Union and Intersection of Schema Mappings

Jorge Pérez[1], Reinhard Pichler[2], Emanuel Sallinger[2], and Vadim Savenkov[2]

[1] Department of Computer Science, Universidad de Chile
[2] Faculty of Informatics, Vienna University of Technology

Abstract. Schema mappings have been extensively studied in database research over the past decade – notably in the areas of data exchange and data integration. Recently, the notion of an information transfer order on schema mappings has been introduced to compare the amount of source information that is actually transferred by two mappings. In this paper, we present two new operators: the *union* and *intersection* of mappings. The union of two mappings allows us to describe the sum of all information transferred by several mappings. The intersection refers to the common part of information transferred by several mappings. As one of our main results we prove that there exists a large class of mappings (containing the class of source-to-target tuple-generating dependencies) that forms a complete lattice with respect to these two operators.

1 Introduction

Schema mappings allow us to describe the relationship between two schemas. As such, schema mappings have been extensively studied in Data Exchange [9] and Data Integration [12]. Bernstein and Melnik [5, 6, 15] have proposed several fundamental operators on schema mappings, with composition [14, 7, 16, 2] and inverse [8, 10, 4, 3] being the most prominent ones. Recently, new concepts have been introduced [1, 10] to compare schema mappings in terms of the amount of source information transferred by the mappings. In this work, we present two new operators on mappings, which we consider as fundamental for studying the information transferred by several mappings. We thus introduce the *union* and *intersection* of mappings. The union of two mappings allows us to describe the sum of all information transferred by the mappings. The intersection of two mappings refers to the common part of information transferred by the mappings.

Before providing more details on our new schema mapping operators, we recall the notion of *information transfer* introduced by Arenas et al. [1]. Intuitively, the authors define a criterion to compare the amount of information that two mappings transfer from source to target. As an example consider the following two mappings given by source-to-target tuple-generating dependencies (st-tgds):

$$\mathcal{M}_1 : A(x, y, z) \to \exists u\, P(x, u)$$
$$\mathcal{M}_2 : A(x, y, z) \to S(x, y)$$

Intuitively, $\mathcal{M}_2$ *transfers more information* than $\mathcal{M}_1$ since the first and second component of tuples in A are being transferred to the target under $\mathcal{M}_2$, while only the first component is being transferred under $\mathcal{M}_1$ [1]. This intuition was formalized in [1] and several algorithmic properties were studied. In particular the authors show that given two mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ specified by st-tgds, it can be decided whether $\mathcal{M}_2$ transfers more information than $\mathcal{M}_1$ [1]. A similar notion of information transfer was proposed by Fagin et al. [10], and it has been shown that both notions coincide for the important case of mappings specified by st-tgds [1, 10].

A possible application of the information transfer notion is in automatic mapping-generation tools [11]. As described in [10], if two possible mappings are automatically generated by different tools, then a plausible criterion to decide which mapping is the better to be used, is to choose the mapping that transfers more information from source to target. But what happens if both tools generate *incomparable mappings* in terms of information transfer? Then the criterion presented in [1, 10] can no longer be used to decide which mapping to choose. This is one of the questions that motivate our research.

To illustrate our new schema mapping operators of *intersection* and *union*, consider the following mapping with the same source schema as $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$\mathcal{M}_3 : A(x, y, z) \to T(x, z)$$

