered decision problems take instances over some fixed schema as inputs. For two such decision problems P_1, P_2, we write $P_1 \preceq_p P_2$ if P_1 reduces to P_2 by a polynomial time reduction. We write $P_1 \preceq_{\mathsf{FO}} P_2$ if P_1 reduces to P_2 by an *FO-reduction*, defined as in [27]. However, most of our FO-reductions are of the following simple form. For problems P_1 and P_2 over schemas $\mathbf{S}_1$ and $\mathbf{S}_2$, a map T that assigns to each $\mathbf{S}_1$-instance I an $\mathbf{S}_2$-instance $T(I)$ is *FO-definable* if for every k-ary relation symbol R in $\mathbf{S}_2$, there is an FO-formula φ_R over $\mathbf{S}_1$ (equality and constants allowed) with k free variabes such that $R(\mathbf{a}) \in T(I)$ iff $I \models \varphi_R[\mathbf{a}]$, for all $\mathbf{a}$ in $\mathsf{adom}(I)$. Such a map gives rise to an FO-reduction of P_1 to P_2 if for all $\mathbf{S}_1$-instances I, we have $I \in P_1$ iff $T(I) \in P_2$. FO-reductions of this simple form differ from the general case in that (i) no arithmetic operations are admitted in the formulas φ_R and (ii) the domain of the $T(I)$ cannot be larger than the domain of I. With *FO-equivalence* and PTIME-*equivalence* of two problems P_1 and P_2, denoted $P_1 \approx_{\mathsf{FO}} P_2$ and $P_1 \approx_p P_2$, we mean that there are reductions between P_1 and P_2 in both directions. For two classes of decision probems $\mathcal{C}_1$ and $\mathcal{C}_2$, we write $\mathcal{C}_1 \preceq_{\mathsf{FO}} \mathcal{C}_2$ if for every problem $p_1 \in \mathcal{C}_1$, there is a problem $p_2 \in \mathcal{C}_2$ such that $p_1 \preceq_{\mathsf{FO}} p_2$ and $p_2 \preceq_{\mathsf{FO}} p_1$. We write $\mathcal{C}_1 \approx_{\mathsf{FO}} \mathcal{C}_2$ and say that $\mathcal{C}_1$ and $\mathcal{C}_2$ are *FO-equivalent* if $\mathcal{C}_1 \preceq_{\mathsf{FO}} \mathcal{C}_2$ and $\mathcal{C}_2 \preceq_{\mathsf{FO}} \mathcal{C}_1$. The definition of $\mathcal{C}_1 \preceq_p \mathcal{C}_2$ and $\mathcal{C}_1 \approx_p \mathcal{C}_2$ is analogous, but based on polynomial time reductions.

Let A be an instance over schema $\mathbf{S}$. The *constraint satisfaction problem* $\text{CSP}(A)$ is to decide, given an instance I over $\mathbf{S}$, whether there is a homomorphism from I to A, which we denote with $I \rightarrow A$. In this context, A is called the *template* of $\text{CSP}(A)$. We will generally and w.l.o.g. assume that the template is a core, that is, every automorphism is an isomorphism. It is often useful to further assume that the template A *admits precoloring*,[1] that is, for each $a \in \mathsf{adom}(A)$, there is a unary relation symbol $P_a \in \mathbf{S}$ such that $P_a(b) \in A$ iff $b = a$ [19]. It is known that for every template A (which is a core), there is a template A' that admits precoloring such that $\text{CSP}(A) \approx_{\mathsf{FO}} \text{CSP}(A')$ [32]. We use $\text{coCSP}(A)$ to denote the complement problem of $\text{CSP}(A)$ and coCSP to denote the set of all problems $\text{coCSP}(A)$ whose template A admits precoloring.

The logic MMSNP was introduced by Feder and Vardi as a descriptive complexity counterpart of CSPs [21]. Since we will mostly be concerned with the *complement* of MMSNP, we refrain from giving the original definition and directly introduce its complement,

[1] This property is also known as 'full idenpotence' and 'pointedness'.

which can conveniently be defined as (negation-free) monadic disjunctive Datalog [10]. A *monadic disjunctive Datalog (MDDLog) rule* ρ has the form $R_1(\mathbf{y}_1) \wedge \cdots \wedge R_n(\mathbf{y}_n) \to S_1(x_1) \vee \cdots \vee S_m(x_m)$ with $m \geq 0$, $n > 0$, and $S_1, \ldots, S_m$ monadic. Every variable that occurs in the *head* $S_1(x_1) \vee \cdots \vee S_m(x_m)$ of ρ is also required to occur in ρ's *body* $R_1(\mathbf{y}_1) \wedge \cdots \wedge R_n(\mathbf{y}_n)$. Empty rule heads are denoted $\bot$. An *MDDLog program* Π is a finite set of MDDLog rules with a selected nullary *goal relation* goal that does not occur in rule bodies and only in *goal rules* of the form $R_1(\mathbf{x}_1) \wedge \cdots \wedge R_n(\mathbf{x}_n) \to \mathsf{goal}()$. Relation symbols that occur in the head of at least one rule of Π are *intensional (IDB)*, and all remaining relation symbols in Π are *extensional (EDB)*. Every MDDLog program Π defines a decision problem: given an $\mathbf{S}$-instance I, where $\mathbf{S}$ is the set of EDB relations in Π, decide whether $I \models \Pi$, that is, whether in every extension of I that satisfies all non-goal rules in Π, the body of at least one goal rule applies. We use coMMSNP to denote the class of all these decision problems and MMSNP to denote the class of their complements.

3 Relating CQA and CSP

We identify fragments of (DC, UCQ) and (MGAV, UCQ) that are FO-equivalent to coCSP. They involve restrictions on the admitted constraints as well as on the admitted queries. A denial constaint is called a *monadic disjointness constraint (MDiC)* if all relation symbols that occur in it are monadic and it contains only a single variable; a GAV is called *monadic* or an *MGAV* if it satisfies the same conditions. MDiCs and MGAVs are clearly rather restricted classes of constraints. Note, however, that they are useful for speaking about *type systems*. For example, the MDiC $\forall x \, \neg(\mathsf{person}(x) \wedge \mathsf{process}(x))$ asserts disjointness of the types person and $\mathsf{process}$ while the MGAV $\forall x \, (\mathsf{professor}(x) \to \mathsf{person}(x))$ ensures a subsumption between the types $\mathsf{professor}$ and person.

While we restrict constraints to MDiCs and MGAVs, UCQs are required to contain only CQs that have the shape of a hypertree. The *incidence graph* of a CQ q is the bipartite undirected graph whose nodes are the variables and the atoms in q and whose edges connect a variable x with each atom $R(\mathbf{x})$ such that $x \in \mathbf{x}$. A CQ is a *hypertree conjunctive query (tCQ)* if its incidence graph is a tree (without multi-edges). A *hypertree UCQ (tUCQ)* is a UCQ in which every CQ is a tCQ. Our notion of hypertree CQ is rather restrictive, see e.g. [6, 26] for more general forms of hypertrees; our form of hypertrees, though, is known to be intimately connected to the expressive power of CSPs [35].

Before proceeding, we observe the following connection between MDiCs and MGAVs.

▶ **Lemma 1.** *(MDiC, $\mathcal{Q}$) $\preceq_{\mathsf{FO}}$ (MGAV, $\mathcal{Q}$) for $\mathcal{Q} \in \{$UCQ, tUCQ$\}$. Moreover, given a problem CQA(C, q) from (MDiC, $\mathcal{Q}$) one can construct a problem CQA(C', q') in (MGAV, $\mathcal{Q}$) such that CQA$(C, q) \approx_{\mathsf{FO}}$ CQA(C', q') in polynomial time.*

Proof. Let CQA(C, q) be a problem from (MDiC, $\mathcal{Q}$) over schema $\mathbf{S}$. Define $C' = \{\forall x \, \varphi(x) \to M(x) \mid \forall x \, \neg\varphi(x) \in C\}$ where M is a fresh monadic relation symbol and $q' = q \vee \exists x \, M(x)$. We show that CQA$(C, q) \preceq_{\mathsf{FO}}$ CQA(C', q') and vice versa. For the former, we observe that for all $\mathbf{S}$-instances I, we have $I \models_C q$ iff $I \models_{C'} q'$. For the latter, it is easy to show that, for all $\mathbf{S} \cup \{M\}$-instances I, we have $I \models_{C'} q'$ iff $I \models \exists x \, M(x)$ or $I \models_C q$. Clearly, these reductions can be implemented as FO-reductions. ❑

Our aim is to show that (MDiC, tUCQ) $\approx_{\mathsf{FO}}$ (MGAV, tUCQ) $\approx_{\mathsf{FO}}$ coCSP. By Lemma 1, it suffices to show that coCSP $\preceq_{\mathsf{FO}}$ (MDiC, tUCQ) and (MGAV, tUCQ) $\preceq_{\mathsf{FO}}$ coCSP. We start with the former, improving upon a reduction by Fontaine [23] which shows that coCSP $\preceq_p$ (GAV, UCQ). Consider CSP(A) over schema $\mathbf{S}$ where A admits precoloring. We construct a problem $\mathrm{CQA}(C_A, q_A)$ from (MDiC, tUCQ) over schema $\mathbf{S}'$ which extends $\mathbf{S}$ with unary relation symbols Q_a, $a \in \mathsf{adom}(A)$ (these symbols should be distinguished from the monadic relation symbols P_a in $\mathbf{S}$, $a \in \mathsf{adom}(A)$, which exist since A admits precoloring). C_A contains one monadic disjointness constraint:

$$\forall x \neg \bigwedge_{a \in \mathsf{adom}(A)} Q_a(x).$$

For each $a \in \mathsf{adom}(A)$, we use $\mathsf{con}_a(x)$ to denote the conjunction $\bigwedge_{e \in \mathsf{adom}(A) \setminus \{a\}} Q_e(x)$. The tUCQ q_A contains the following tCQ for each $R \in \mathbf{S}$ of arity n and each $\mathbf{a} = a_1 \cdots a_n \in \mathsf{adom}(A)^n$ such that $R(\mathbf{a}) \notin A$:

$$\exists x_1 \cdots \exists x_n \, (\mathsf{con}_{a_1}(x_1) \wedge \cdots \wedge \mathsf{con}_{a_n}(x_n) \wedge R(x_1, \ldots, x_n)).$$

To understand the construction of $\mathrm{CQA}(C_A, q_A)$, consider the reduction from $\mathrm{CSP}(A)$ to the complement of $\mathrm{CQA}(C_A, q_A)$. Given an $\mathbf{S}$-instance I that is an input to $\mathrm{CSP}(A)$, we construct an $\mathbf{S}'$-instance $T^{\uparrow}(I)$ by adding $Q_a(b)$ for all $b \in \mathsf{adom}(I)$ and all $a \in \mathsf{adom}(A)$. Then each minimal repair J of $T^{\uparrow}(I)$ must satisfy, for each $b \in \mathsf{adom}(I)$, $J \models \mathsf{con}_a[b]$ for a unique $a \in \mathsf{adom}(A)$. In this way, J represents a function h that assigns to each $b \in \mathsf{adom}(I)$ the unique $a \in \mathsf{adom}(A)$ such that $Q_a(b) \notin J$. If $J \not\models q_A$, then h must clearly be a homomorphism. Indeed, we show in the appendix that $I \to A$ iff $T^{\uparrow}(I) \not\models_{C_A} q_A$.

▶ **Lemma 2.** *coCSP*$(A) \preceq_{\mathsf{FO}}$ *CQA*(C_A, q_A) *and CQA*$(C_A, q_A) \preceq_{\mathsf{FO}}$ *coCSP*(A).

Proof. The first reduction was already described above. Clearly, $T^{\uparrow}$ is FO-definable and thus the reduction is an FO-reduction. For the converse reduction, let I be an $\mathbf{S}'$-instance. Denote by X the set of $b \in \mathsf{adom}(I)$ such that there are at least two distinct $a_1, a_2 \in \mathsf{adom}(A)$ with neither $Q_{a_1}(b) \in I$ nor $Q_{a_2}(b) \in I$. Then $T^{\downarrow}(I)$ is obtained from I by dropping all facts that involve a constant from X or a relation symbol in $\mathbf{S}' \setminus \mathbf{S}$, and adding all facts $P_a(b)$ such that $Q_e(b) \in I$ iff $e \neq a$ for all $e \in \mathsf{adom}(A)$. Note that it is crucial here that A admits precoloring as we use the relation symbols P_a. Clearly $T^{\downarrow}(I)$ is FO-definable. We show in the appendix that $I \not\models_{C_A} q_A$ iff $T^{\downarrow}(I) \to A$. ❑

It is easy to see that, given A, we can construct $\mathrm{CQA}(C_A, q_A)$ in polynomial time. We have thus established the following.

▶ **Theorem 3.** *For every CSP template A that admits precoloring, there is a problem CQA(C, q) from (MDiC,tUCQ) that satisfies CQA$(C, q) \approx_{\mathsf{FO}}$ coCSP(A) and can be constructed in polynomial time.*

We now show that (MGAV, tUCQ) $\preceq_{\mathsf{FO}}$ coCSP. Let $\mathrm{CQA}(C, q_0)$ be over schema $\mathbf{S}$ with C a set of MGAVs and q_0 a hypertree UCQ. We use $\mathbf{S}^{(1)}$ to denote the restriction of $\mathbf{S}$ to monadic relation symbols and $\mathbf{S}^{(>1)}$ for the restriction of $\mathbf{S}$ to non-monadic symbols. In the following, we define a CSP template A_{C,q_0} over schema $\mathbf{S}' = \mathbf{S}^{(>1)} \cup \{P_\Gamma \mid \Gamma \subseteq \mathbf{S}^{(1)}\}$. Note that there is an obvious natural translation of $\mathbf{S}$-instances into corresponding $\mathbf{S}'$-instances and vice versa; for example, an $\mathbf{S}'$-instance I is translated to an $\mathbf{S}$-instance J by replacing every fact $P_\Gamma(a)$ with $P(a)$ for all $P \in \Gamma$. The change of schema is used to deal with the complication that the 'yes'-instances of each CSP are closed under homomorphic pre-images while the 'no'-instances of CQA problems from (MGAV, tUCQ) are not (unless the schema is modified in the described way). For any set $\Gamma \subseteq \mathbf{S}^{(1)}$, we use $\mathsf{rep}(\Gamma)$ to denote the set of all sets $\Gamma' \subseteq \mathbf{S}^{(1)}$ such that $\{P(a) \mid P \in \Gamma'\}$ is a minimal repair of the instance $\{P(a) \mid P \in \Gamma\}$.

Let Q be the set of all connected subqueries of CQs in q_0 (which are again hypertrees) and of all queries of the form $P(x)$, $P \in \mathbf{S}^{(1)}$. A *place* is a pair (q, x) with $q \in Q$ and $x \in \mathsf{var}(q)$.

A *type* t is a set of places. The type $\mathsf{tp}_I(a)$ *realized by constant* a *in instance* I is the set of all places (q, x) such that there is a homomorphism h from q to I that takes x to a. We say that a type t is *realizable* if there is an instance I such that I satisfies all constraints in C and t is realized by some constant in I; we say that t *avoids* q_0 if $(q, x) \notin t$ for any disjunct q of q_0 and $x \in \mathsf{var}(q)$. For $R \in \mathbf{S}$ of arity n, we say that a tuple $(t_1, \ldots, t_n)$ of types is R*-coherent* if for any instance I and tuples of constants $(a_1, \ldots, a_n)$ such that $\mathsf{tp}_I(a_i) = t_i$ for $1 \le i \le n$, after adding to I the fact $R(a_1, \ldots, a_n)$, we still have $\mathsf{tp}_I(a_i) = t_i$ for $1 \le i \le n$. Now, the template A_{C,q_0} is defined as follows:

- the constants in A_{C,q_0} are the pairs $\langle t, \Gamma\rangle$ with t a realizable type that avoids q_0 and $\Gamma \subseteq \mathbf{S}^{(1)}$ such that $t|_{\mathbf{S}^{(1)}} \in \mathsf{rep}(\Gamma)$ where $t|_{\mathbf{S}^{(1)}}$ is the restriction of t to schema $\mathbf{S}^{(1)}$;
- A_{C,q_0} contains all facts of the form $P_\Gamma(\langle t, \Gamma\rangle)$;
- A_{C,q_0} contains all facts $R(\langle t_1, \Gamma_1\rangle, \ldots, \langle t_n, \Gamma_n\rangle)$, R of arity n, if $(t_1, \ldots, t_n)$ is R-coherent.

To understand the construction, consider the reduction from the complement of CQA(C, q_0) to CSP(A_{C,q_0}) and let I be an $\mathbf{S}$-instance that is an input to the former. We replace I with the corresponding $\mathbf{S}'$-instance $T^\uparrow(I)$ obtained from I by dropping all facts that involve a monadic relation and adding $P_{\Gamma_a}(a)$ for every element $a \in I$, where $\Gamma_a = \{P \mid P(a) \in I\}$. Then, a homomorphism h from $T^\uparrow(I)$ to A_{C,q_0} defines the repair of I that is obtained by repairing the monadic relations at each $a \in \mathsf{adom}(I)$ as suggested by $h(a) = \langle t, \Gamma\rangle$, namely to remove all $P(a)$ with $(P(x), x) \notin t$. Indeed, we show in the appendix that $I \not\models_C q_0$ iff $T^\uparrow(I) \to A_{C,q_0}$. It is again easy to see that $T^\uparrow(I)$ is FO-definable. For later use, we note explicitly that an FO-formula φ_{P_Γ} for defining P_Γ in $T^\uparrow(I)$ is given by

$$\varphi_{P_\Gamma}(x) = \bigwedge_{P\in\Gamma} P(x) \wedge \bigwedge_{P\in\mathbf{S}^{(1)}\setminus\Gamma} \neg P(x).$$

It is also interesting to note that the $\mathbf{S}$-instance I_A obtained from the template A_{C,q_0} by dropping all facts $P_\Gamma(\langle t, \Gamma\rangle)$ and adding $P(\langle t, \Gamma\rangle)$ for all $(P(x), x) \in t$ is a *universal minimal repair* in the following sense: (i) it is a minimal repair of the $\mathbf{S}$-instance that corresponds to A_{C,q_0} and (ii) any minimal repair of any $\mathbf{S}$-instance homomorphically maps to I_A.

▶ **Lemma 4.** *CQA$(C, q_0) \preceq_{\mathsf{FO}}$ coCSP(A_{C,q_0}) and coCSP$(A_{C,q_0}) \preceq_{\mathsf{FO}}$ CQA(C, q_0).*

Proof. The first reduction was described above. Correctness is proved in full detail in the appendix. For the second reduction, let an $\mathbf{S}'$-instance I be given. If there exists $a \in \mathsf{adom}(I)$ with $P_\Gamma(a), P_\Delta(a) \in I$ for some $\Gamma \neq \Delta$, then there is no homomorphism from I to A_{C,q_0} and "false" is returned. Otherwise define an $\mathbf{S}$-instance $T^\downarrow(I)$ by dropping all facts of the form $P_\Gamma(a)$ and adding $P(a)$ whenever $P_\Gamma(a) \in I$ with $P \in \Gamma$. We show in the appendix that $T^\downarrow(I) \to A_{C,q_0}$ iff $I \not\models_C q_0$. Clearly $T^\downarrow(I)$ is FO-definable. ❑

▶ **Theorem 5.** *For every problem CQA(C, q) from (MGAV,tUCQ), there is a CSP template A that satisfies coCSP$(A) \approx_{\mathsf{FO}}$ CQA(C, q) and can be constructed in single exponential time.*

Summarizing Theorems 3 and 5, we thus obtain the following FO-equivalences.

▶ **Corollary 6.** *(MDiC, tUCQ) $\approx_{\mathsf{FO}}$ (MGAV, tUCQ) $\approx_{\mathsf{FO}}$ coCSP.*

It follows that there is a dichotomy between PTime and coNP for (MDiC, tUCQ) and (MGAV, tUCQ) if and only if the Feder-Vardi conjecture is true, and more generally that classifying the complexity of these classes of CQA problems is equivalent to classifying the complexity of CSPs (up to FO reductions).

4 FO- and Datalog-Rewritability

We exploit the reductions from the previous section and recent results concerning the FO- and Datalog-definability of CSPs to show that FO-rewritability and Datalog-rewritability are decidable in $(\text{MDiC}, \text{tUCQ})$ and $(\text{MGAV}, \text{tUCQ})$. A problem $\text{CQA}(C, q)$ over schema $\mathbf{S}$ is *FO-rewritable* if there is a Boolean FO-query $\widehat{q}_{C,q}$ such that for all $\mathbf{S}$-instances I, we have $I \models_C q$ iff $I \models \widehat{q}_{C,q}$. Thus, FO-rewritability ensures that $\text{CQA}(C, q)$ can be implemented using a conventional RDBMS. Datalog-rewritability is defined accordingly, where we refer to the version of Datalog that admits *negated monadic EDB atoms* in rule bodies. Without such atoms, Datalog-rewritability would be an extremely elusive property, as illustrated by the trivial problem $\text{CQA}(C, q)$ where $C = \{\forall x \, \neg(A_1(x) \wedge A_2(x))\}$ and $q = \exists x \, A_1(x)$, which can be rewritten into $A_1(x) \wedge \neg A_2(x) \rightarrow \mathsf{goal}()$, but not into any Datalog program without negated monadic EDB atoms. Note that in constrast to consistent query answering problems, a class $\text{coCSP}(A)$ is Datalog-definable with negated EDB atoms in rule bodies if and only if it is definable by a negation-free Datalog program [22]. We rely on the following results from [31, 5, 24].

▶ **Theorem 7.** *Given a CSP template A, it is* NP*-complete to decide whether coCSP(A) is FO-definable. The same is true for Datalog-definability.*

Since the reductions between $\text{CQA}(C, q)$ and $\text{coCSP}(A_{C,q})$ used in the proof of Theorem 5 are FO-reductions, it follows immediately that $\text{CQA}(C, q)$ is FO-rewritable if and only if $\text{coCSP}(A_{C,q})$ is FO-definable [27]. Given that $A_{C,q}$ can be constructed in single exponential time, we obtain a NExpTime upper bound for deciding FO-rewritability of CQA problems in $(\text{MDiC}, \text{tUCQ})$ and $(\text{MGAV}, \text{tUCQ})$.

▶ **Theorem 8.** *Given (C, q) such that $CQA(C, q)$ is from $(MDiC, tUCQ)$ or $(MGAV, tUCQ)$, it is decidable in* NExpTime *whether $CQA(C, q)$ is FO-rewritable. Both problems are* PSpace*-hard.*

Proof.(Sketch) The PSpace lower bounds are proved in the appendix by a non-trivial reduction of the word problem of polynomially space-bounded Turing machines. Similar reductions have been used to establish PSpace-hardness of boundedness in linear monadic datalog [20] and of certain FO-rewritability problems in ontology-based data access [9]. The general idea is to start with a DTM M that solves a PSpace-complete problem and to first modify it so that M terminates when started in *any* configuration (which must not even be reachable from an initial configuration). Then, one crafts a problem $\text{CQA}(C, q)$ such that if M accepts an input x, then any FO-rewriting of $\text{CQA}(C, q)$ would have to query for the existence of unboundedly long paths of facts that represent an accepting computation of M on x, repeated over and over again. Clearly, this contradicts the locality of FO-queries. For technical reasons, we actually cannot ensure that the first computation on the mentioned paths starts with the initial configuration for x. However, M terminates from any configuration, and the second computation on the paths (as well as all subsequent ones) are guaranteed to start with the initial configuration for x. Details are given in the appendix. We note that C consists only of a single constraint, which is of the form $\forall x \, \neg(B(x) \wedge B'(x))$. ❏

The exact complexity remains open, but we speculate that the problems in Theorem 8 are at least ExpTime-hard. To obtain complexity bounds for Datalog-rewritability, we inspect the FO-formulas required to define $T^{\uparrow}(I)$ and $T^{\downarrow}(I)$ in the proof of Lemma 4.

▶ **Lemma 9.** *Let CQA*(C, q_0) *and* A_{C,q_0} *be as in Lemma 4. Then CQA*(C, q_0) *is Datalog-rewritable iff coCSP*(A_{C,q_0}) *is Datalog-definable.*

Proof. Let Π be a Datalog program that defines coCSP(A_{C,q_0}). Replace in Π all body atoms $P_\Gamma(x)$ with $\bigwedge_{P\in\Gamma} P(x) \wedge \bigwedge_{P\in\mathbf{S}^{(1)}\setminus\Gamma} \neg P(x)$ and denote the resulting program by Π'. It can be verified using the proof of Lemma 4 that $I \models_C q_0$ iff $I \models \Pi'$ for all $\mathbf{S}$-instances I. Thus Π' is a Datalog-rewriting of CQA(C, q_0).

Conversely, assume that Π is a Datalog-rewriting of CQA(C, q_0). Replace every rule ρ in Π by a set of rules as follows. For every variable x in ρ, replace the set S of conjuncts in the body of ρ that are of the form $P(x)$ and $\neg P(x)$, $P \in \mathbf{S}^{(1)}$, by any atom $P_\Gamma(x)$ such that $S \subseteq \{P(x) \mid P \in \Gamma\} \cup \{\neg P(x) \mid P \in \mathbf{S}^{(1)} \setminus \Gamma\}$. Moreover, add $P_\Gamma(x) \wedge P_\Lambda(x) \rightarrow \mathsf{goal}()$ as a new rule for all $\Gamma, \Lambda \subseteq \mathbf{S}^{(1)}$ with $\Gamma \neq \Lambda$. Denote the resulting program by Π'. It can be verified using the proof of Lemma 4 that Π' defines coCSP(A_{C,q_0}). ❑

The following is a consequence of Lemma 4, Lemma 9, and Theorem 7.

▶ **Theorem 10.** *Given* (C, q) *such that CQA*(C, q) *is from* $(MDiC, tUCQ)$ *or* $(MGAV, tUCQ)$*, it is decidable in* NExpTime *whether CQA*(C, q) *is Datalog-rewritable.*

We do not have any non-trivial lower bound. Finally, we observe that, thanks to allowing negation in Datalog-rewritings as described above, FO-rewritability implies Datalog-rewritability.

▶ **Theorem 11.** *In* $(MDiC, tUCQ)$ *and* $(MGAV, tUCQ)$*, FO-rewritability implies (non-recursive) Datalog-rewritability.*

Proof. It was observed by Atserias (and follows from Rossman's homomorphism preservation theorem) that for any CSP template A, coCSP(A) being FO-definable implies that coCSP(A) is UCQ-definable [4, 36]. Thus FO-rewritability of CQA(C, q) implies FO-definability of coCSP$(A_{C,q})$. The latter implies UCQ-definability of coCSP$(A_{C,q})$ which implies (non-recursive) Datalog-definability of coCSP$(A_{C,q})$. The latter implies (non-recursive) Datalog-rewritability of CQA(C, q). ❑

5 A Dichotomy Result

For restricted classes of CSPs, a dichotomy between PTime and NP has been established. The most notable results include Schaefer's famous dichotomy theorem for templates with at most two elements [37] and its generalization to three element templates obtained much later by Bulatov [14]. It is natural to ask whether these dichotomies transfer to restricted classes of CQA problems. In this section, we identify a subclass of (MDiC, tUCQ) for which this is the case, thus obtaining a dichotomy between PTime and coNP.

We consider the class of problems CQA(C, q_0) such that C consists of a single MDiC of the form $\forall x\, \neg(A_1(x) \wedge A_2(x))$ and q_0 is a tUCQ such that, in every tCQ in q_0, there is at most one atom with a relation symbol distinct from A_1, A_2. Let us call a problem of this form a *restricted binary CQA problem* where 'binary' refers to the number of atoms allowed in the MDiC. Note that the resulting class r2CQA of CQA problems is not trivial since a straightforward analysis of the proof of Theorem 3 shows that 2coCSP $\preceq_{\mathsf{FO}}$ r2CQA, where $i$coCSP is the class of complements of all problems that can be defined by a CSP template with at most i elements. Consequently, establishing a PTime / coNP dichotomy for r2CQA implies Schaefer's theorem; we actually find it remarkable that a class of CQA problems as simple as 2rCQA turns out to be that complex. In the following, we show that r2CQA $\preceq_{\mathsf{FO}}$ 3coCSP and thus obtain a dichotomy between PTime and coNP for r2CQA by Bulatov's result.

Let CQA(C, q_0) be a restricted binary CQA problem over schema $\mathbf{S}$. We define a CSP template A_{C,q_0} over schema $\mathbf{S}' = (\mathbf{S} \setminus \{A_1, A_2\}) \cup \{P_\Gamma \mid \Gamma \subseteq \{A_1, A_2\}\}$ as follows:

1. the constants in A_{C,q_0} are $0, 1, 2$;
2. A_{C,q_0} contains the facts $P_\emptyset(0)$, $P_{\{A_1\}}(1)$, $P_{\{A_2\}}(2)$, $P_{\{A_1,A_2\}}(1)$, and $P_{\{A_1,A_2\}}(2)$;
3. A_{C,q_0} contains the fact $R(i_1, \dots, i_k)$, $R \in \mathbf{S} \setminus \{A_1, A_2\}$ and $i_j \in \{1, 2\}$, when q_0 does not evaluate to true on the $\mathbf{S}$-instance $\{R(i_1, \dots, i_k)\} \cup \{A_{i_j}(i_j) \mid 1 \leq j \leq k\}$.

The general idea is the same as in the proof of Theorem 5, that is, a homomorphism h from an $\mathbf{S}'$-instance J to A_{C,q_0} defines a repair of the corresponding $\mathbf{S}$-instance I. In fact, for any situation $A_1(a), A_2(a) \in I$ we must have $h(a) \in \{1, 2\}$ and in the repair of I described by h we then keep $A_{h(a)}$ and remove $A_{3-h(a)}$. The element 0 in A_{C,q_0} is needed as a homomorphism target for constants $a \in \mathsf{adom}(I)$ such that neither $A_1(a)$ nor $A_2(a)$ are in I.

▶ **Lemma 12.** *CQA(C, q_0) $\preceq_{\mathsf{FO}}$ coCSP(A_{C,q_0}) and coCSP(A_{C,q_0}) $\preceq_{\mathsf{FO}}$ CQA(C, q_0).*

The FO-reductions used for proving Lemma 12 are identical to those in the proof of Lemma 4, but of course the correctness proofs differ. Details are given in the appendix.

▶ **Theorem 13.** *Every binary restricted CQA problem is in* PTime *or* coNP*-complete.*

6 Relating CQA and MMSNP

The classes of CQA problems identified so far all require queries to be hypertree UCQs. We now show that the transition from hypertree UCQs to unrestricted UCQs corresponds to the transition from CSP to MMSNP: while classifying the complexity of (MDiC, tUCQ) and (MGAV, tUCQ) is equivalent to classifying the complexity of coCSP (up to FO-reductions), classifying (MDiC, UCQ) and (MGAV, UCQ) is equivalent in the same sense to classifying coMMSNP. Since it is known that MMSNP $\approx_p$ CSP, the results in this section also imply that there is a dichotomy between PTime and coNP for (MDiC, tUCQ) if and only if there is such a dichotomy for (MDiC, UCQ) if and only if the Feder-Vardi conjecture holds (and likewise for the corresponding CQA languages based on MGAVs). They also yield some insight into the problem of deciding FO-rewritability in (MDiC, UCQ) and (MGAV, UCQ): these problems are decidable if and only if FO-definability in MMSNP is decidable, which is an open problem.

Recall that coMMSNP is the class of problems definable by an MDDLog program. Thus, let Π be an MDDLog program over schema $\mathbf{S}$ and assume that $\mathbf{Q}$ is the set of IDB relations in Π. A $\mathbf{Q}$-*type* is a subset $t \subseteq \mathbf{Q}$. We say that Π *admits precoloring* if the EDB schema $\mathbf{S}$ includes a monadic relation symbol S_t for each $\mathbf{Q}$-type t and Π includes rules (i) $S_t(x) \rightarrow Q(x)$ for all $\mathbf{Q}$-types t and all $Q \in t$ and (ii) $S_t(x) \wedge Q(x) \rightarrow \bot$ for all $\mathbf{Q}$-types t and all $Q \in \mathbf{Q} \setminus t$; the S_t relations are not allowed to be used in any other rule. We use coMMSNP$^{\mathsf{pre}}$ to denote the class of all problems defined by a MDDLog program that admits precoloring and MMSNP$^{\mathsf{pre}}$ to denote the class of their complements. Recall that we can w.l.o.g. assume CSPs to admit precoloring. The following result says that the same is true for (co)MMSNP.

▶ **Theorem 14** (Bodirsky and Madelaine [13])**.** *MMSNP $\preceq_{\mathsf{FO}}$ MMSNP$^{\mathsf{pre}}$.*

We start with establishing a counterpart of Theorem 3. Let Π be an MDDLog program over schema $\mathbf{S}$ that admits precoloring with IDB relations $\mathbf{Q}$, as above. We use tp to denote the set of all $\mathbf{Q}$-types. Construct a problem CQA(C_Π, q_Π) in (MDiC, UCQ) over schema $\mathbf{S}'$ which extends $\mathbf{S}$ with unary relation symbols Q_t, $t \in \mathsf{tp}$ (to be distinguished from the EDB relations S_t required because Π admits precoloring). C_Π contains one monadic disjointness constraint:

$$\forall x \neg \bigwedge_{t \in \mathsf{tp}} Q_t(x).$$

For each $\mathbf{Q}$-type t, we use $\mathsf{con}_t(x)$ to denote the conjunction $\bigwedge_{t' \in \mathsf{tp} \setminus \{t\}} Q_{t'}(x)$. We now construct the UCQ q_Π, which contains the following two kinds of CQs.

1. Consider each non-goal rule

$$\rho = \bigwedge_{1 \leq i \leq n} R_i(\mathbf{x}_i) \wedge \bigwedge_{1 \leq i \leq m} S_i(y_i) \rightarrow \bigvee_{1 \leq i \leq \ell} S'_i(z_i)$$

   where all R_i are from $\mathbf{S}$ and all S_i and S'_i are from $\mathbf{Q}$. Let $x_1, \dots, x_k$ be the variables in ρ. Then include in q_Π the following CQ, for all sequences of $\mathbf{Q}$-types $t_1, \dots, t_k$ such that (i) if $S(x_i)$ occurs the body of ρ with $S \in \mathbf{Q}$, then $S \in t_i$ and (ii) if $S(x_i)$ occurs in the head of ρ, then $S \notin t_i$:

$$\bigwedge_{1 \leq i \leq n} R_i(\mathbf{x}_i) \wedge \bigwedge_{1 \leq i \leq k} \mathsf{con}_{t_i}(x_i) \wedge \bigwedge_{1 \leq i \leq \ell} \mathsf{con}_{t'_i}(z_i);$$

2. Consider each goal rule

$$\rho = \bigwedge_{1 \leq i \leq n} R_i(\mathbf{x}_i) \wedge \bigwedge_{1 \leq i \leq m} S_i(y_i) \rightarrow \mathsf{goal}()$$

   where all R_i are from $\mathbf{S}$ and all S_i are from $\mathbf{Q}$. Let $x_1, \dots, x_k$ be the variables in ρ. Then include in q_Π the following CQ, for all sequences of $\mathbf{Q}$-types $t_1, \dots, t_k$ such that if $S(x_i)$ occurs in the body of ρ with $S \in \mathbf{Q}$, then $S \in t_i$:

$$\bigwedge_{1 \leq i \leq n} R_i(\mathbf{x}_i) \wedge \bigwedge_{1 \leq i \leq k} \mathsf{con}_{t_i}(x_i).$$

The above construction parallels the one used in the proof of Theorem 3, with $\mathbf{Q}$-types playing the role of elements of the CSP template. In the reduction from Π to CQA(C_Π, q_Π), we are given an $\mathbf{S}$-instance I and construct an $\mathbf{S}'$-instance $T^\uparrow(I)$ by adding $Q_t(a)$ for all $a \in \mathsf{adom}(I)$ and all $\mathbf{Q}$-types t. Then each minimal repair J of $T^\uparrow(I)$ must satisfy, for each $a \in \mathsf{adom}(I)$, $J \models \mathsf{con}_t[a]$ for a unique $\mathbf{Q}$-type t and thus assigns to each $a \in \mathsf{adom}(I)$ a $\mathbf{Q}$-type t_a. In this way, it describes a unique $\mathbf{S} \cup \mathbf{Q}$-instance I' that is obtained from I by adding $P(a)$ whenever $P \in t_a$. If $J \not\models q_\Pi$, then I' satisfies all non-goal rules of Π, but no goal rule applies. Indeed, we show in the appendix that $I \models \Pi$ iff $T^\uparrow(I) \models_{C_\Pi} q_\Pi$. We also give an FO-reduction from CQA(C_Π, q_Π) to Π, which is again similar to what is done in the proof of Theorem 3.