It can be shown that $\mathcal{M}_2$ and $\mathcal{M}_3$ are incomparable with respect to the information that both mappings transfer from source to target. Assume now that $\mathcal{M}_2$ and $\mathcal{M}_3$ are mappings that have been generated independently by two different tools. Since $\mathcal{M}_2$ and $\mathcal{M}_3$ are incomparable in terms of the information transfer from the source, a conservative approach would be to *synthesize* from both mappings a new mapping $\mathcal{M}'$ that only transfers the shared source information that is being mapped by both $\mathcal{M}_2$ and $\mathcal{M}_3$. Since $\mathcal{M}_2$ is transferring the first and second component of relation A, and $\mathcal{M}_3$ is transferring the first and third component of relation A, then $\mathcal{M}'$ can be defined as:

$$\mathcal{M}' : A(x, y, z) \to R(x)$$

that is, $\mathcal{M}'$ only maps the first component of A. In our framework, mapping $\mathcal{M}'$ is the *intersection* of $\mathcal{M}_2$ and $\mathcal{M}_3$. Formally, the intersection of $\mathcal{M}_2$ and $\mathcal{M}_3$ is a new mapping $\mathcal{N}$ that transfers less information than $\mathcal{M}_2$ and $\mathcal{M}_3$ and such that any other mapping that transfer less information than $\mathcal{M}_2$ and $\mathcal{M}_3$ also transfers less information than $\mathcal{N}$. That is, the intersection is the *greatest lower bound* with respect to the information transfer order. It can be shown that the intersection is not unique, in fact, both $\mathcal{M}'$ and $\mathcal{M}_1$ are possible intersections of $\mathcal{M}_2$ and $\mathcal{M}_3$. We formalize this notion and study several of its properties. In the above example, computing the intersection was an easy task, but we show that in general, intersecting mappings is not trivial. In fact, we prove that even for mappings specified by st-tgds, the intersection may not be expressible in First-Order logic (FO). On the other hand, we prove that Existential Second-Order logic (ESO) suffices to express the intersection of such mappings.

The dual operator is the *union* of schema mappings. Intuitively, the union of two mappings is a new mapping that transfers the *sum* of all the information transferred by both initial mappings. In our example above, the union of $\mathcal{M}_2$ and $\mathcal{M}_3$ is the mapping

$$\mathcal{M}' : A(x, y, z) \to S(x, y) \wedge T(x, z)$$

Formally, the union of $\mathcal{M}_2$ and $\mathcal{M}_3$ is a new mapping $\mathcal{N}$ that transfers more information than $\mathcal{M}_2$ and $\mathcal{M}_3$ and such that any other mapping that transfers more information than $\mathcal{M}_2$ and $\mathcal{M}_3$ also transfers more information than $\mathcal{N}$. That is, the union is the *least upper bound* with respect to the information transfer order. As for the case of the intersection, dealing with union is not always trivial. For example, one might be tempted to state that the following mapping $\mathcal{M}''$ is also a union of $\mathcal{M}_2$ and $\mathcal{M}_3$:

$$\mathcal{M}'' : A(x, y, z) \to R(x, y, z)$$

but it can be shown that $\mathcal{M}''$ is strictly more informative than $\mathcal{M}'$ and thus does not define the union for $\mathcal{M}_2$ and $\mathcal{M}_3$. However, if we are given the functional dependencies $\{A[1] \longrightarrow A[2], A[1] \longrightarrow A[3]\}$ over the source schema, then $\mathcal{M}''$ becomes the union of $\mathcal{M}_2$ and $\mathcal{M}_3$. We show that, in the absence of source constraints, the union is considerably easier to handle compared with the intersection. In particular, it can be shown that given mappings specified by a set of st-tgds their union can also be specified by a set of st-tgds.

Organisation of the paper and summary of results. In Section 2, we recall some basic notions and results. A conclusion is given in Section 6. The main results of the paper are detailed in Sections 3 – 5:

• *New operators.* In Section 3, we introduce the union and intersection operators of schema mappings and state our main results, namely: for a large class of mappings (containing the class of st-tgds) the union and intersection always exist. More precisely, this class of mappings is the class Rec of all mappings that have a *maximum recovery* [4] (we recall the definition of maximum recovery in Section 2). Our new operators allow us to define a lattice of the mappings in Rec w.r.t. to the information transfer order, s.t. the union (resp. intersection) corresponds to the least upper bound (resp. greatest lower bound) of two mappings.