▶ **Theorem 15.** *For every MDDLog program Π that admits precoloring, there is a problem CQA(C, q) from (MDiC,UCQ) that satisfies CQA$(C, q) \approx_{\mathsf{FO}} \Pi$ and can be constructed in polynomial time.*

We now show that (MGAV, UCQ) $\preceq_{\mathsf{FO}}$ coMMSNP. Let CQA(C, q) be from (MGAV,UCQ) with schema $\mathbf{S}$. We define an MDDLog program $\Pi_{C,q}$ over EDB schema $\mathbf{S}' = \mathbf{S}^{(>1)} \cup \{P_\Gamma \mid \Gamma \subseteq \mathbf{S}^{(1)}\}$ where the P_Γ are fresh monadic relation symbols. Recall from Section 3 that for each $\Gamma \subseteq \mathbf{S}^{(1)}$, $\mathsf{rep}(\Gamma)$ denotes the corresponding set of repairs. The IDB relations of $\Pi_{C,q}$

are $\mathbf{Q} = \mathbf{S}^{(1)} \cup \{Q_\Gamma \mid \Gamma \subseteq \mathbf{S}^{(1)}\}$ where the Q_Γ are monadic. The rules of $\Pi_{C,q}$ are as follows:

$$\begin{array}{rcll} P_\Gamma(x) & \rightarrow & \bigvee_{\Lambda \in \mathsf{rep}(\Gamma)} Q_\Lambda(x) & \text{for each } \Gamma \subseteq \mathbf{S}^{(1)} \\ Q_\Gamma(x) & \rightarrow & P(x) & \text{for each } \Gamma \subseteq \mathbf{S}^{(1)},\ P \in \Gamma \\ P_\Gamma(x) \wedge P_\Lambda(x) & \rightarrow & \mathsf{goal}() & \text{for all distinct } \Gamma, \Lambda \subseteq \mathbf{S}^{(1)} \\ q' & \rightarrow & \mathsf{goal}() & \text{for each disjunct } q' \text{ of } q. \end{array}$$

▶ **Lemma 16.** *$CQA(C, q) \preceq_{\mathsf{FO}} \Pi_{C,q}$ and $\Pi_{C,q} \preceq_{\mathsf{FO}} CQA(C, q)$.*

The reductions used for establishing this lemma are *exactly* the FO-reductions from the proof of Lemma 4. In the appendix, we prove that these reductions are correct also in this case.

▶ **Theorem 17.** *For every CQA problem in (MGAV, UCQ), there is an FO-equivalent MDDLog program Π that can be constructed in single exponential time.*

We thus obtain the following equivalences.

▶ **Corollary 18.**

1. *$(MDiC, UCQ) \approx_{\mathsf{FO}} (MGAV, UCQ) \approx_{\mathsf{FO}}$ coMMSNP;*
2. *$(MDiC, UCQ) \approx_p (MDiC, tUCQ)$ and $(MGAV, UCQ) \approx_p (MGAV, tUCQ)$.*

Since the FO-reductions in the proof of Lemma 16 are identical to those in the proof of Lemma 4, all arguments used in Section 4 for relating FO- and Datalog-rewritability of CQA problems with hypertree UCQs to the FO- and Datalog-definability of CSPs can also be used to relate, in exactly the same way, CQA problems with unrestricted UCQs to MMSNP. Consequently, we obtain that FO-rewritability in (MDiC, UCQ) and (MGAV, UCQ) is decidable iff FO-definability in MMSNP is decidable and the former is at most exponentially harder than the latter.

7 Relating CQA and GMSNP

Monadic disjointness constraints are surely a rather restricted class of denial constraints. In this section, we analyze a slightly more powerful class obtained by dropping the monadicity of MDiCs while still retaining some 'uniformity' accross the atoms in the constraint. More precisely, a *disjointness constraint (DiC)* has the form $\forall \mathbf{x}\, \neg(R_1(\mathbf{x}) \wedge \cdots \wedge R_n(\mathbf{x}))$ where all relations R_i have the same arity and all atoms $R_i(\mathbf{x})$ use the same variables from $\mathbf{x}$ in the same order (multiple occurrences are allowed). We show that the connection between (MDiC, UCQ) and coMMSNP established in Section 6 can be lifted to (DiC, UCQ) and a generalization of coMMSNP called coGMSNP that corresponds to replacing MDDLog with frontier-guarded disjunctive Datalog [10]. It is straightforward to define a corresponding generalization of GAV constraints and to establish analogous results for them, but for simplicity we refrain from doing so.

Recall that we defined coMMSNP in terms of MDDLog. A *frontier-guarded disjunctive Datalog (GDDLog) rule* takes the form $R_1(\mathbf{y}_1) \wedge \cdots \wedge R_k(\mathbf{y}_k) \rightarrow S_1(\mathbf{z}_1) \vee \cdots \vee S_m(\mathbf{z}_m)$ where the IDB predicates need not be monadic and for each head atom $S_i(\mathbf{z}_i)$, there must be a body atom $R_j(\mathbf{y}_j)$ with $\mathbf{z}_i \subseteq \mathbf{y}_j$. A GDDLog program is then defined in the obvious way (with nullary goal predicate) and we use coGMSNP to denote the class of decision problems that are defined by a GDDLog program. GMSNP, which is also known as MMSNP_2 [33], is considered an interesting candidate for a generalization of MMSNP that is still PTIME-equivalent to CSP. It is, however, an open problem whether this is really the case. We nevertheless believe that relating (DiC, UCQ) and coGMSNP provides interesting additional insight into the computational complexity of CQA. Note that coGMSNP is also closely related to ontology-based data access with the guarded fragment of FO [10].

What we actually show in this section is $\text{coGMSNP}^{\mathsf{pre}} \preceq_{\mathsf{FO}} (\text{DiC}, \text{UCQ}) \preceq_{\mathsf{FO}} \text{coGMSNP}$ where $\text{GMSNP}^{\mathsf{pre}}$ is a version of GMSNP that admits precoloring defined as follows in analogy with $\text{MMSNP}^{\mathsf{pre}}$. Let Π be a GDDLog program over schema $\mathbf{S}$, assume that $\mathbf{Q}$ is the set of IDB relations in Π, and let m be the maximal arity of relations in $\mathbf{Q}$. Fix variables $x_1, \ldots, x_m$. For $i \leq m$, an *i-type* is a set t of relational atoms using relation symbols from $\mathbf{Q}$ and variables from $\mathbf{x}_i := x_1 \cdots x_i$. We say that Π *admits precoloring* if the EDB schema $\mathbf{S}$ includes a relation symbol S_t of arity i for each i-type t, $i \leq m$, and Π includes rules (i) $S_t(\mathbf{x}_i) \to R(x_{i_1}, \ldots, x_{i_\ell})$ for all i-types t and all $R(x_{i_1}, \ldots, x_{i_\ell}) \in t$ and (ii) $S_t(\mathbf{x}_i) \wedge R(x_{i_1}, \ldots, x_{i_\ell}) \to \bot$ for all i-types t and all $R(x_{i_1}, \ldots, x_{i_\ell}) \notin t$ with $R \in \mathbf{Q}$ and $x_{i_1}, \ldots, x_{i_\ell} \in \mathbf{x}_i$; the S_t relations are not allowed to be used in any other rule. We leave it open whether $\text{coGMSNP} \preceq_p \text{coGMSNP}^{\mathsf{pre}}$, but consider it likely that this is the case given the corresponding result for MMSNP mentioned in Section 6.

We start with proving $\text{coGMSNP}^{\mathsf{pre}} \preceq_{\mathsf{FO}} (\text{DiC}, \text{UCQ})$, first observing that it suffices to consider GDDLog programs of a certain form, which we introduce next. Let $\mathbf{Q}$ be a schema that consists of relations which all have the same arity m. A *$\mathbf{Q}$-type* is a set of relational atoms $R(\mathbf{x})$ with $R \in \mathbf{Q}$ and where $\mathbf{x}$ is a permutation of $x_1 \cdots x_m$. We say that a GDDlog program Π with IDB relations $\mathbf{Q}$ is *normalized* if it satisfies the following conditions:

1. all EDB and IDB relations (except the goal relation) have the same arity m, each variable may occur at most once in each head and body atom, and Π includes all rules of the form $R(\ldots, x, \ldots, x, \ldots) \to \bot$ for each EDB and IDB relation R;
2. the EDB schema $\mathbf{S}$ includes a relation symbol S_t of arity m for each $\mathbf{Q}$-type t and Π includes rules (i) $S_t(x_1, \ldots, x_m) \to R(x_{i_1}, \ldots, x_{i_m})$ for all $\mathbf{Q}$-types t and all $R(x_{i_1}, \ldots, x_{i_m}) \in t$ and (ii) $S_t(x_1, \ldots, x_m) \wedge R(x_{i_1}, \ldots, x_{i_m}) \to \bot$ for all $\mathbf{Q}$-types t and atoms $R(x_{i_1}, \ldots, x_{i_m}) \notin t$, where $R \in \mathbf{Q}$ and $x_{i_1} \ldots x_{i_m}$ is a permutation of $x_1 \ldots x_m$; the S_t relations are not allowed to be used in any other rule.[2]

Working with normalized programs will simplify the subsequent constructions.

▶ **Lemma 19.** *For every GDDLog program Π that admits precoloring, there is a normalized GDDLog-Program Π' with $\Pi \approx_{\mathsf{FO}} \Pi'$.*

Let Π be a normalized GDDLog program over schema $\mathbf{S}$, let $\mathbf{Q}$ be the set of IDB relations in Π and m the unique arity of relations in Π. We define a CQA setup $\text{CQA}(C_\Pi, q_\Pi)$ from (DiC, UCQ) over schema $\mathbf{S}'$, that is, $\mathbf{S}$ extended with one relation Q_t of arity m for each $\mathbf{Q}$-type t. For an $\mathbf{S}'$-instance J, and a tuple $\mathbf{a} \in \mathsf{adom}(J)^m$, we say that $\mathbf{a}$ is *assigned $\mathbf{Q}$-type t in J* if there is a permutation $\mathbf{b}$ of $\mathbf{a}$ and a $\mathbf{Q}$-type $\widehat{t}$ such that (i) $Q_{t'}(\mathbf{b}) \in J$ if $t' \neq \widehat{t}$ for all $\mathbf{Q}$-types t' and (ii) $\widehat{t}$ is obtained from t by permuting the variables $x_1 \cdots x_m$ in the same way in which the constants in $\mathbf{a}$ are permuted in $\mathbf{b}$, that is, if $\mathbf{a} = a_1 \cdots a_m$ and $\mathbf{b} = a_{i_1} \cdots a_{i_m}$, then $\widehat{t} = t[x_{i_1} \cdots x_{i_m} / x_1 \cdots x_m]$. We say that J is *proper* if every $\mathbf{S}$-guarded tuple[3] $\mathbf{a} \in \mathsf{adom}(J)^m$ is assigned a unique $\mathbf{Q}$-type. A proper $\mathbf{S}'$-instance J represents an $\mathbf{S}$-instance and an $\mathbf{S} \cup \mathbf{Q}$-instance: the former is the reduct of J to schema $\mathbf{S}$ and the latter is obtained by starting with that reduct and then adding the fact $R(a_{i_1}, \ldots, a_{i_m})$ whenever

[2] Note that this is different from requiring Π to admit precoloring because admitting precoloring is about i-types whereas for normalized programs we use $\mathbf{Q}$-types. In fact, a program in normal form cannot admit precoloring in the original sense because the conditions on the rules required for normality and admitting precoloring are incompatible.

[3] A tuple $\mathbf{a} \in \mathsf{adom}(I)^m$ is $\mathbf{S}$-*guarded* if I contains some fact $R(\mathbf{a})$ with $R \in \mathbf{S}$.

$\mathbf{a} = a_1 \cdots a_m \in \mathsf{adom}(J)^m$ is assigned $\mathbf{Q}$-type t and $R(x_{i_1}, \dots, x_{i_m}) \in t$. Intuitively, the latter instance is supposed to represent an extension of the former instance obtained by applying the rules in Π.

We now define CQA(C_Π, q_Π). The set C_Π contains one disjointness constraint, namely

$$\forall x_1 \cdots \forall x_m \neg (\bigwedge_{t \text{ a } \mathbf{Q}\text{-type}} Q_t(x_1, \dots, x_m)).$$

For a $\mathbf{Q}$-type t, we use $C_t(\mathbf{x})$ to denote the conjunction $\bigwedge_{t' \text{ a } \mathbf{Q}\text{-type}, t' \neq t} Q_{t'}(\mathbf{x})$. A CQ q over schema $\mathbf{S}'$ is *forbidden* if for each proper $\mathbf{S}'$-instance J such that the $\mathbf{S} \cup \mathbf{Q}$-instance I of J satisfies all non-goal rules in Π and no goal-rule of Π applies in I, we have $J \not\models q$. The UCQ q_Π consists of the following CQs:

1. all forbidden CQs q over schema $\mathbf{S}'$ such that
   a. the number of variables in q is bounded by the maximum number of variables in a rule body in Π;
   b. for every atom $R(\mathbf{x})$ there is a $\mathbf{Q}$-type t and a permutation $\mathbf{y}$ of $\mathbf{x}$ such that $C_t(\mathbf{y})$ is a subconjunction of q;
2. all CQs $C_t(\mathbf{x}) \wedge C_{t'}(\mathbf{y})$ such that $\mathbf{y}$ is a permutation of $\mathbf{x}$ and $t' \neq t[\mathbf{y}/\mathbf{x}]$.

▶ **Lemma 20.** $\Pi \preceq_{\mathsf{FO}}$ *CQA*(C_Π, q_Π) *and CQA*$(C_\Pi, q_\Pi) \preceq_{\mathsf{FO}} \Pi$.

Proof. For the first reduction, assume that an $\mathbf{S}$-instance I is given. Define an $\mathbf{S}'$-instance $T^\uparrow(I)$ as the extension of I with all facts $Q_t(\mathbf{a})$ such that t is a $\mathbf{Q}$-type and $\mathbf{a} \in \mathsf{adom}(I)^m$. We show in the appendix that $I \models \Pi$ iff $T^\uparrow(I) \models_{C_\Pi} q_\Pi$. Clearly, $T^\uparrow$ is FO-definable.

For the second reduction, let I be an $\mathbf{S}'$-instance. Let $T^\downarrow(I)$ be the $\mathbf{S}$-instance obtained from the reduct of I to schema $\mathbf{S}$ by the following sequence of operations:
1. drop each fact $R(\mathbf{a})$ such that for every permutation $\mathbf{b}$ of $\mathbf{a}$, there are distinct $\mathbf{Q}$-types t and t' such that neither $Q_t(\mathbf{b})$ nor $Q_{t'}(\mathbf{b})$ are in I;
2. add $S_t(\mathbf{a})$ for each $\mathbf{S}$-guarded tuple $\mathbf{a} \in \mathsf{adom}(I)^m$ that is assigned $\mathbf{Q}$-type t in I (the assignment need not be unique).

We show in the appendix that $I \models_{C_\Pi} q_\Pi$ iff $T^\downarrow(I) \models \Pi$. Again $T^\downarrow$ is clearly FO-definable. ❏

▶ **Theorem 21.** *For every GDDLog program* Π *that admits precoloring, there is an FO-equivalent problem CQA*(C, q) *from (DiC,UCQ).*

We now turn towards showing that (DiC, UCQ) $\preceq_{\mathsf{FO}}$ coGMSNP. Let CQA(C, q) over schema $\mathbf{S}$ be given with C a set of disjointness constraints. Let m be the maximum arity of a relation in $\mathbf{S}$. Fix variables $x_1, \dots, x_m$. For $i \leq m$, an *i-type* is a set of atoms $R(x_1, \dots, x_i)$ with $R \in \mathbf{S}$ (all variables occur in fixed order and are distinct). We use tp_i to denote the set of all i-types. For an $\mathbf{S}$-instance I and $\mathbf{a} \in \mathsf{adom}(I)^i$, we use $\mathsf{tp}_I(\mathbf{a})$ denote the i-type realized at $\mathbf{a}$ in I, that is, the set of all atoms $R(x_1, \dots, x_i)$ such that $R(\mathbf{a}) \in I$.

We define a GDDLog program $\Pi_{C,q}$ over schema $\mathbf{S}' = \{R_t \mid t \in \mathsf{tp}_i, i \leq m\}$. The quantified relations in $\Pi_{C,q}$ are $\mathbf{S} \cup \{Q_t \mid t \in \mathsf{tp}_i, i \leq m\}$ where the arities are defined in the obvious way. The disjointness constraints in C assign to each i-type t and each sequence of variables $\mathbf{x} \in \{x_1, \dots, x_m\}^i$ (repetitions allowed) a set of possible minimal repairs (which are also i-types), denoted with $\mathsf{rep}_{\mathbf{x}}(t)$. Formally, $\mathsf{rep}_{\mathbf{x}}(t)$ are the minimal repairs of $\{R(\mathbf{x}) \mid R(x_1, \dots, x_i) \in t\}$ viewed as an instance. Note that different sequences $\mathbf{x}$ may give rise to different repair sets for the same i-type t since the constraints in C might use variables multiple times in the same atom. The rules of $\Pi_{C,q}$ are as follows:

1. for each i-type t, $i \leq m$, and each $\mathbf{x} \in \{x_1, \ldots, x_m\}^i$: $R_t(\mathbf{x}) \rightarrow \bigvee_{t' \in \mathsf{rep}_{\mathbf{x}}(t)} Q_{t'}(\mathbf{x})$
2. for each i-type t, $i \leq m$, and each $R(x_1, \ldots, x_i) \in t$: $Q_t(x_1, \ldots, x_i) \rightarrow R(x_1, \ldots, x_i)$
3. for all distinct i-types t, t', $i \leq m$: $R_t(x_1, \ldots, x_i) \wedge R_{t'}(x_1, \ldots, x_i) \rightarrow \mathsf{goal}()$
4. for each CQ q' in q: $q' \rightarrow \mathsf{goal}()$

▶ **Lemma 22.** *$CQA(C, q) \preceq_{\mathsf{FO}} \Pi_{C,q}$ and $\Pi_{C,q} \preceq_{\mathsf{FO}} CQA(C, q)$.*

Proof. For the first reduction, let an $\mathbf{S}$-instance I be given. Define an $\mathbf{S}'$-instance $T^{\uparrow}(I)$ that consists of all facts $R_{\mathsf{tp}_I(\mathbf{a})}(\mathbf{a})$ with $\mathbf{a} \in \mathsf{adom}(I)^i$, $i \leq m$. Clearly, $T^{\uparrow}$ is FO-definable. We show in the appendix that $I \models_C q$ iff $T^{\uparrow}(I) \models \Pi_{C,q}$.