• *Existence of the union.* In Section 4, we show for the class Rec that the union of two mappings always exists. The proof is constructive in that we describe how to obtain the union. For mappings defined by a set of st-tgds we show that the union can also be expressed by st-tgds. This allows us to prove an NEXPTIME upper bound for checking if some mapping is the union of two other mappings for the case of st-tgds.

• *Existence of the intersection.* In Section 5, we show several fundamental results on the existence of the intersection. First, for two mappings from the class Rec, the intersection always exists. However, in general, even for the restricted case of st-tgds, this intersection is not expressible in First-Order logic (FO). On the other hand, in Existential Second-Order logic (ESO) it is always possible to express the intersection of mappings defined by st-tgds.

Due to lack of space, proofs are only sketched. Details are given in the appendix.

2 Preliminaries

2.1 Schemas and schema mappings

A *schema* $\mathbf{S}$ is a finite set $\{R_1, \dots, R_k\}$ of relation symbols, with each R_i having a fixed arity $n_i \geq 0$. Let $\mathbf{D}$ be a countably infinite domain. An *instance* I of $\mathbf{S}$ assigns to each relation symbol R_i of $\mathbf{S}$ a finite relation $R_i^I \subseteq \mathbf{D}^{n_i}$. $\mathrm{Inst}(\mathbf{S})$ denotes the set of all instances of $\mathbf{S}$. We denote by $\mathrm{dom}(I)$ the set of all elements that occur in any of the relations R_i^I. We say that $R_i(t)$ is a *fact* of I if $t \in R_i^I$. We sometimes denote an instance by its set of facts.

Given schemas $\mathbf{S}$ and $\mathbf{T}$, a *schema mapping* (or just *mapping*) from $\mathbf{S}$ to $\mathbf{T}$ is a subset of $\mathrm{Inst}(\mathbf{S}) \times \mathrm{Inst}(\mathbf{T})$. We say that J is a *solution for I under $\mathcal{M}$* whenever $(I, J) \in \mathcal{M}$. The set of all solutions for I under $\mathcal{M}$ is denoted by $\mathrm{Sol}_{\mathcal{M}}(I)$. For a mapping $\mathcal{M}$ from $\mathbf{S}$ to $\mathbf{T}$, we denote by $\mathrm{dom}(\mathcal{M})$ the set of all instances $I \in \mathrm{Inst}(\mathbf{S})$ such that $\mathrm{Sol}_{\mathcal{M}}(I) \neq \emptyset$. Moreover, $\mathcal{M}$ is said to be *total* if $\mathrm{dom}(\mathcal{M}) = \mathrm{Inst}(\mathbf{S})$.

Notice that mappings are binary relations, and thus one can define the composition of mappings as for the composition of binary relations [15, 7]. Let $\mathcal{M}_{12}$ be a mapping from schema $\mathbf{S}_1$ to schema $\mathbf{S}_2$ and $\mathcal{M}_{23}$ a mapping from $\mathbf{S}_2$ to schema $\mathbf{S}_3$. Then $\mathcal{M}_{12} \circ \mathcal{M}_{23}$ is a mapping from $\mathbf{S}_1$ to $\mathbf{S}_3$ given by the set $\{(I, J) \in \mathrm{Inst}(\mathbf{S}_1) \times \mathrm{Inst}(\mathbf{S}_3) \mid \text{there exists } K \text{ such that } (I, K) \in \mathcal{M}_{12} \text{ and } (K, J) \in \mathcal{M}_{23}\}$ [15, 7].

2.2 Dependencies and definability of mappings

Given disjoint schemas $\mathbf{S}$ and $\mathbf{T}$, a *source-to-target tuple-generating dependency* (st-tgd) from $\mathbf{S}$ to $\mathbf{T}$ is a sentence of the form $\forall \bar{x} \forall \bar{y}(\varphi(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\ \psi(\bar{x}, \bar{z}))$, where $\varphi(\bar{x}, \bar{y})$ is a conjunctive query (CQ) over $\mathbf{S}$, and $\psi(\bar{x}, \bar{z})$ is a CQ over $\mathbf{T}$. The left-hand side of the implication in an st-tgd is called the *premise*, and the right-hand side the *conclusion*. A *full st-tgd* is an st-tgd with no existentially quantified variables in its conclusion. We usually omit the universal quantifiers when writing st-tgds. Suppose that we are given a set Σ of logical formulas over the schemas $\mathbf{S}$ and $\mathbf{T}$, e.g., a set of st-tgds from $\mathbf{S}$ and $\mathbf{T}$, a set of First-Order formulas or of Existential Second-Order formulas over the schemas $\mathbf{S}$ and $\mathbf{T}$. We say that a mapping $\mathcal{M}$ is *specified* by Σ, denoted by $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$, if for every $(I, J) \in \mathrm{Inst}(\mathbf{S}) \times \mathrm{Inst}(\mathbf{T})$, we have $(I, J) \in \mathcal{M}$ if and only if (I, J) satisfies Σ.

As is customary in the data exchange literature, we assume the existence of two disjoint sets of elements: constant values $\mathbf{C}$ and null values $\mathbf{N}$. Thus, for a mapping defined by st-tgds, we assume that source instances are constructed by using only elements from $\mathbf{C}$, while target instances are constructed by using elements from $\mathbf{C} \cup \mathbf{N}$.

2.3 Maximum recovery

The notion of maximum recovery proposed in [4] is fundamental to our study. It provides a natural notion for inverting schema mappings. In [4] the authors first define recoveries of schema mappings and then restrict them to maximum recoveries. Given a mapping $\mathcal{M}$ from $\mathbf{S}_1$ to $\mathbf{S}_2$, we say that $\mathcal{M}'$ from $\mathbf{S}_2$ to $\mathbf{S}_1$ is a *recovery* of $\mathcal{M}$ if for every instance I in $\mathbf{S}_1$, it holds that $(I, I) \in \mathcal{M} \circ \mathcal{M}'$.

In symbols, $\mathcal{M}'$ is a recovery of $\mathcal{M}$ if $\mathrm{Id} \subseteq \mathcal{M} \circ \mathcal{M}'$, where Id is the *identity mapping* $\{(I, I) \mid I \in \mathrm{Inst}(\mathbf{S}_1)\}$. Moreover, $\mathcal{M}'$ is said to be a *maximum recovery* of $\mathcal{M}$ if for every other recovery $\mathcal{M}''$ of $\mathcal{M}$, it holds that $\mathcal{M} \circ \mathcal{M}' \subseteq \mathcal{M} \circ \mathcal{M}''$. Intuitively, $\mathcal{M}'$ is a maximum recovery of $\mathcal{M}$ if $\mathcal{M} \circ \mathcal{M}'$ is as close as possible to the identity mapping Id. We write REC to denote the class of all total mappings that admit a maximum recovery.

2.4 Information Transfer

In [1] a notion of *information transfer* for schema mappings was defined to compare the amount of information that two mappings transfer from source to target. Formally, given mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ with the same source schema $\mathbf{S}$, mapping $\mathcal{M}_1$ *transfers at least as much source information as* $\mathcal{M}_2$, denoted by $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$ if there exists a mapping $\mathcal{N}$ such that $\mathcal{M}_1 \circ \mathcal{N} = \mathcal{M}_2$ [1]. That is, $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$ if $\mathcal{M}_2$ can be constructed from $\mathcal{M}_1$ via mapping composition. Notice that $\preceq_{inf}$ is a *pre-order*, i.e., $\preceq_{inf}$ is a reflexive and transitive, but not antisymmetric relation. Thus, we say that $\mathcal{M}_1$ and $\mathcal{M}_2$ transfer the same information from the source, denoted by $\mathcal{M}_1 \equiv_{inf} \mathcal{M}_2$ if $\mathcal{M}_1 \preceq_{inf} \mathcal{M}_2$ and $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$. By slight abuse of notation we consider $\preceq_{inf}$ as an order (rather than a pre-order) by identifying a mapping $\mathcal{M}$ with the equivalence class of all mappings $\equiv_{inf}$-equivalent with $\mathcal{M}$.