For the second reduction, let an $\mathbf{S}'$-instance I be given. If $R_t(\mathbf{a}), R_{t'}(\mathbf{a}) \in I$ for any $\mathbf{a} \in \mathsf{dom}(I)^i$, $i \leq m$, and $t \neq t'$, then answer 'inconsistent'. Otherwise, replace each fact $R_t(\mathbf{a})$ with the set of facts $\{R(\mathbf{a}) \mid R(\mathbf{x}) \in t\}$, and call the result $T^{\downarrow}(I)$. Clearly, $T^{\downarrow}$ is FO-definable. We show in the appendix that $I \models \Pi_{C,q}$ iff $T^{\downarrow}(I) \models_C q$. ❏

▶ **Theorem 23.** *For every CQA problem in (DiC,UCQ), there is an FO-equivalent GDDLog program Π.*

8 Conclusion

We find it intriguing that very simple integrity constraints such as MDiCs and MGAVs combined with structurally simple queries such as tUCQs result in classes of CQA problems that are as difficult to analyze as CSPs, and that slight generalizations (such as DiCs) result in entering essentially unknown terrain (in the form of GMSNP). We believe that the CSP-CQA connection is not yet explored to the end. For example, the known non-dichotomy between PTime and NP for MMSNP extended with inequality [21] gives rise to the speculation that $(\text{MDiC}, \text{UCQ}^{\neq})$ might have no dichotomy between PTime and coNP either, where $\text{UCQ}^{\neq}$ denotes the class of UCQs that also admit inequality atoms. The proof of such a result will not be entirely simple, though, as it seems to require a version of Ladner's theorem that relates to Ladner's original theorem in a similar way in which the mortality problem of Turing machines relates to the halting problem. We leave this as future work.

Another interesting question is whether larger and more natural classes of CQA problems, such as those based on unrestricted denial constraints or on some form of functional dependency also have natural counterparts in the world of CSP and MMSNP. In fact, it is not even clear whether the correspondence between $(\text{MDiC}, \text{tUCQ})$ and coCSP can be extended from monadic disjointness constraints to monadic *denial* constraints in which relations are still required to be monadic, but where more than one variable can be used. For example, the constraint $\forall x \forall y \neg(A(x) \wedge B(y))$ is a monadic denial constraint, but not an MDiC. The main challenge is to deal with the problem that the 'yes'-instances of each CSP are closed under homomorphic pre-images while the 'no'-instances of CQA problems are not. Simply changing the monadic relations in the schema as in the proof of Theorem 5 is no longer sufficient because, in contrast to MDiCs, monadic denial constraints are not local to a single constant. We believe, however, that even if it should turn out that such differences prevent the CQA-CSP connection from gracefully extending beyond the classes of CQA problems considered here, it might still be possible to carry over techniques and intuitions from the CSP/MMSNP world, which has seen frantic development in the last decade.

Acknowledgements We want to thank Manuel Bodirsky, Hubie Chen, and André Hernich for interesting and helpful discussions.

References

1 Foto N. Afrati and Phokion G. Kolaitis. Repair checking in inconsistent databases: algorithms and complexity. In *Proc. of ICDT*, volume 361 of *ACM International Conference Proceeding Series*, pages 31–41. ACM, 2009.

2 Marcelo Arenas and Leopoldo E. Bertossi. On the decidability of consistent query answering. In *Proc. of AMW*, volume 619 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2010.

3 Marcelo Arenas, Leopoldo E. Bertossi, and Jan Chomicki. Consistent query answers in inconsistent databases. In *Proc. of PODS*, pages 68–79. ACM Press, 1999.

4 Albert Atserias. On digraph coloring problems and treewidth duality. In *Proc. of LICS*, pages 106–115, 2005.

5 Libor Barto and Marcin Kozik. Constraint satisfaction problems of bounded width. In *Proc. of FOCS*, pages 595–603, 2009.

6 Catriel Beeri, Ronald Fagin, David Maier, and Mihalis Yannakakis. On the desirability of acyclic database schemes. *J. ACM*, 30(3):479–513, 1983.

7 Leopoldo E. Bertossi. Consistent query answering in databases. *SIGMOD Record*, 35(2):68–76, 2006.

8 Leopoldo E. Bertossi, Loreto Bravo, Enrico Franconi, and Andrei Lopatenko. The complexity and approximation of fixing numerical attributes in databases under integrity constraints. *Inf. Syst.*, 33(4-5):407–434, 2008.

9 Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. First-order rewritability of atomic queries in horn description logics. In *Proc. of IJCAI*. IJCAI/AAAI, 2013.

10 Meghyn Bienvenu, Balder ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: a study through disjunctive datalog, CSP, and MMSNP. In *Proc. of PODS*, pages 213–224. ACM, 2013.

11 Manuel Bodirsky, Hubie Chen, and Tomás Feder. On the complexity of MMSNP. *SIAM J. Discrete Math.*, 26(1):404–414, 2012.

12 Manuel Bodirsky, Jan Kára, and Barnaby Martin. The complexity of surjective homomorphism problems - a survey. *Discrete Applied Mathematics*, 160(12):1680–1690, 2012.

13 Manuel Bodirsky and Florent Madeleine. Feder and Vardi's logic revisited. In preparation.

14 Andrei A. Bulatov. A dichotomy theorem for constraint satisfaction problems on a 3-element set. *J. ACM*, 53(1):66–120, 2006.

15 Andrea Calì, Domenico Lembo, and Riccardo Rosati. On the decidability and complexity of query answering over inconsistent and incomplete databases. In *Proc. of PODS*, pages 260–271. ACM, 2003.

16 Jan Chomicki. Consistent query answering: Five easy pieces. In *Proc. of ICDT*, volume 4353 of *LNCS*, pages 1–17. Springer, 2007.

17 Jan Chomicki and Jerzy Marcinkowski. Minimal-change integrity maintenance using tuple deletions. *Inf. Comput.*, 197(1-2):90–121, 2005.

18 Jan Chomicki, Jerzy Marcinkowski, and Slawomir Staworko. Computing consistent query answers using conflict hypergraphs. In *Proc. of CIKM*, pages 417–426. ACM, 2004.

19 David Cohen and Peter Jeavons. *The complexity of constraint languages*, chapter 8. Elsevier, 2006.

20 Stavros S. Cosmadakis, Haim Gaifman, Paris C. Kanellakis, and Moshe Y. Vardi. Decidable optimization problems for database logic programs (preliminary report). In *Proc. of STOC*, pages 477–490. ACM, 1988.

21 Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through datalog and group theory. *SIAM J. Comput.*, 28(1):57–104, 1998.

22 Tomás Feder and Moshe Y. Vardi. Homomorphism closed vs. existential positive. In *18th IEEE Symposium on Logic in Computer Science (LICS 2003), 22-25 June 2003, Ottawa, Canada, Proceedings*, pages 311–320, 2003.

23 Gaëlle Fontaine. Why is it hard to obtain a dichotomy for consistent query answering? In *Proc. of LICS*, pages 550–559. IEEE Computer Society, 2013.

24 Ralph Freese, Marcin Kozik, Andrei Krokhin, Miklós Maróti, Ralph KcKenzie, and Ross Willard. On Maltsev conditions associated with omitting certain types of local structures. In preparation. Manuscript available from `http://www.math.hawaii.edu/~ralph/Classes/619/OmittingTypesMaltsev.pdf`.

25 Ariel Fuxman and Renée J. Miller. First-order query rewriting for inconsistent databases. In *Proc. of ICDT*, volume 3363 of *LNCS*, pages 337–351. Springer, 2005.

26 Georg Gottlob, Nicola Leone, and Francesco Scarcello. Hypertree decompositions and tractable queries. *J. Comput. Syst. Sci.*, 64(3):579–627, 2002.

27 Neil Immerman. *Descriptive complexity.* Springer, 1999.

28 Phokion G. Kolaitis and Enela Pema. A dichotomy in the complexity of consistent query answering for queries with two atoms. *Inf. Process. Lett.*, 112(3):77–85, 2012.

29 Paraschos Koutris and Dan Suciu. A dichotomy on the complexity of consistent query answering for atoms with simple keys. In *Proc. of ICDT*, pages 165–176. OpenProceedings.org, 2014.

30 Gábor Kun. Constraints, MMSNP and expander relational structures. *Combinatorica*, 33(3):335–347, 2013.

31 Benoit Larose, Cynthia Loten, and Claude Tardif. A characterisation of first-order constraint satisfaction problems. *Logical Methods in Computer Science*, 3(4), 2007.

32 Benoit Larose and Pascal Tesson. Universal algebra and hardness results for constraint satisfaction problems. *Theor. Comput. Sci.*, 410(18):1629–1647, 2009.

33 Florent R. Madelaine. Universal structures and the logic of forbidden patterns. *Logical Methods in Computer Science*, 5(2), 2009.

34 Florent R. Madelaine and Iain A. Stewart. Constraint satisfaction, logic and forbidden patterns. *SIAM J. Comput.*, 37(1):132–163, 2007.

35 Jaroslav Nesetril. Many facets of dualities. In *Bonn Workshop of Combinatorial Optimization*, pages 285–302, 2008.

36 Benjamin Rossman. Homomorphism preservation theorems. *J. ACM*, 55(3), 2008.

37 Thomas J. Schaefer. The complexity of satisfiability problems. In *Proc. of STOC*, pages 216–226. ACM, 1978.

38 Slawomir Staworko and Jan Chomicki. Consistent query answers in the presence of universal constraints. *Inf. Syst.*, 35(1):1–22, 2010.

39 Balder ten Cate, Gaëlle Fontaine, and Phokion G. Kolaitis. On the data complexity of consistent query answering. In *Proc. if ICDT*, pages 22–33. ACM, 2012.

40 Jef Wijsen. On the first-order expressibility of computing certain answers to conjunctive queries over uncertain databases. In *Proc. of PODS*, pages 179–190. ACM, 2010.

41 Jef Wijsen. Charting the tractability frontier of certain conjunctive query answering. In *Proc. of PODS*, pages 189–200. ACM, 2013.

42 Jef Wijsen. A survey of the data complexity of consistent query answering under key constraints. In *Proc. of FoIKS*, volume 8367 of *LNCS*, pages 62–78. Springer, 2014.

A Proofs for Section 3

▶ **Lemma 24.** $coCSP(A) \preceq_{\mathsf{FO}} CQA(C_A, q_A)$.

Proof. Let an $\mathbf{S}$-instance I be given. Convert I into an $\mathbf{S}'$-instance $T^{\uparrow}(I)$ by extending I with $Q_a(b)$ for all $b \in \mathsf{adom}(I)$ and $a \in \mathsf{adom}(A)$. Clearly, $T^{\uparrow}$ is FO-definable. It remains to show that $T^{\uparrow}(I) \not\models_{C_A} q_A$ iff $I \to A$.

"if". Let h be a homomorphism from I to A. Define J as $T^{\uparrow}(I)$ with $Q_a(b)$ removed whenever $h(b) = a$, for all $a \in \mathsf{adom}(A)$ and $b \in \mathsf{adom}(I)$. It is easy to verify that J is a minimal repair of $T^{\uparrow}(I)$. We show that $J \not\models q_A$. For a proof by contradiction assume that there are $R \in \mathbf{S}$ of arity n and $a_1 \cdots a_n \in \mathsf{adom}(A)$ with $R(a_1, \dots, a_n) \notin A$ such that

$$J \models \mathsf{con}_{a_1}(b_1) \wedge \cdots \wedge \mathsf{con}_{a_n}(b_n) \wedge R(b_1, \dots, b_n)$$

for some $b_1, \dots, b_n$. By construction of J, this implies $h(b_1) = a_1, \dots, h(b_n) = a_n$, and thus we have derived a contradiction to the assumption that h is a homomorphism.

"only if". $T^{\uparrow}(I) \not\models_{C_A} q_A$ implies $I \to A$. Let J be a minimal repair of $T^{\uparrow}(I)$ such that $J \not\models q_A$. Then for each $b \in \mathsf{adom}(J)$, there exists a unique $a_b \in \mathsf{adom}(A)$ such that $Q_{a_b}(b) \in J$ iff $e \neq a_b$ for all $e \in \mathsf{adom}(A)$. Define a map h from I to A by setting $h(b) = a_b$. Assume to the contrary of what is to be shown that h is not a homomorphism from I to A. Then there is a fact $R(b_1, \dots, b_n) \in I$ such that $R(h(b_1), \dots, h(b_n)) \notin A$. But then

$$J \models \mathsf{con}_{a_{b_1}}(b_1) \wedge \cdots \wedge \mathsf{con}_{a_{b_n}}(b_n) \wedge R(b_1, \dots, b_n)$$

and we have derived a contradiction to the assumption that $J \not\models q_A$. ❏

▶ **Lemma 25.** $CQA(C_A, q_A) \preceq_{\mathsf{FO}} coCSP(A)$.

Proof. Let I be an $\mathbf{S}'$-instance. Denote by X the set of $b \in \mathsf{adom}(I)$ such that there are at least two distinct $a_1, a_2 \in \mathsf{adom}(A)$ with neither $Q_{a_1}(b) \in I$ nor $Q_{a_2}(b) \in I$. Then $T^{\downarrow}(I)$ is obtained from I by dropping all facts that involve a constant from X or a relation symbol in $\mathbf{S}' \setminus \mathbf{S}$, and adding all facts $P_a(b)$ such that $Q_e(b) \in I$ iff $e \neq a$ for all $e \in \mathsf{adom}(A)$. Note that we rely on A to admit precoloring by using the relations P_a. Clearly $T^{\downarrow}(I)$ is FO-definable. It remains to show that $I \not\models_{C_A} q_A$ iff $T^{\downarrow}(I) \to A$.

"if". Let h be a homomorphism from $T^{\downarrow}(I)$ to A. Let J be obtained from I by removing every fact $Q_a(b)$ with $h(b) = a$. J is a minimal repair of I. To prove this, it suffices to show that whenever $Q_a(b)$ is removed, then $Q_e(b) \in I$ for all $e \in A$. Assume to the contrary that this is not the case. Since $h(b)$ is defined, we must have $b \notin X$, and consequently there is a unique $a' \in \mathsf{adom}(A)$ with $Q_{a'}(b) \notin I$. By construction of $T^{\downarrow}(I)$, we have $P_{a'}(b) \in T^{\downarrow}(I)$. Since h is a homomorphism, we must thus have $h(b) = a'$, consequently $a' = a$ and thus $Q_{a'}(b) \notin I$, a contradiction to it being removed during the construction of J. We now show that $J \not\models q_A$. Let $R \in \mathbf{S}$ be of arity n and $R(a_1, \dots, a_n) \notin A$. Assume to the contrary of what we have to show that

$$J \models \mathsf{con}_{a_1}(b_1) \wedge \cdots \wedge \mathsf{con}_{a_n}(b_n) \wedge R(b_1, \dots, b_n)$$

for some $b_1, \dots, b_n$. From $J \models \mathsf{con}_{a_i}(b_i)$ we get $h(b_i) = a_i$ by construction of J, for $1 \leq i \leq n$. Moreover, $J \models \mathsf{con}_{a_1}(b_1) \wedge \cdots \wedge \mathsf{con}_{a_n}(b_n)$ implies that $b_1, \dots, b_n$ are in X. Consequently, $J \models R(b_1, \dots, b_n)$ yields $R(b_1, \dots, b_n) \in T^{\downarrow}(I)$ in contradiction to h being a homomorphism.

"only if". $I \not\models_{C_A} q_A$ implies $T^{\downarrow}(I) \to A$. Let J be a minimal repair of I such that $J \not\models q_A$. Since J is a minimal repair and by definition of X, for every $b \in \mathsf{adom}(I) \setminus X$ there must be a unique $a_b \in A$ such that $Q_e(a) \in J$ iff $e \neq a_b$, for all $e \in \mathsf{adom}(A)$. Define a map h from $T^{\downarrow}(I)$ to A by setting $h(b) = a_b$ for all $b \in \mathsf{adom}(I) \setminus X$. We show that h is a homomorphism from $T^{\downarrow}(I)$ to A. It is easy to check that h preserves the facts $P_a(b)$ added in the construction of $T^{\downarrow}(I)$. All other facts $R(b_1, \dots, b_n) \in T^{\downarrow}(I)$ are also in J and satisfy $R \in \mathbf{S}$, thus we have $R(b_1, \dots, b_n) \in J$. We must have $R(h(b_1), \dots, h(b_n)) \in A$ because by definition of h we otherwise obtain $J \models q_A$ in contradiction to this not being the case. ❏

To prove Lemma 4, we first establish a technical lemma. Let I_A be the $\mathbf{S}$-instance obtained from A_{C,q_0} by dropping all facts $P_\Gamma(\langle t, \Gamma\rangle)$ and adding $P(\langle t, \Gamma\rangle)$ for all $(P(x), x) \in t$.

▶ **Lemma 26.**

1. *I_A satisfies all constraints in C and $I_A \not\models q_0$.*
2. *Let I be an $\mathbf{S}$-instance $R(a_1, \dots, a_n) \in I$, $n > 1$. Then $(t_1, \dots, t_n)$ is R-coherent, where $t_i = \mathsf{tp}_I(a_i)$.*

Proof. We start with the proof of Point 1. It is clear by construction of A_{C,q_0} and I_A that the latter satisfies all constraints in C. It thus remains to show that $I_A \not\models q_0$. We show that for all places (q, x) and elements $\langle t, \Gamma\rangle$ in I_A, $(q, x) \notin t$ implies that $(q, x) \notin \mathsf{tp}_{I_A}(\langle t, \Gamma\rangle)$. Since the type t in any element $\langle t, \Gamma\rangle$ of A_{C,q_0} avoids q_0, this implies $I_A \not\models q_0$ as desired. The proof is by induction on the number of atoms in q. The induction start and step are identical and treated together.

Take a place (q, x) and an element $\langle t, \Gamma\rangle$ with $(q, x) \notin t$. Assume to the contrary of what is to be shown that $(q, x) \in \mathsf{tp}_{I_A}(\langle t, \Gamma\rangle)$, that is, there is a homomorphism h from q to I_A such that $h(x) = \langle t, \Gamma\rangle$. Take an atom $R(x_1, \dots, x_n) \in q$ with $x = x_\ell$ and let $h(x_i) = \langle t_i, \Gamma_i\rangle$ for $1 \leq i \leq n$. To obtain a contradiction, it suffices to show that $(t_1, \dots, t_n)$ is not R-coherent because then $R(h(x_1), \dots, h(x_n)) \notin I_A$ despite h being a homomorphism. To show that $(t_1, \dots, t_n)$ is not R-coherent, take disjoint instances $I_1, \dots, I_n$ with distinguished elements $a_1, \dots, a_n$ such that $\mathsf{tp}_{I_i}(a_i) = t_i$ for $1 \leq i \leq n$, which exist since the t_i are all realizable by construction of A_{C,q_0}. Let U be the disjoint union of $I_1, \dots, I_n$. Since all queries used in places are connected, we have $\mathsf{tp}_U(a_i) = t_i$ for $1 \leq i \leq n$. In particular, $(q, x) \notin \mathsf{tp}_U(a_\ell)$. Let $q_1, \dots, q_k$ be the queries obtained from q as maximally connected components when removing the atom $R(x_1, \dots, x_n)$—in the induction start, this sequence of queries is empty. Since q is a hypertree, we can associate with each q_i a unique j_i, $1 \leq j_i \leq n$, such that the variable x_{j_i} occurs in q_i. Because of the homomorphism h, we have $(q_i, x_{j_i}) \in \mathsf{tp}_{I_A}(h(x_{j_i}))$ for $1 \leq i \leq k$. By IH, this yields $(q_i, x_{j_i}) \in t_{j_i}$, consequently $(q_i, x_{j_i}) \in \mathsf{tp}_U(a_{j_i})$. Now let U' be obtained from U by adding the fact $R(a_1, \dots, a_n)$. It is easy to assemble the homomorphisms witnessing $(q_i, x_{j_i}) \in \mathsf{tp}_U(a_{j_i})$ into a homomorphism h' from q to U' with $h'(x_\ell) = a_\ell$. Consequently, $(q, x) \in \mathsf{tp}_{U'}(a_\ell)$. Thus, the instances U and U' witness that $(t_1, \dots, t_n)$ is not R-coherent.

We now prove Point 2. Take an instance J and a tuple of constants $(b_1, \dots, b_n)$ such that $\mathsf{tp}_J(b_i) = t_i$ for $1 \leq i \leq n$. We have to show that, after extending J to a new instance J' by adding the tuple $R(b_1, \dots, b_n)$, we have $\mathsf{tp}_{J'}(b_i) = t_i$ for $1 \leq i \leq n$. This amounts to showing that for all places (q, x) and all elements a of J, $(q, x) \in \mathsf{tp}_{J'}(a)$ implies $(q, x) \in \mathsf{tp}_J(a)$. We do this by induction on the number of atoms in q. The induction start and step are identical and treated together.

Take a place (q, x) and a constant a from J such that $(q, x) \in \mathsf{tp}_{J'}(a)$. Then there is a homomorphism h from q to J' that takes x to a. Choose an atom $S(x_1, \dots, x_m) \in q$ with $x = x_\ell$ and, therefore, $h(x_\ell) = a$ or some ℓ. Let $q_1, \dots, q_k$ be the (hypertree) CQs obtained as the maximal connected components when removing from q the atom $S(x_1, \dots, x_m)$—in the induction start, this sequence of queries is empty. Since q is a hypertree, for every q_i

there is a unique variable x_{j_i} from $x_1, \dots, x_m$ that occurs in q_i. IH yields that for $1 \leq i \leq k$, we have $(q_i, x_{j_i}) \in \mathsf{tp}_J(h(x_{j_i}))$. We distinguish two cases.

First assume that h does not map $S(x_1, \dots, x_m)$ to $R(b_1, \dots, b_n)$. Then h must map $S(x_1, \dots, x_m)$ to a fact that was already in J. Combining the restriction of h to variables $x_1, \dots, x_m$ with homomorphisms witnessing $(q_i, x_{j_i}) \in \mathsf{tp}_J(h(x_{j_i}))$, it is now straightforward to assemble a homomorphism from q to J that maps x to a, witnessing $(q, x) \in \mathsf{tp}_J(a)$ as required.

Now assume that h maps $S(x_1, \dots, x_m)$ to $R(b_1, \dots, b_n)$. Then $S = R$, $m = n$ and $h(x_i) = b_i$ for $1 \leq i \leq n$. By the latter and since $\mathsf{tp}_J(b_{i_k}) = t_{j_i}$, $(q_i, x_{j_i}) \in \mathsf{tp}_J(h(x_{j_i}))$ yields $(q_i, x_{j_i}) \in t_{j_i}$ for $1 \leq i \leq k$. Thus there is a homomorphism from q_i to J that maps x_{j_i} to b_{j_i}, for $1 \leq i \leq k$. From $R(a_1, \dots, a_n) \in I$ and $t_i = \mathsf{tp}_I(a_i)$ we obtain $(R(x_1, \dots, x_n), x_i) \in t_i$. We can thus assemble a homomorphism from q to J that maps x_i to b_i for $1 \leq i \leq n$. Therefore, $(q, x_i) \in t_i$ for $1 \leq i \leq n$. Since $h(x_\ell) = a = b_\ell$, we obtain $(q, x_\ell) \in \mathsf{tp}_I(a)$ and are done. ❑

▶ **Lemma 27.** *$CQA(C, q_0) \preceq_{\mathsf{FO}} coCSP(A_{C,q_0})$.*

Proof. Let an $\mathbf{S}$-instance I be given. Define an $\mathbf{S}'$-instance $T^\uparrow(I)$ by dropping all facts that involve a monadic predicate and adding $P_{\Gamma_a}(a)$ for every element $a \in I$, where $\Gamma_a = \{P \mid P(a) \in I\}$. Clearly, $T^\uparrow$ is FO-definable. Then we have $I \not\models_C q_0$ iff $T^\uparrow(I) \to A_{C,q_0}$.

"if". Assume that h is a homomorphism from $T^\uparrow(I)$ to A_{C,q_0}. Define an $\mathbf{S}$-instance I' by starting with I and dropping each fact $P(a)$ such that $h(a) = \langle t, \Gamma \rangle$ and $(P(x), x) \notin t$. It suffices to show that I' is a minimal repair of I with $I' \not\models q_0$. The former is obvious by definition of $T^\uparrow(I)$, A_{C,q_0}, and I'. The latter follows from Point 1 of Lemma 26 since any homomorphism from a CQ in q_0 to I' gives rise to a homomorphism from that CQ to I_A via composition with h.

"only if". Assume that $I \not\models_C q_0$ and let I' be a minimal repair of I with $I' \not\models q_0$. Define a map h from $T^\uparrow(I)$ to A_{C,q_0} by setting

$$h(a) = \langle \mathsf{tp}_{I'}(a), \{P \in \mathbf{S}^{(1)} \mid P(a) \in I\} \rangle$$

for all elements a of $T^\uparrow(I)$. Note that, by definition of $T^\uparrow(I)$ and choice of I', $h(a)$ is indeed a constant of A_{C,q_0} for each a. It remains to show that h is a homomorphism. First let $P_\Gamma(a) \in T^\uparrow(I)$. Then $P_\Gamma(h(a)) \in A_{C,q_0}$ by definition of the monadic relations in A_{C,q_0}. Now let $R(a_1, \dots, a_n) \in T^\uparrow(I)$ with $n > 1$. It follows from $R(a_1, \dots, a_n) \in I'$ and Point 2 of Lemma 26 that $(t_1, \dots, t_n)$ is R-coherent, where $t_i = \mathsf{tp}_{I'}(a_i)$. By definition of A_{C,q_0} and h, we thus have $R(h(a_1), \dots, h(a_n)) \in A_{C,q_0}$, as required. ❑

▶ **Lemma 28.** *$coCSP(A_{C,q_0}) \preceq_{\mathsf{FO}} CQA(C, q_0)$.*

Proof. Let an $\mathbf{S}'$-instance I be given. If there exists $a \in \mathsf{adom}(I)$ with $P_\Gamma(a), P_\Delta(a) \in I$ for some $\Gamma \neq \Delta$, then there is no homomorphism from I to A_{C,q_0} and "false" is returned. Otherwise define an $\mathbf{S}$-instance $T^\downarrow(I)$ by dropping all facts of the form $P_\Gamma(a)$ and adding $P(a)$ whenever $P_\Gamma(a) \in I$ with $P \in \Gamma$. Then we have $I \to A_{C,q_0}$ iff $T^\downarrow(I) \not\models_C q_0$.

"if". Assume that $T^\downarrow(I) \not\models_C q_0$ and let J be a minimal repair of $T^\downarrow(I)$ with $J \not\models q_0$. Define a map h from I to A_{C,q_0} by setting

$$h(a) = \langle \mathsf{tp}_J(a), \{P \in \mathbf{S}^{(1)} \mid P(a) \in T^\downarrow(I)\} \rangle$$

for all $a \in \mathsf{adom}(I)$. Note that, by definition of $T^\downarrow(I)$ and choice of J, $h(a)$ is indeed a constant of A_{C,q_0} for each a. It remains to show that h is a homomorphism. First let

$P_\Gamma(a) \in I$. Then $P_\Gamma(h(a)) \in A_{C,q_0}$ by definition of the monadic relations in A_{C,q_0}. Now let $R(a_1, \ldots, a_n) \in I$ with $n > 1$. It follows from $R(a_1, \ldots, a_n) \in J$ and Point 2 of Lemma 26 that $(t_1, \ldots, t_n)$ is R-coherent, where $t_i = \mathsf{tp}_J(a_i)$. By definition of A_{C,q_0} and h, we thus have $R(h(a_1), \ldots, h(a_n)) \in A_{C,q_0}$, as required.

"only if". Assume that h is a homomorphism from I to A_{C,q_0}. Define an $\mathbf{S}$-instance J by starting with $T^{\downarrow}(I)$ and dropping each fact $P(a)$ such that $h(a) = \langle t, \Gamma \rangle$ and $(P(x), x) \notin t$. It suffices to show that J is a minimal repair of I with $J \not\models q_0$. The former is obvious by definition of $T^{\downarrow}(I)$, A_{C,q_0}, and J. The latter follows from Point 1 of Lemma 26 since any homomorphism from q_0 to J gives rise to a homomorphism from q_0 to I_A via composition with h. ❏

B Proofs for Section 4

We show the PSpace lower bound claimed in Theorem 8 for (MDiC, tUCQ) using a reduction of the word problem of polynomially space-bounded Turing machines. Similar reductions have been used to establish PSpace-hardness of boundedness in linear monadic datalog [20] and of certain FO-rewritability problems in ontology-based data access [9]. It is easy to adapt the proof to (MGAV, tUCQ). Let $M = (Q, \Omega, \Gamma, \delta, q_0, q_{\mathsf{acc}}, q_{\mathsf{rej}})$ be a DTM that solves a PSpace-complete problem and $p(\cdot)$ its polynomial space bound. Here, Q is the set of states, Ω is the input alphabet, Γ the tape alphabet, $\delta : (Q \times \Gamma) \to \{L, N, R\} \times Q \times \Gamma$ the transition function, $q_0 \in Q$ the initial state, and q_{acc} q_{rej} the accepting and rejecting state, respectively. We assume that the transition function is total except on q_{acc} and q_{rej} where it is undefined for every tape symbol. The tape is assumed to be two-side infinite. We make the following additional assumptions on M. We assume that M never writes the blank symbol and with the *left (resp. right) end of the tape* we mean the first tape cell to the left (resp. right) of the head labeled with a blank. We also assume that M always terminates with the head on the right-most tape cell and that it never attempts to move left on the left-most end of the tape.

Finally and most importantly, we assume that, when started in *any* (not necessarily initial) configuration C, then the computation of M terminates. Any TM M_0 with polynomial space bound $p(\cdot)$ can be converted into a TM M that satisfies this condition as follows. Using padding, we can achieve that $2^n - 1 \geq |\Gamma|^{p(n)} \cdot p(n) \cdot |Q|$, that is, a binary counter with n bits can be used to count the number of tape cells used by M_0 started on an input of length n. We then replace every tape symbol a with symbols $(a, 0), (a, 1), (a, \bot)$, that is, add a bit value to every tape symbol; the symbol $\bot$ indicates that this tape cell is not part of the counter. In this way, the tape cells between the left end of the tape and the (right end or the) first cell carrying a symbol of the form $(a, \bot)$ represents a number encoded in binary. After each step of M_0, M traverses the tape to the left end, then to the right end, and then back to M_0's head position, incrementing the counter value of the tape (but leaving its content unchanged otherwise). If the counter has maximum value (all bits one), then M stops. Otherwise, M executes the next step of M_0. Whenever M_0 overwrites a blank with a non-blank symbol a, M writes the symbol $(a, \bot)$ instead, that is, the counter length remains unchanged. It can be verified that a straightforward implementation of the counter incrementation in terms of concrete TM transitions results in M having the desired property, that is, even if M is started in a state that is used for counter incrementation, and with *arbitrary tape content* (that can not necessarily be generated from any initial configuration), termination is still guaranteed.

Now let M be a TM that satisfies the conditions above and let $x \in \Omega^*$ be an input to M of length n. Our aim is to construct a problem CQA(C, q) in (MDiC, tUCQ) such that

CQA(C, q) is *not* FO-rewritable iff M accepts x.

A fundamental idea of the reduction is that when M accepts x, then CQA(C, q) is not FO-rewritable because any FO-rewriting would have to query for longer and longer paths that represent the accepting computation of M on x, repeated over and over again; this clearly contradicts the *locality* of an FO-query. In the reduction, we use a single constraint

$$C = \{\forall z\, \neg(B(z) \wedge B'(z))\}$$

where B, B' are monadic relation symbols. To understand the source for non-FO-rewritability that we build on, consider CQA(C, q) with $\widehat{q} = B(x) \wedge R(x, y) \wedge B'(y)$. Non-FO-rewritability is witnessed by path-shaped instances of the form

$$I_m := \{B'(b_0), R(b_1, b_0), R(b_2, b_1), \ldots, R(b_m, b_{m-1}), B(b_m)\} \cup \{B(b_i), B'(b_i) \mid 0 < i < m\}.$$

In fact, it can be verified that $I_m \models_C \widehat{q}$ for all $m > 0$, but whenever we drop a fact from I_m resulting in instance I'_m, then $I'_m \not\models_C \widehat{q}$. We are going to modify the above paths so that they describe a (repeated) accepting computation of M on x. To this end, the tape contents, the current state, and the head position are represented using the elements of $\Gamma \cup (\Gamma \times Q)$ as monadic relation symbols. Each constant on the path represents one tape cell of one configuration, the binary relation R is used to move between consecutive tape cells, the binary relation S is used to move between successor configurations inside the same computation, and the binary relation T is used to separate computations. To illustrate, suppose the computation of M on $x = ab$ consists of the two configurations qab and $aq'b$.[4] The corresponding path of length m that describes this computation (repeatedly) is

$$B'(b_0), R(b_1, b_0), S(b_2, b_1), R(b_3, b_2), T(b_4, b_3), R(b_5, b_4), \ldots, R(b_m, b_{m-1}), B(b_m)$$

with the additional monadic facts $(a, q)(c)$ for $c = b_0, b_4, b_8, \ldots,$ $b(c)$ for $c = b_1, b_5, b_9, \ldots,$ $a(c)$ for $c = b_2, b_6, b_{10}, \ldots,$ and $(b, q')(c)$ for $c = b_3, b_7, b_{11}, \ldots$. We now assemble the tUCQ q for our CQA problem. To ensure that every constant on the path is labeled with at least one symbol from $\Gamma \cup (\Gamma \times Q)$ (and since we now have three relations R, S, T instead of only a single one), we modify the query $\widehat{q}$ from above. While doing this, we also ensure that T-steps can only occur exactly after the accepting state was reached:

(r-pr) $B(x) \wedge A(x) \wedge R(x, y) \wedge A'(y) \wedge B'(y)$, for all $A \in \Gamma \cup (\Gamma \times Q)$ and all $A' \in \Gamma \cup (\Gamma \times (Q \setminus \{q_{\mathsf{acc}}, q_{\mathsf{rej}}\}))$;

(s-pr) $B(x) \wedge A(x) \wedge S(s, y) \wedge A'(y) \wedge B'(y)$, for all $A \in \Gamma \cup (\Gamma \times Q)$ and all $A' \in \Gamma \cup (\Gamma \times (Q \setminus \{q_{\mathsf{acc}}, q_{\mathsf{rej}}\}))$;

(t-pr) $B(x) \wedge A(x) \wedge T(x, y) \wedge A'(y) \wedge B'(y)$ for all $A \in \Gamma \cup (\Gamma \times Q)$ and all $A' \in \Gamma \times \{q_{\mathsf{acc}}\}$.