3 The union and intersection operators

Below, we make use of $\preceq_{inf}$ to define the union and the intersection of two mappings. Intuitively the union is a mapping that transfers the *sum* of all the information transferred by the two initial mappings. Analogously, we define the intersection of two mappings as a mapping that transfers only information which is transferred *by each* of the initial mappings.

Definition 1. *Let $\mathcal{C}$ be a class of mappings and $\mathcal{M}_1$ and $\mathcal{M}_2$ two mappings in $\mathcal{C}$ with the same source schema. The* union *of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ such that:*

1. *$\mathcal{M}_1 \preceq_{inf} \mathcal{M}$,*
2. *$\mathcal{M}_2 \preceq_{inf} \mathcal{M}$, and*
3. *if $\mathcal{N}$ is a mapping in $\mathcal{C}$ with $\mathcal{M}_1 \preceq_{inf} \mathcal{N}$ and $\mathcal{M}_2 \preceq_{inf} \mathcal{N}$, then $\mathcal{M} \preceq_{inf} \mathcal{N}$.*

The union of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$ is denoted by $\mathcal{M}_1 \sqcup_{\mathcal{C}} \mathcal{M}_2$.

Definition 2. *Let $\mathcal{C}$ be a class of mappings and $\mathcal{M}_1$ and $\mathcal{M}_2$ two mappings in $\mathcal{C}$ with the same source schema. The* intersection *of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\mathcal{C}$, is a mapping $\mathcal{M} \in \mathcal{C}$ such that:*

1. *$\mathcal{M} \preceq_{inf} \mathcal{M}_1$,*
2. *$\mathcal{M} \preceq_{inf} \mathcal{M}_2$, and*
3. *if $\mathcal{N}$ is a mapping in $\mathcal{C}$ with $\mathcal{N} \preceq_{inf} \mathcal{M}_1$ and $\mathcal{N} \preceq_{inf} \mathcal{M}_2$, then $\mathcal{N} \preceq_{inf} \mathcal{M}$.*

The intersection of $\mathcal{M}_1$ and $\mathcal{M}_2$ is denoted by $\mathcal{M}_1 \sqcap_{\mathcal{C}} \mathcal{M}_2$.

When the class $\mathcal{C}$ is clear from the context, we just write $\mathcal{M}_1 \sqcup \mathcal{M}_2$ (resp. $\mathcal{M}_1 \sqcap \mathcal{M}_2$) for the union (resp. for the intersection) of two mappings. Notice that the definition of the union of mappings is just the *least upper bound* (supremum) of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\preceq_{inf}$ (inside the class of mappings $\mathcal{C}$). Analogously, the intersection of mappings is just the *greatest lower bound* (infimum) of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. $\preceq_{inf}$ (inside the class $\mathcal{C}$). Also notice that the union and the intersection as defined above are unique up to the equivalence relation $\equiv_{inf}$. This is why we speak of *the* union resp. intersection of two mappings.

Notice that with the definition of union and intersection based on the order $\preceq_{inf}$ it is by no means evident that for any two mappings the union or intersection always exists. Thus, a first important question that needs to be answered for these operators is for which classes of mappings the existence of the intersection or the union is guaranteed. As we will show next, the class Rec of mappings having a maximum recovery will play a fundamental role in determining the existence of the union and intersection.

Beside existence, there are two other important questions regarding these operators that need to be addressed. One is the question of expressiveness: what is the mapping language needed to express the union/intersection when it exists? Another main question is about computing these operators: is there an algorithm to compute the union/intersection? One of our main results is the following general result that gives a positive answer to the existence question.