If we simpy use the disjunction of the above three queries as the tUCQ in our CQA problem, then that problem is not FO-rewritable. This is witnessed by paths as above in which every element is labeled with some relation symbol from $\Gamma \cup (\Gamma \times Q)$. However, these labeled witness paths need not represent proper computations of M on x since the transition relation need not be satisfied, there need not be any state, etc. We fix these problems by including additional tCQs in the tUCQ q that discover 'defects' in the computation. These queries rule out labeled path that do not describe proper computations as witnesses for non-FO-rewritability of the defined CQA problem: paths with defects are 'yes'-instances, but can be identified

[4] $uqv \in \Gamma^* Q \Gamma^*$ means that M is in state q, the tape left of the head is labeled with u, and starting from the head position, the remaining tape is labeled with v.

by an FO-query. In fact, the following queries do not mention B and B' and thus apply in a repair of an instance if and only if they apply in the original instance. They thus do not require any rewriting. The first set of additional CQs ensures that every tape cell has a unique label.

(uni) $A(x) \wedge A'(x)$ for all distinct $A, A' \in \Gamma \cup (\Gamma \times Q)$.

The next CQ enforces that there is not more than one head position per configuration:

(h1) $\bigwedge_{0 \le l < i}\{R(x_l, x_{l+1})\} \wedge (a,q)(x_i) \wedge \bigwedge_{0 \le l < j} R(y_l, y_{l+1}) \wedge (a',q')(y_j)$, for all $i < j < p(n)$, $(a,q), (a',q') \in \Gamma \times Q$, and $x_0 = y_0$.

and that there is at least one head position per configuration:

(h2) $R(x_0, x_1) \wedge \ldots \wedge R(x_{p(n)-2}, x_{p(n)-1}) \wedge a_1(x_0) \wedge \ldots \wedge a_{p(n)-1}(x_{p(n)-1})$, for all sequences $a_0, \ldots, a_{p(n)-1} \in \Gamma$.

We ensure that configurations have at most length $p(n)$ using the CQ

(l1) $R(x_0, x_1) \wedge \ldots \wedge R(x_{p(n)-1}, x_{p(n)})$.

We also ensure that configurations are not shorter than $p(n)$ (with the possible exception of the first configuration, which can be shorter):

(l2) $\rho(x_0, x_1) \wedge R(x_1, x_2) \wedge \ldots \wedge R(x_i, x_{i+1}) \wedge \rho'(x_{i+1}, x_{i+2})$ for all $i < p(n) - 1$ and $\rho, \rho' \in \{S, T\}$.

We now enforce that the transition function is respected and that the content of tape cells which are not under the head does not change. Let forbid denote the set of all tuples (A_1, A_2, A_3, A) with $A_i \in \Gamma \cup (\Gamma \times Q)$ such that whenever three consecutive tape cells in a configuration c are labeled with A_1, A_2, A_3, then in the successor configuration c' of c, the tape cell corresponding to the middle cell *cannot* be labeled with A:

(con) $A(x_0) \wedge R(x_0, x_1) \wedge \ldots \wedge R(x_{i-1}, x_i) \wedge S(x_i, y_0) \wedge T(y_0, y_1) \wedge \ldots \wedge$
$R(y_{p(n)-i-3}, y_{p(n)-i-2}) \wedge A_3(y_{p(n)-i-2}) \wedge R(y_{p(n)-i-2}, y_{p(n)-i-1}) \wedge$
$A_2(y_{p(n)-i-1}) \wedge R(y_{p(n)-i-1}, y_{p(n)-i}) \wedge A_1(y_{p(n)-i})$
for all $0 \le i < p(n)$ and $(A_1, A_2, A_3, A) \in \mathsf{forbid}$.

It remains to set up the initial configuration. Recall that witness instances consist of repeated computations of M, which ideally we would all like to start in the initial configuration for input x. It does not seem possible to enforce this for the first computation in the instance, so we live with this computation starting in some unknown configuration, relying on our assumption that M terminates also when started in an arbitrary configuration. Then, we utilize the final states q_{acc} and q_{rej} to enforce that all computations in the instance except the first one must start with the initial configuration for x. Let $A_0^{(0)}, \ldots, A_{p(n)-1}^{(0)}$ be the monadic relation symbols that describe the initial configuration, i.e., when the input x is $x_0 \cdots x_{n-1}$, then $A_0^{(0)} = (x_0, q_0)$, $A_i^{(0)} = x_i$ for $1 \le i < n$, and $A_i^{(0)} = x_i$ is the blank symbol for $n \le i < p(n)$. Now take

(in) $\bigwedge_{0 \le l < i} R(x_l, x_{l+1}) \wedge T(x_i, x_{i+1}) \wedge A(x_0)$ for all $0 \le i < p(n)$ and $A \ne A_i^{(0)}$.

The query q used in the CQA problem is the UCQ defined by taking the union of all CQs given above. The following lemma establishes the correctness of our reduction.

▶ **Lemma 29.** *CQA(C, q) is* not *FO-rewritable iff M accepts x.*

Proof. "if". Assume that M accepts x. By using standard locality arguments (e.g., Hanf's Theorem), it is enough to show that there exist arbitary large k and instances I_k with domain $\{b_0, \ldots, b_k\}$ such that

- for all $i, j \le k$: if $\rho(b_i, b_j) \in I_k$ for some $\rho \in \{R, S, T\}$, then $i = j + 1$ or $j = i + 1$;

- $I_k \models_C q$;
- $J \not\models_C q$, where J is the disjoint union of the instances I_k^0 and I_k^k, where I_k^0 is obtained from I_k by removing all facts involving b_0 and I_k^k is obtained from I_k by removing all facts involving b_k.

Assume $k > 0$ is given. Let $C_1, \ldots, C_m$ be a sequence of configurations of length $p(n)$ obtained by sufficiently often repeating the accepting computation of M on x so that $|C_1| + \cdots + |C_m| \geq k$. We can convert $C_1, \ldots, C_m$ into the desired witness instance I_k in a straightforward way: introduce one individual name for each tape cell in each configuration and computation, use R to connect cells within the same configuration, S to connect configurations, and T to connect computations, and the relation symbols from $\Gamma \cup (\Gamma \times Q)$ to indicate the tape inscription, current state, and head position. We obtain an instance satisfying the conditions above by identifying the individuals with $b_0, \ldots, b_k$ assuming that b_0 stands for the first cell of the first configuration of C_1. Finally add the facts $\{B'(b_0)\} \cup \{B(b_k))\} \cup \{B(b_i), B'(b_i) \mid 0 < i < k\}$ to obtain I_k. It can be verified that I_k is as required. To see that $I_k \models_C q$ observe that in any minimal repair I'_k of I_k there is some i with $0 \leq i < k$ such that $B'(b_i) \in I'_k$ and $B(b_{i+1}) \in I'_k$. To see that $J \not\models_C q$ for the disjoint union J of I_k^0 and I_k^k, observe that one obtains a minimal repair of J by satisfying B' everywhere in I_k^0 and B everywhere in I_k^k.

"only if". Assume that $\mathrm{CQA}(C, q)$ is not FO-rewritable. Note that all CQs in q that are distinct from (r-pr), (s-pr), and (t-pr) have a match in an instance I iff they have a match in all minimal repairs of I w.r.t. $\mathrm{CQA}(C, q)$. Thus, they are FO-rewritable. Now consider the following

Observation. Assume I is an instance such that no CQ in q distinct from the CQs (r-pr), (s-pr), and (t-pr) has a match in I. Then $I \models_C q$ iff there exists $k > 0$ such that the following condition $(*_k)$ holds: there are

$$\rho_0(b_1, b_0), \ldots, \rho_{k-1}(b_k, b_{k-1}), A_0(b_0), \ldots, A_k(b_k) \in I$$

with $\rho_i \in \{R, S, T\}$ for all $i < k$ and $A_i \in \Gamma \cup (\Gamma \times Q)$ for all $i \leq k$ such that

- $B'(b_0) \in I_k, B(b_0) \notin I_k$,
- $B(b_k) \in I_k$, $B'(b_k) \notin I_k$,
- $B(b_i), B'(b_i) \in I_k$ for all $0 < i < k$,
- if $\rho_{i+1} \in \{R, S\}$, then $A_i \in \Gamma \cup (\Gamma \times (Q \setminus \{q_{\mathsf{acc}}, q_{\mathsf{rej}}\}))$,
- if $\rho_{i+1} = T$, then $A_i \in \Gamma \cup (\Gamma \times \{q_{\mathsf{acc}}\})$.

Clearly, for every $k > 0$ condition $(*_k)$ can be expressed in FO. Thus, if $\mathrm{CQA}(C, q)$ is not FO-rewritable, then for every $k > 0$ there exists an instance I satisfying $(*_k)$. Now let m_0 be the maximum number of steps M makes starting from any configuration of length $p(n)$ before entering the final state. One can prove that any I satifying $(*_k)$ for $k \geq 2m_0(p(n) + 1) + 1$ encodes an accepting computation of M for input x, as required. ❏

C Proofs for Section 5

The following small observation shows that we can w.l.o.g. assume that every tCQ in q_0 contains *exactly* one atom with a relation symbol distinct from A_1, A_2.

▶ **Lemma 30.** *Let $CQA(C, q_0)$ be a restricted binary CQA problem. If q_0 contains a tCQ that refers only to the relation symbols A_1 and A_2, then $CQA(C, q_0)$ is in AC_0.*

Proof. (sketch). Let CQA(C, q_0) be a binary restricted CQA problem. Note that every tCQ q in q_0 that refers only to the relation symbols A_1 and A_2 can only contain a single variable x since q is a hypertree. We thus have the following cases. (i) q_0 contains the tCQ $A_1(x) \wedge A_2(x)$. Then the answer is always 'false'. (ii) q_0 contains both the tCQ $A_1(x)$ and the tCQ $A_2(x)$, but not the tCQ $A_1(x) \wedge A_2(x)$. Then we can simply answer the query q_0' over the original instance, where q_0' is obtained from q_0 by removing from any disjunct of q_0 containing an atom distinct from A_1 and A_2 the atoms involving A_1 and A_2. (iii) q_0 contains the tUCQ $A_i(x)$, but neither the tUCQ $A_{\bar{i}}(x)$ nor the tCQ $A_1(x) \wedge A_2(x)$ where $i \in \{1, 2\}$, $\bar{i} = 3 - i$. Then there is only one candidate instance for a repair, namely where $A_{\bar{i}}(a)$ is removed whenever $A_1(x)$ and $A_2(x)$ are in the original instance. In AC$_0$, we can clearly compute this repair and check whether q applies in it. ❑

▶ **Lemma 31.** *CQA*$(C, q_0) \preceq_{\mathsf{FO}}$ *coCSP*(A_{C,q_0}).

Proof. Let I be an $\mathbf{S}$-instance. Construct an $\mathbf{S}'$-instance $T^{\uparrow}(I)$ by removing all facts $A_i(a)$ and adding $P_\Gamma(a)$ for each a, where $\Gamma = \{A_i \mid A_i(a) \in I\}$. Clearly, $T^{\uparrow}$ is FO-definable. Moreover, we have $I \not\models_C q_0$ iff $T^{\uparrow}(I) \rightarrow A_{C,q_0}$.

"if". Let h be a homomorphism from $T^{\uparrow}(I)$ to A_{C,q_0}. Define an $\mathbf{S}$-instance I' by starting with I and dropping all facts $A_{\bar{i}}(a)$ such that $A_1(a), A_2(a)$ are in I and $h(a) = a_i$ (note that we cannot have $h(a) = 0$ then because 0 is not labeled with $P_{\{A_1,A_2\}}$ in A_{C,q_0}). Clearly, I' is a minimal repair of I. We show that $I' \not\models q_0$. Assume to the contrary that $I' \models q_0$. Then there is a tCQ q in q_0 with $I' \models q$. Let h' be a homomorphism from q to I' and let $R(x_1, \ldots, x_n)$ be the unique atom in q with R distinct from A_1, A_2. Obviously, we have $R(h'(x_1), \ldots, h'(x_k)) \in T^{\uparrow}(I)$ and thus $R(h(h'(x_1)), \ldots, h(h'(x_k))) \in A_{C,q_0}$. Since q is a tCQ and thus connected, all atoms in q that are of the form $A_i(x)$ satisfy $x \in \{x_1, \ldots, x_n\}$. For each atom $A_i(x_j) \in q$, we must have $A_i(h'(x_j)) \in I'$ and $A_{\bar{i}}(h'(x_j)) \notin I'$. Consequently, $A_i(x_j) \in q$ implies $h(h'(x_j)) = i$: if $A_1(a)$ and $A_2(a)$ are in I, then this is a consequence of the construction of I'; if only $A_i(a)$ is in I, then $P_{\{A_i\}}(a) \in T^{\uparrow}(I)$ and thus $h(a) = i$. It is thus easy to verify that the instance in Point 4 in the definition of A_{C,q_0} for the fact $R(h(h'(x_1)), \ldots, h(h'(x_k)))$ implies q, thus q_0. This is a contradiction to $R(h(h'(x_1)), \ldots, h(h'(x_k)))$ being in A_{C,q_0}.

"only if". Assume that $I \not\models_C q_0$ and let I' be a minimal repair of I such that $I' \not\models q_0$. Define a function h by setting $h(a) = i$ if $A_i(a) \in I'$ and $A_{\bar{i}}(a) \notin I'$, $i \in \{1, 2\}$, and $h(a) = 0$ if $A_i(a) \notin I'$ and $A_{\bar{i}}(a) \notin I'$. We show that h is a homomorphism from $T^{\uparrow}(I)$ to A_{C,q_0}. Take a $P_\Gamma(a) \in T^{\uparrow}(I)$. If $\Gamma = \emptyset$, then neither $A_1(a)$ nor $A_2(a)$ are in I', thus $h(a) = 0$ and $P_\emptyset(0) \in A_{C,q_0}$. If $\Gamma = \{A_i\}$ for $i \in \{1, 2\}$, then $A_i(a)$ is in I' but $A_{\bar{i}}(a)$ is not, thus $h(a) = i$ and $P_{\{A_i\}}(i) \in A_{C,q_0}$. If $\Gamma = \{A_1, A_2\}$, then there is an $i \in \{1, 2\}$ such that $A_i(a)$ is in I' but $A_{\bar{i}}(a)$ is not. Thus $h(a) = i$ and $P_{\{A_1,A_2\}}(i) \in A_{C,q_0}$. Finally, let $R(a_1, \ldots, a_k) \in T^{\uparrow}(I)$. Assume to the contrary of what is to be shown that $R(h(a_1), \ldots, h(a_k)) \notin A_{C,q_0}$. Then the instance in Point 4 in the definition of A_{C,q_0} for the fact $R(h(a_1), \ldots, h(a_k))$ implies q_0. Note that if that instance contains $A_j(h(a_i))$, then $h(a_i) = j$. By definition of h, we then have $A_j(a_i) \in I'$. Additionally using that $R(a_1, \ldots, a_k) \in I'$ and that no atom in q_0 uses the same variable multiple times since each CQ in q_0 is a hypertree, it is thus easy to verify that I' entails q_0, a contradiction. ❑

▶ **Lemma 32.** *coCSP*$(A_{C,q_0}) \preceq_{\mathsf{FO}}$ *CQA*(C, q_0).

Proof. Let an $\mathbf{S}'$-instance I be given. If I contains facts $P_\Gamma(a)$ and $P_\Lambda(a)$ with $\Gamma \neq \Lambda$, then answer 'false' as there is clearly no homomorphism to A_{C,q_0}. Otherwise, replace every fact

$P_\Gamma(a)$ with $A_i(a)$ for all $A_i \in \Gamma$. Call the resulting $\mathbf{S}$-instance $T^\downarrow(I)$. Clearly, the map $T^\downarrow$ is FO-definable. Moreover, $I \rightarrow A_{C,q_0}$ iff $T^\downarrow(I) \not\models_C q_0$.

"if". Assume that $T^\downarrow(I) \not\models_C q_0$ and let J be a minimal repair of $T^\downarrow(I)$ such that $J \not\models q_0$. Define a function h from $\mathsf{adom}(I)$ to $\mathsf{adom}(A_{C,q_0})$ by setting $h(a) = i$ if $A_i(a) \in J$ and $A_{\bar{i}}(a) \notin J$, $i \in \{1,2\}$, and $h(a) = 0$ if $A_i(a) \notin J$ and $A_{\bar{i}}(a) \notin J$. We show that h is a homomorphism from I to A_{C,q_0}. Take a $P_\Gamma(a) \in I$. If $\Gamma = \emptyset$, then[5] neither $A_1(a)$ nor $A_2(a)$ are in J, thus $h(a) = 0$ and $P_\emptyset(0) \in A_{C,q_0}$. If $\Gamma = \{A_i\}$ for $i \in \{1,2\}$, then $A_i(a)$ is in J but $A_{\bar{i}}(a)$ is not, thus $h(a) = i$ and $P_{\{A_i\}}(i) \in A_{C,q_0}$. If $\Gamma = \{A_1, A_2\}$, then there is an $i \in \{1,2\}$ such that $A_i(a)$ is in J but $A_{\bar{i}}(a)$ is not. Thus $h(a) = i$ and $P_{\{A_1,A_2\}}(i) \in A_{C,q_0}$. Finally, let $R(a_1,\dots,a_k) \in I$. Assume to the contrary of what is to be shown that $R(h(a_1),\dots,h(a_k)) \notin A_{C,q_0}$. Then the instance in Point 4 in the definition of A_{C,q_0} for the fact $R(h(a_1),\dots,h(a_k))$ entails q_0. Note that if that instance contains $A_j(h(a_i))$, then $h(a_i) = j$. By definition of h, we then have $A_j(a_i) \in J$. Additionally using that $R(a_1,\dots,a_k) \in J$ and that no atom in q uses the same variable multiple times since each CQ in q is a hypertree, it is thus easy to verify that J entails q_0, a contradiction.

"only if". Let h be a homomorphism from I to A_{C,q_0}. Define an $\mathbf{S}$-instance J by starting with $T^\downarrow(I)$ and dropping all facts $A_{\bar{i}}(a)$ such that $A_1(a), A_2(a)$ are in $T^\downarrow(I)$ and $h(a) = a_i$ (note that we cannot have $h(a) = 0$ then because 0 is not labeled with $P_{\{A_1,A_2\}}$ in A_{C,q_0}). Clearly, J is a minimal repair of $T^\downarrow(I)$. We show that $J \not\models q_0$. Assume to the contrary that $J \models q_0$. Then there is a tCQ q in q_0 with $J \models q$. Let h' be a homomorphism from q to J and let $R(x_1,\dots,x_n)$ be the unique atom in q with R distinct from A_1, A_2. Obviously, we have $R(h(h'(x_1)),\dots,h(h'(x_k))) \in A_{C,q_0}$. Since q is a tCQ and thus connected, all atoms in q that are of the form $A_i(x)$ satisfy $x \in \{x_1,\dots,x_n\}$. For each atom $A_i(x_j) \in q$, we must have $A_i(h'(x_j)) \in J$ and $A_{\bar{i}}(h'(x_j)) \notin J$. Consequently, $A_i(x_j) \in q$ implies $h(h'(x_j)) = i$: if $A_1(a)$ and $A_2(a)$ are in $T^\downarrow(I)$, then this is consequence of the construction of J; if only $A_i(a)$ is in $T^\downarrow(I)$, then $P_{\{A_i\}}(a) \in I$ and thus $h(a) = i$. It is thus easy to verify that the instance in Point 4 in the definition of A_{C,q_0} for the fact $R(h(h'(x_1)),\dots,h(h'(x_k)))$ implies q, thus q_0. This is a contradiction to $R(h(h'(x_1)),\dots,h(h'(x_k)))$ being in A_{C,q_0}. ❏

D Proofs for Section 6

▶ **Lemma 33.** $\Pi \preceq_{\mathsf{FO}} \mathit{CQA}(C_\Pi, q_\Pi)$.

Proof. Let an $\mathbf{S}$-instance I be given. Convert I into an $\mathbf{S}'$-instance $T^\uparrow(I)$ by extending I with $Q_t(b)$ for all $b \in \mathsf{adom}(I)$ and $t \in \mathsf{tp}$. Clearly, $T^\uparrow$ is FO-definable. We show that $T^\uparrow(I) \models_{C_\Pi} q_\Pi$ iff $I \models \Pi$.

"if". Assume that $T^\uparrow(I) \not\models_{C_\Pi} q_\Pi$. Let J be a minimal repair of $T^\uparrow(I)$ such that $J \not\models q_\Pi$. Then for each $a \in \mathsf{adom}(J)$, there exists a unique $t_a \in \mathsf{tp}$ such that $Q_t(a) \in J$ iff $t \neq t_a$ for all $t' \in \mathsf{tp}$. Define an extension of I to schema $\mathbf{S} \cup \mathbf{Q}$ by adding $P(a)$ whenever $P \in t_a$. Since $T^\uparrow(I) \not\models_{C_\Pi} q_\Pi$, I satisfies all non-goal rules in Π, but no goal rule applies. Consequently, we have $I \not\models \Pi$ as required.

"only if". Assume that $I \not\models \Pi$ and let I' be an extension of I to schema $\mathbf{S} \cup \mathbf{Q}$ that satisfies all non-goal rules in Π but where no goal rule applies. Define J as $T^\uparrow(I)$ with $Q_t(a)$ removed whenever the $\mathbf{Q}$-type realized by a in I' is t, that is, $t = \{P \mid P \in \mathbf{Q} \text{ and } P(a) \in I'\}$.

[5] Here and in the following, we implicitly make use of the fact that for all $a \in \mathsf{adom}(I)$, $P_\Gamma(a), P_\Delta(a) \in I$ implies $\Gamma = \Delta$.

It is easy to verify that J is a minimal repair of $T^{\uparrow}(I)$. It thus remains to show that $J \not\models q_\Pi$, which is also easy. Since all non-goal rules of Π are satisfied in I', no CQ under Point 1 of the definition of q_Π applies in J; and since no goal rule applies in I', no CQ under Point 2 applies either. ❑

▶ **Lemma 34.** *$CQA(C_\Pi, q_\Pi) \preceq_{\mathsf{FO}} \Pi$.*

Proof. Let I be an $\mathbf{S}'$-instance. Denote by X the set of $a \in \mathsf{adom}(I)$ such that there are at least two distinct $t_1, t_2 \in \mathsf{tp}$ with neither $Q_{t_1}(a) \in I$ nor $Q_{t_2}(a) \in I$. Then $T^{\downarrow}(I)$ is obtained from I by dropping all facts that involve a constant from X or a relation symbol that is not from $\mathbf{S}$, and adding all facts $S_t(a)$ such that $Q_{t'}(b) \in I$ iff $t' \neq t$ for all $t' \in \mathsf{tp}$. Note that we rely on Π to admit precoloring by using the relations S_t. Clearly $T^{\downarrow}(I)$ is FO-definable. We show that $I \models_{C_\Pi} q_\Pi$ iff $T^{\downarrow}(I) \models \Pi$.

"if". Assume that $I \not\models_{C_\Pi} q_\Pi$. Then there is a minimal repair I' of I with $I' \not\models q_\Pi$. Since I' is a minimal repair and by definition of X, for every $a \in \mathsf{adom}(I) \setminus X$ there must be a unique $t_a \in \mathsf{tp}$ such that $Q_t(a) \in I'$ iff $t \neq t_a$. Define an extension J of $T^{\downarrow}(I)$ to schema $\mathbf{S} \cup \mathbf{Q}$ by adding $P(a)$ whenever $a \in \mathsf{adom}(T^{\downarrow}(I))$ and $P \in t_a$. It is easy to see that whenever $S_t(a)$ was added during the construction of $T^{\downarrow}(I)$, then $t = t_a$. Therefore, J satisfies all non-goal rules in Π that involve one of the S_t relations. All other non-goal rules in Π are satisfied in J since $I' \not\models q_\Pi$. For the same reason, no goal rule of Π applies in J. Consequently, $T^{\downarrow}(I) \not\models \Pi$.

"only if". Assume that $T^{\downarrow}(I) \not\models \Pi$. Then there is an extension J of $T^{\downarrow}(I)$ to schema $\mathbf{S} \cup \mathbf{Q}$ in which all non-goal rules of Π are satisfied and no goal rule applies. For each $a \in \mathsf{adom}(J)$, let t_a be the $\mathbf{Q}$-type realized by a in J. Let I' be obtained from I by removing every fact $Q_{t_a}(a)$ for all $a \in \mathsf{adom}(J)$. Then I' is a minimal repair of I. To prove this, it suffices to show that whenever $Q_{t_a}(a)$ has been removed from I during the construction of I', then $Q_t(a) \in I$ for all $t \in \mathsf{tp}$. Assume that this is not the case. Since $a \in \mathsf{adom}(J)$, we must have $a \notin X$, and thus there is a unique $t \in \mathsf{tp}$ with $Q_t(a) \notin I$. By construction of $T^{\downarrow}(I)$, we have $S_t(a) \in T^{\downarrow}(I)$ and since J satisfies all non-goal rules of Π that involve an S_t relation, we must have $t_a = t$. Thus $Q_{t_a}(a) \notin I$, in contradiction to it being removed during the construction of I'. It remains to show that $I' \not\models q_\Pi$. Note that the CQs in q_Π are constructed such that every body variable x is 'guarded' by a conjunction of the form $\mathsf{con}_t(x)$, $t \in \mathsf{tp}$. It follows that no match of a CQ in q_Π can involve a constant from X. However, all remaining potential matches involve only constants from J, and consequently each such match implies that some non-goal rule of Π is violated in J or that some goal rule of Π applies in J, which is a contradiction. ❑

▶ **Lemma 35.** *$CQA(C, q) \preceq_{\mathsf{FO}} \Pi_{C,q}$.*

Proof. Let an $\mathbf{S}$-instance I be given. Define $T^{\uparrow}(I)$ exactly as in the proof of Lemma 4: drop all facts that involve a monadic relation symbol and add $P_{\Gamma_a}(a)$ for every $a \in \mathsf{adom}(I)$, where $\Gamma_a = \{P \mid P(a) \in I\}$. It remains to show that $I \models_C q$ iff $T^{\uparrow}(I) \models \Pi_{C,q}$.

"if". Assume $I \not\models_C q$. Then there is a minimal repair I' of I such that $I' \not\models q$. Clearly, for every $a \in \mathsf{adom}(I)$, there must be a $\Lambda_a \in \mathsf{rep}(\Gamma_a)$ such that $P(a) \in I'$ iff $P \in \Lambda_a$ for all $P \in \mathbf{S}^{(1)}$. Extend the $\mathbf{S}'$-instance $T^{\uparrow}(I)$ to schema $\mathbf{S}' \cup \mathbf{Q}$ by adding the facts $Q_{\Lambda_a}(a)$ as well as $P(a)$ whenever $P \in \Lambda_a$, for each $a \in \mathsf{adom}(I)$. Call the resulting instance J. It is readily checked that J satisfies all non-goal rules of Π. Moreover, no goal rule applies. In particular, I' and J satisfy exactly the same $\mathbf{S}$-facts, and thus $J \not\models q$ because $I' \not\models q$.

"only if". Assume $T^{\uparrow}(I) \not\models \Pi_{C,q}$. Then there is an extension J of $T^{\uparrow}(I)$ to schema $\mathbf{S}' \cup \mathbf{Q}$ such that all non-goal rules in $\Pi_{C,q}$ are satisfied, but no goal rule applies. Assume that J is a minimal such extension regarding set inclusion. Note that $J \not\models q$. By the rules in $\Pi_{C,q}$, we find, for each $a \in \mathsf{adom}(J)$, a $\Lambda_a \in \mathsf{rep}(\Gamma_a)$ such that $Q_{\Lambda_a}(a) \in J$. By the minimality of J, we have $P(a) \in J$ iff $P \in \Lambda_a$, for all $P \in \mathbf{S}^{(1)}$. Define I' by removing from I all facts $P(a)$ that are not in J. Clearly, I' is a minimal repair of I. Moreover, $I' \not\models q$ because I' and J make true the same $\mathbf{S}$-facts and $J \not\models q$. ❑

▶ **Lemma 36.** $\Pi_{C,q} \preceq_{\mathsf{FO}} CQA(C, q)$.

Proof. Let an $\mathbf{S}'$-instance I be given. If $P_\Gamma(a), P_\Lambda(a) \in I$ for any a and two distinct Γ, Λ, then answer 'yes'. Otherwise define $T^\downarrow(I)$ exactly as in the proof of Lemma 4: for each constant $a \in \mathsf{adom}(I)$, let Γ_a be the unique subset of $\mathbf{S}^{(1)}$ such that $P_{\Gamma_a}(a) \in I$. Replace each fact $P_{\Gamma_a}(a)$ with the facts $P(a)$, $P \in \Gamma_a$ and call the result $T^\downarrow(I)$. We show that $I \models \Pi_{C,q}$ iff $T^\downarrow(I) \models_C q$.

"if". Assume $I \not\models \Pi_{C,q}$. Then there is an extension I' of I with to schema $\mathbf{S}' \cup \mathbf{Q}$ such that I' satisfies all non-goal rules of $\Pi_{C,q}$, but no goal rule applies in I'. Assume that I' is a minimal such extension regarding set inclusion. Note that $I' \not\models q$. By the rules in $\Pi_{C,q}$, we find, for each constant a in I, a $\Lambda_a \in \mathsf{rep}(\Gamma_a)$ such that $Q_{\Lambda_a}(a) \in I'$. By the minimality of I' and the uniqueness of Γ_a, we have $P(a) \in I'$ iff $P \in \Lambda_a$, for all $P \in \mathbf{S}^{(1)}$ and $a \in \mathsf{adom}(I)$. Define J by removing from $T^\downarrow(I)$ all facts $P(a)$ that are not in I'. Clearly, J is a minimal repair of $T^\downarrow(I)$. Moreover, $J \not\models q$ because J and I' make true the same $\mathbf{S}$-facts and $I' \not\models q$.

"only if". Assume $T^\downarrow(I) \not\models_C q$. Then there is a minimal repair J of $T^\downarrow(I)$ such that $J \not\models q$. Clearly, for every $a \in \mathsf{adom}(I)$, there must be a $\Lambda_a \in \mathsf{rep}(\Gamma_a)$ such that $P(a) \in J$ iff $P \in \Lambda_a$ for all $P \in \mathbf{S}^{(1)}$. Extend I to schema $\mathbf{S}' \cup \mathbf{Q}$ by adding the fact $Q_{\Lambda_a}(a)$ for each a and $P(a)$ whenever $P \in \Lambda_a$. Call the resulting instance I'. It is readily checked that I' satisfies all non-goal rules of $\Pi_{C,q}$. Moreover, no body of a goal rule applies. In particular, I' and J make true exactly the same $\mathbf{S}$-facts, and thus $J \not\models q$ because $I' \not\models q$. ❑

E Proofs for Section 7

Lemma 19. For every GDDLog program Π that admits precoloring, there is a normalized GDDLog-Program Π' with $\Pi \approx_{\mathsf{FO}} \Pi'$.

Proof. Let Π be a GDDLog program over schema $\mathbf{S}$ that admits precoloring and let m be the maximum arity of EDB and IDB relation symbols in Π. We may assume that the rules of Π are closed under identifying variables, that is, if ρ' can be obtained from $\rho \in \Pi$ by identifying variables, then $\rho' \in \Pi$. For any tuple of variables $\mathbf{x}$, we use $\delta(\mathbf{x})$ to denote the tuple of variables obtained from $\mathbf{x}$ by deleting each but the first occurrence of each variable. We use $f(\mathbf{x})$ to denote the *footprint* of $\mathbf{x}$, that is, $f(\mathbf{x})$ is the tuple obtained from $\mathbf{x}$ by replacing each variable with its position in $\delta(\mathbf{x})$. A *footprint of length* i is a tuple $f \in \{1, \ldots, i\}^i$ such that $(1, 1) \in f$ and from $(k, \ell) \in f$ and $k > 1$ it follows that $(k', \ell - 1)$ for some $k' < k$. Let schema $\widehat{\mathbf{S}}$ consist of relation symbols R^f where $R \in \mathbf{S}$ is of arity i and f is a footprint of length i.

We next construct a GDDLog program $\widehat{\Pi}$ over schema $\widehat{\mathbf{S}}$ that satisfies the first two items of Condition 1 of normalized GDDLog programs. In a second step, we extend $\widehat{\Pi}$ to satisfy also the third item of Condition 1 and Condition 2. To construct $\widehat{\Pi}$, start with Π and modify every rule

$$R_1(\mathbf{y}_1) \wedge \cdots \wedge R_k(\mathbf{y}_k) \rightarrow S_1(\mathbf{z}_1) \vee \cdots \vee S_m(\mathbf{z}_m)$$

as follows:

(i) replace every $R_i(\mathbf{y}_i)$ with $R_i^{f(\mathbf{y}_i)}(\mathbf{y})$ where $\mathbf{y}$ is obtained from $\delta(\mathbf{y}_i)$ by filling up to length m with fresh variables;

(ii) consider each head atom $S_i(\mathbf{z}_i)$ with $S_i \neq \mathsf{goal}$; due to Step (i), there is a body atom $R(\mathbf{y})$ with R of arity m and such that $\mathbf{z}_i \subseteq \mathbf{y}$; replace $S_i(\mathbf{z}_i)$ with $S_i^{f(\mathbf{z}_i)}(\mathbf{z})$ where $\mathbf{z}$ is obtained from $\delta(\mathbf{z}_i)$ by filling up to length m with the variables from $\mathbf{y} \setminus \mathbf{z}_i$ (in any order and without repetitions).

Let $\mathbf{Q}'$ denote the set of IDB relations in $\widehat{\Pi}$. Let $\mathbf{S}'$ be the extension of $\widehat{\mathbf{S}}$ with an m-ary relation symbol S_t for each $\mathbf{Q}'$-type t. Let Π' be obtained from $\widehat{\Pi}$ by adding the rules required for the third item of Condition 1 and the rules that involve the S_t symbols and are required by Condition 2 of normalized GDDLog programs.

It remains to show that $\Pi \preceq_{\mathsf{FO}} \Pi'$ and vice versa. We confine ourselves to a sketch. For $\Pi \preceq_{\mathsf{FO}} \Pi'$, let an $\mathbf{S}$-instance I be given. Construct an $\mathbf{S}'$-instance $T^{\uparrow}(I)$ by selecting elements $a_1, \dots, a_m \in \mathsf{adom}(I)$ and replacing every tuple $R(\mathbf{a})$ in I with $R^{f(\mathbf{a})}(\mathbf{b})$ where $\mathbf{b}$ is obtained from $\delta(\mathbf{a})$ by appending distinct elements from $a_0, \dots, a_m$ until length m is reached and where $f(\mathbf{a})$ and $\delta(\mathbf{a})$ are defined in exactly the same way as for sequences of variables.[6] It can be verified that $I \models \Pi$ iff $T^{\uparrow}(I) \models \Pi'$.

To show $\Pi' \preceq_{\mathsf{FO}} \Pi$, let an $\mathbf{S}'$-instance I be given. If I contains a fact $R^f(\mathbf{a})$ where $\mathbf{a}$ contains repititions, then return "$I \models \Pi$". Otherwise construct an $\mathbf{S}$-instance $T^{\downarrow}(I)$ by applying the following operations:

- replace every fact $R^f(\mathbf{a})$ where $f = \ell_1 \cdots \ell_i$ and $\mathbf{a} = a_1 \cdots a_m$ with $R(a_{\ell_1}, \dots, a_{\ell_i})$;
- replace every fact $S_t(\mathbf{a})$ (where t is a $\mathbf{Q}'$-type) with $S_{t'}(\mathbf{a})$, where t' is the following m-type:
$$t' = \{R(x_{i_1}, \dots, x_{\ell_i}) \mid R^f(x_1, \dots, x_m) \in t \text{ and } f = \ell_1 \cdots \ell_i\}.$$

Again, one can verify that $T^{\downarrow}(I) \models \Pi$ iff $I \models \Pi'$. ❏

▶ **Lemma 37.** $\Pi \preceq_{\mathsf{FO}} \mathit{CQA}(C_\Pi, q_\Pi)$.

Proof. Assume that an $\mathbf{S}$-instance I is given. Define an $\mathbf{S}'$-instance $T^{\uparrow}(I)$ as the extension of I with all facts $Q_t(\mathbf{a})$ such that t is a $\mathbf{Q}$-type and $\mathbf{a} \in \mathsf{adom}(I)^m$. We show that $I \models \Pi$ iff $T^{\uparrow}(I) \models_{C_\Pi} q_\Pi$.

"if". Assume that $I \not\models \Pi$. Then there is an extension I' of I to schema $\mathbf{S} \cup \mathbf{Q}$ such that all non-goal rules in Π are satisfied in I' and no goal rule of Π applies in I'. For each tuple $\mathbf{a} \in \mathsf{adom}(I)^m$, we use $\mathsf{tp}_{I'}(\mathbf{a})$ to denote the $\mathbf{Q}$-type realized by $\mathbf{a} = a_1 \cdots a_m$ in I', that is, $\mathsf{tp}_{I'}(\mathbf{a})$ is the set of all atoms $R(\mathbf{x})$ such that $R \in \mathbf{Q}$, and $\mathbf{x} = x_{i_1}, \dots, x_{i_m}$ is a permutation of $x_1 \cdots x_m$ such that $R(a_{i_1}, \dots, a_{i_m}) \in I'$. Let J be the $\mathbf{S}'$-instance obtained from $T^{\uparrow}(I)$ by dropping all facts $Q_t(\mathbf{a})$ such that $\mathsf{tp}_{I'}(\mathbf{a}) = t$. We show that J is a minimal repair of $T^{\uparrow}(I)$ with $J \not\models q_\Pi$. Regarding the former, J is a minimal repair since, whenever I contains $Q_t(\mathbf{a})$ for all $\mathbf{Q}$-types t thus violating a constraint in C_Π, then exactly one of those facts is removed in J, the one with $t = \mathsf{tp}_{I'}(\mathbf{a})$. Regarding the latter, we start with noting that J is proper and that I' is the $\mathbf{S} \cup \mathbf{Q}$-instance of J. Since I' satisfies all goal rules in Π and no body of a goal rule of Π applies in I', the CQs under Point 1 of the definition of q_Π do not apply in J. The CQs under Point 2 of the definition q_Π do not apply by construction of J.

"only if". Assume that $T^{\uparrow}(I) \not\models_{C_\Pi} q_\Pi$. Then there is a minimal repair J of $T^{\uparrow}(I)$ such that $J \not\models q_\Pi$. By construction of $T^{\uparrow}(I)$ and since J satisfies all constraints in C_Π, every $\mathbf{a} \in \mathsf{adom}(J)^m$ is assigned a $\mathbf{Q}$-type $t_\mathbf{a}$ in J. Since none of the CQs under Point 2 of the

[6] We assume here that $\mathsf{adom}(I)$ contains at least m constants. Instances of smaller cardinality can be dealt with in a straightfoward way.

definition of q_Π applies in J, this $\mathbf{Q}$-type is unique. Define an extension I' of I to schema $\mathbf{S}\cup\mathbf{Q}$ by adding $R(a_{i_1},\ldots,a_{i_m})$ whenever $R(x_{i_1},\ldots,x_{i_m})\in t_{\mathbf{a}}$ and $\mathbf{a}=a_1\cdots a_m$. Note that I' is the $\mathbf{S}\cup\mathbf{Q}$-instance in J. To establish that $I\not\models\Pi$ as desired, it suffices to show that I' satisfies all non-goal rules of Π and that no goal-rule of Π applies in I'. Both proofs are very similar, so we concentrate on the former. Take a rule $R_1(\mathbf{y}_1)\wedge\cdots\wedge R_k(\mathbf{y}_k)\rightarrow S_1(\mathbf{z}_1)\vee\cdots\vee S_m(\mathbf{z}_m)$ from Π and assume to the contrary of what is to be shown that $R_1(\mathbf{a}_1),\ldots,R_k(\mathbf{a}_k)\in I'$, but none of the atoms in the rule head is satisfied. For $i\in\{1,\ldots,k\}$ let q_i be the restriction of J to the facts of the form $R(\mathbf{a}_i)$, $R\in\mathbf{S}'$, viewed as relational atoms of a CQ (consistently substitute the constants with variables, maintaining consistency also across queries q_i). It is not hard to verify that whenever $\widehat{J}$ is a proper $\mathbf{S}'$-instance and $\widehat{J}\models q_i[\mathbf{b}_i]$ for $1\le i\le k$, then the $\mathbf{S}\cup\mathbf{Q}$-instance $\widehat{I}$ in $\widehat{J}$ contains $R_1(\mathbf{b}_1),\ldots,R_k(\mathbf{b}_k)$, but again none of the atoms in the rule head is satisfied; in fact, $R(\mathbf{a}_i)\in I'$ implies $R(\mathbf{b}_i)\in\widehat{I}$ for all $R\in\mathbf{S}$ and $1\le i\le k$, and $R(\mathbf{b}_i)\in I'$ iff $R(\mathbf{b}_i)\in\widehat{I}$ for all $R\in\mathbf{Q}$, $1\le i\le k$, and permutations $\mathbf{b}_i$ of $\mathbf{a}_i$. Consequently, $\widehat{I}$ also violates the selected rule in Π and thus the CQ obtained from $\bigwedge_{1\le i\le k} q_i$ by existentially quantifying all variables is a forbidden CQ. In fact, it satisfies the two conditions under Point 1 of the definition of q_Π and thus the application of this query in J yields a contradiction to J not satisfying any of the CQs in q_Π. ❏

► **Lemma 38.** *$CQA(C_\Pi, q_\Pi)\preceq_{\mathsf{FO}}\Pi$.*

Proof. Let I be an $\mathbf{S}'$-instance. Let $T^{\downarrow}(I)$ be the $\mathbf{S}$-instance obtained from the reduct of I to schema $\mathbf{S}$ by the following sequence of operations:

1. drop each fact $R(\mathbf{a})$ such that for every permutation $\mathbf{b}$ of $\mathbf{a}$, there are distinct Q-types t and t' such that neither $Q_t(\mathbf{b})$ nor $Q_{t'}(\mathbf{b})$ are in I;
2. add $S_t(\mathbf{a})$ for each $\mathbf{S}$-guarded tuple $\mathbf{a}\in\mathsf{adom}(I)^m$ that is assigned $\mathbf{Q}$-type t in I (the assignment need not be unique).

We show that $I\models_{C_\Pi} q_\Pi$ iff $T^{\downarrow}(I)\models\Pi$.

"if". Assume that $I\not\models_{C_\Pi} q_\Pi$. Then there is a minimal repair I' of I such that $I'\not\models q_\Pi$. Let I'' be obtained from I' by dopping all facts that are removed during Step 1 of the construction of $T^{\downarrow}(I)$ from I. Clearly, I'' satisfies all constraints from C_Π and we have $I''\not\models q_\Pi$. By construction of I'' and since I'' satisfies all constraints in C_Π, every $\mathbf{a}\in\mathsf{adom}(I'')^m$ is assigned a $\mathbf{Q}$-type $t_{\mathbf{a}}$ in I''. Since none of the CQs under Point 2 of the definition of q_Π applies in I'', this $\mathbf{Q}$-type is unique (and thus I'' is a proper $\mathbf{S}'$-instance); moreover, if a $\mathbf{Q}$-type is assigned to $\mathbf{a}$ in I, then this type is (unique and) $t_{\mathbf{a}}$. Define an extension J of $T^{\downarrow}(I)$ to schema $\mathbf{S}\cup\mathbf{Q}$ by adding $R(a_{i_1},\ldots,a_{i_m})$ whenever $R(x_{i_1},\ldots,x_{i_m})\in t_{\mathbf{a}}$ and $\mathbf{a}=a_1\cdots a_m\in\mathsf{adom}(T^{\downarrow}(I))^m$. Note that J is the $\mathbf{S}\cup\mathbf{Q}$-instance in I'' extended with the facts added in Point 2 of the construction of $T^{\downarrow}(I)$. To establish that $T^{\downarrow}(I)\not\models\Pi$ as desired, it suffices to show that J satisfies all non-goal rules of Π and that no goal-rule of Π applies in J. Both proofs are very similar, so we concentrate on the former.

We start with non-goal rules $R_1(\mathbf{y}_1)\wedge\cdots\wedge R_k(\mathbf{y}_k)\rightarrow S_1(\mathbf{z}_1)\vee\cdots\vee S_m(\mathbf{z}_m)$ from Π that do not contain any of the S_t relations. Assume to the contrary of what is to be shown that $R_1(\mathbf{a}_1),\ldots,R_k(\mathbf{a}_k)\in J$, but none of the atoms in the rule head is satisfied. For $i\in\{1,\ldots,k\}$, let q_i be the restriction of I'' to the facts of the form $R(\mathbf{a}'_i)$, $R\in\mathbf{S}'$ and $\mathbf{a}'_i$ a permutation of $\mathbf{a}_i$, viewed as relational atoms of a CQ (consistently substitute the constants with variables, maintaining consistency also across queries q_i). It is not hard to verify that whenever $\widehat{J}$ is a proper $\mathbf{S}'$-instance and $\widehat{J}\models q_i[\mathbf{b}_i]$ for $1\le i\le k$, then the $\mathbf{S}\cup\mathbf{Q}$-instance $\widehat{I}$ in $\widehat{J}$ contains $R_1(\mathbf{b}_1),\ldots,R_k(\mathbf{b}_k)$, but again none of the atoms in the rule head is satisfied; in fact, $R(\mathbf{a}_i)\in J$ implies $R(\mathbf{b}_i)\in\widehat{I}$ for all $R\in\mathbf{S}$ that are not of the form S_t and $1\le i\le k$, and we have $R(\mathbf{b}'_i)\in J$ iff $R(\mathbf{b}'_i)\in\widehat{I}$ for all $R\in\mathbf{Q}$ and all permutations

$\mathbf{b}_i'$ of $\mathbf{b}_i$, $1 \le i \le k$. Consequently, $\widehat{I}$ also violates the selected rule in Π and thus the CQ obtained from $\bigwedge_{1 \le i \le k} q_i$ by existentially quantifying all variables is a forbidden CQ. In fact, it satisfies the two conditions under Point 1 of the definition of q_Π and thus the applicability of this query in I'' yields a contradiction to I'' not satisfying any of the CQs in q_Π.

Now for non-goal rules in Π that mention an S_t relation, which are of the form (i) $S_t(\mathbf{x}) \to R(x_{i_1}, \dots, x_{i_m})$ with $\mathbf{x} = x_1 \cdots x_m$, $R(x_{i_1}, \dots, x_{i_m}) \in t$ and (ii) $S_t(\mathbf{x}) \wedge R(\mathbf{y}) \to \bot$ with $R(\mathbf{y}) \notin t$, $R \in \mathbf{Q}$, and $\mathbf{y}$ a permutation of $\mathbf{x}$. For brevity, we concentrate on the latter. Assume that $S_t(\mathbf{a}), R(\mathbf{b}) \in J$ with $\mathbf{b}$ a permutation of $\mathbf{a}$. No matter whether $S_t(\mathbf{a})$ is already in I or was added during the construction of J, t is the (unique, as argued above) $\mathbf{Q}$-type assigned to $\mathbf{a}$ in I; in the case where $S_t(\mathbf{a})$ is already in I, assuming that $\mathbf{a}$ is assigned $\mathbf{Q}$-type t assigned to $\mathbf{a}$ in I in fact gives rise to a forbidden query satisfied by I'', which is a contradiction. As noted above, the $\mathbf{Q}$-type t assigned to $\mathbf{a}$ in I is actually $t_\mathbf{a}$, that is, the $\mathbf{Q}$-type assigned to $\mathbf{a}$ in I''. The construction of J thus yields a contradiciton to $R(\mathbf{b}) \in J$.

"only if". Assume that $T^\downarrow(I) \not\models \Pi$. Then there is an extension J of $T^\downarrow(I)$ to schema $\mathbf{S} \cup \mathbf{Q}$ such that all non-goal rules in Π are satisfied in J and no goal rule of Π applies in J. Let I' be the $\mathbf{S}'$-instance obtained from I by dropping all facts $Q_t(\mathbf{a})$ such that $\mathsf{tp}_J(\mathbf{a}) = t$. We show that I' is a minimal repair of $T^\downarrow(I)$ with $I' \not\models q_\Pi$. Regarding the former, I' is a minimal repair since, whenever I contains $Q_t(\mathbf{a})$ for all $\mathbf{Q}$-types t thus violating a constraint in C_Π, then exactly one of those facts is removed in I', the one with $t = \mathsf{tp}_J(\mathbf{a})$; also note that when I assigns $\mathbf{Q}$-type t to some $\mathbf{a} \in \mathsf{adom}(I)^m$ (and thus the constraint C_Π is not violated at $\mathbf{a}$ in I), then $S_t(\mathbf{a})$ was added in Step 2 of the construction of $T^\downarrow(I)$, and consequently $\mathsf{tp}_{T(I')}(\mathbf{a}) = t$, thus no tuple $Q_{t'}(\mathbf{a})$ is removed from I during the construction of I'. Regarding $I' \not\models q_\Pi$, we start with noting that the $\mathbf{S}'$-instance I'' obtained from I' by dropping all facts removed in Point 1 of the construction of $T^\downarrow(I)$ and adding all facts added in Point 2 of the construction is proper and that J is the $\mathbf{S} \cup \mathbf{Q}$-instance of I''. Since J satisfies all goal rules in Π and no body of a goal rule of Π applies in I', the CQs under Point 1 of the definition of q_Π do not apply in I''. By Point (c) of the construction of those CQs and construction of I'' from I', they also do not apply in I'. The CQs under Point 2 of the definition q_Π do not apply in I' by construction of I'. ❏

▶ **Lemma 39.** $CQA(C, q) \preceq_{\mathsf{FO}} \Pi_{C,q}$.

Proof. Let an $\mathbf{S}$-instance I be given. Define an $\mathbf{S}'$-instance $T^\uparrow(I)$ that consists of all facts $R_{\mathsf{tp}_I(\mathbf{a})}(\mathbf{a})$ with $\mathbf{a} \in \mathsf{adom}(I)^i$, $i \le m$. Clearly, $T^\uparrow$ is FO-definable. We show that $I \models_C q$ iff $T^\uparrow(I) \models \Pi_{C,q}$.

"if". First assume $I \not\models_C q$. Then there is a minimal repair I' of I such that $I' \not\models q$. Let $\mathbf{Q}$ be the IDB relations of $\Pi_{C,q}$ and extend $T^\uparrow(I)$ to an $\mathbf{S}' \cup \mathbf{Q}$-instance J as follows: if $\mathsf{tp}_{I'}(\mathbf{a}) = t$ for a guarded tuple $\mathbf{a} \in \mathsf{dom}(I')^i$, $i \le m$, then include $Q_t(\mathbf{a})$ as well as $R(\mathbf{a})$ for all $R(x_1, \dots, x_i) \in t$. One can check that J satisfies all non-goal rules in $\Pi_{C,q}$ and that the goal rules under Point 3 of the definition of $\Pi_{C,q}$ do not apply. The goal rules under Point 4 do not apply since $I' \not\models q$ and by construction of J. Thus $T^\uparrow(I) \not\models \Pi_{C,q}$.

"only if". Now assume that $T^\uparrow(I) \not\models \Pi_{C,q}$ and let J be a minimal extension of $T^\uparrow(I)$ (w.r.t. set inclusion) that satisfies all rules in $\Pi_{C,q}$ and where no body of a goal rule applies. Define I' as the restriction of J to $\mathbf{S}$. Then I' is a minimal repair of I. In fact, it suffices to show that for each guarded set $\mathbf{a} \in \mathsf{dom}(I')$, $i \le m$, the set of facts in I' of the form $R(\mathbf{a})$ is a minimal repair of the corresponding restriction of I. This is a consequence of the first two types of rules in $\Pi_{C,q}$ being satisfied in J, the minimality of J, and the body of the first type of goal rule not applying in J. It thus remains to argue that $I' \not\models q$, which is an immediate consequence of the second type of goal rule not applying in J. ❏

▶ **Lemma 40.** $\Pi_{C,q} \preceq_{\mathsf{FO}} CQA(C, q)$.

Proof. Let an $\mathbf{S}'$-instance I be given. If $R_t(\mathbf{a}), R_{t'}(\mathbf{a}) \in I$ for any $\mathbf{a} \in \mathsf{dom}(I)^i$, $i \leq m$, and $t \neq t'$, then answer 'inconsistent'. Otherwise, replace each fact $R_t(\mathbf{a})$ with the set of facts $\{R(\mathbf{a}) \mid R(\mathbf{x}) \in t\}$, and call the result $T^{\downarrow}(I)$. Clearly, $T^{\downarrow}$ is FO-definable. We show that $I \models \Pi_{C,q}$ iff $T^{\downarrow}(I) \models_C q$.

"if". Assume $I \not\models \Pi_{C,q}$ and let I' be an extension of I to the IDB relations that satisfies all non-goal rules in $\Pi_{C,q}$ and in which no body of a goal rule applies; we may further assume w.l.o.g. that I' is minimal (w.r.t. set inclusion) with this property. Let J be the restriction of I' to $\mathbf{S}$. Then J is a repair of $T^{\downarrow}(I)$. In fact, it suffices to show that for each guarded set $\mathbf{a} \in \mathsf{dom}(J)$, $i \leq m$, the set of facts in J of the form $R(\mathbf{a})$ is a minimal repair of the corresponding restriction of $T^{\downarrow}(I)$. Since for each $\mathbf{a} \in \mathsf{dom}(I')^i$, $i \leq m$, I contains is at most one fact of the form $R_t(\mathbf{a})$ and due to the minimality of I' and the fact that I' satisfies the non-goal rules under Points 1 and 2, we have that whenever $R_t(\mathbf{a}) \in I$, then the set of facts in I' of the form $R(\mathbf{a})$ is a minimal repair of the corresponding restriction of $T^{\downarrow}(I)$. By construction of J, the same is true for J as desired.

"only if". Now assume $T^{\downarrow}(I) \not\models_C q$. Then there is a minimal repair J of $T^{\downarrow}(I)$ such that $J \not\models q$. Extend J to an instance I' by adding the following IDB facts: for each $\mathbf{a} \in \mathsf{dom}(J)^i$, $i \leq m$ with $\mathsf{tp}_J(\mathbf{a}) = t$, add $Q_t(\mathbf{a})$ as well as $R(\mathbf{a})$ for all $R(\mathbf{x}) \in t$. One can check that I' satisfies all non-goal rules in $\Pi_{C,q}$ and that no body of a goal rule applies, thus $I \not\models \Pi_{C,q}$. ❏