Theorem 1. *There exists a class $\mathcal{R}$ of mappings (that contains the class of mappings specified by st-tgds), such that for every pair of mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ in $\mathcal{R}$ the union $\mathcal{M}_1 \sqcup_{\mathcal{R}} \mathcal{M}_2$ and the intersection $\mathcal{M}_1 \sqcap_{\mathcal{R}} \mathcal{M}_2$ always exist.*

We will show that Rec is such a class $\mathcal{R}$ that satisfies the statement of Theorem 1. By using notions of lattice theory, Theorem 1 can be restated as follows. Recall that given an order relation $\leq$, a lattice is a structure $\langle \mathcal{A}, \leq \rangle$ such that every two elements $X, Y \in \mathcal{A}$ have a least upper bound (supremum) and a greatest lower bound (infimum) in $\mathcal{A}$. Then Theorem 1 can be formulated as follows.

Theorem 2. *Let $\mathbf{S}$ be a relational schema. There exists a class $\mathcal{R}_{\mathbf{S}}$ that contains the class of all mappings specified by st-tgds having $\mathbf{S}$ as source schema, such that $\langle \mathcal{R}_{\mathbf{S}}, \preceq_{inf} \rangle$ is a lattice (up to $\equiv_{inf}$-equivalence).*

Theorem 1 is proved by combining the results in the following sections for union and intersection. Theorem 2 is an immediate consequence of Theorem 1 given the following proposition:

Proposition 1. *The union and intersection of mappings are invariant under $\equiv_{inf}$-equivalence. Formally, let $\mathcal{M}_1, \mathcal{M}'_1, \mathcal{M}_2$, and $\mathcal{M}'_2$ be mappings from some class $\mathcal{C}$ with $\mathcal{M}_1 \equiv_{inf} \mathcal{M}'_1$ and $\mathcal{M}_2 \equiv_{inf} \mathcal{M}'_2$. Then the following relations hold:*

(1) If $\mathcal{M}_1 \sqcup_{\mathcal{C}} \mathcal{M}_2$ exists, then $\mathcal{M}'_1 \sqcup_{\mathcal{C}} \mathcal{M}'_2$ exists as well and the equivalence $\mathcal{M}_1 \sqcup_{\mathcal{C}} \mathcal{M}_2 \equiv_{inf} \mathcal{M}'_1 \sqcup_{\mathcal{C}} \mathcal{M}'_2$ holds.

(2) If $\mathcal{M}_1 \sqcap_{\mathcal{C}} \mathcal{M}_2$ exists, then $\mathcal{M}'_1 \sqcap_{\mathcal{C}} \mathcal{M}'_2$ exists as well and the equivalence $\mathcal{M}_1 \sqcap_{\mathcal{C}} \mathcal{M}_2 \equiv_{inf} \mathcal{M}'_1 \sqcap_{\mathcal{C}} \mathcal{M}'_2$ holds.

Intuitively, Proposition 1 states that union and intersection of mappings are preserved under $\equiv_{inf}$-equivalence. The proposition follows immediately from the definition of $\sqcup_{\mathcal{C}}$ and $\sqcap_{\mathcal{C}}$.

4 Existence of the union

In this section we propose a straightforward method to compute the union of mappings specified by st-tgds (w.r.t. the class of mappings specified by st-tgds). This method will allow us to provide a positive answer to all of the questions concerning the existence, expressiveness, and computation of the union for this class of mappings. In contrast, we will show in Section 5 that dealing with the intersection operator is considerably more difficult.

The procedure to compute the union is very simple. Let $\mathcal{M}_1 = (\mathbf{S}, \mathbf{T}_1, \Sigma_1)$ and $\mathcal{M}_2 = (\mathbf{S}, \mathbf{T}_2, \Sigma_2)$ be two mappings specified by st-tgds. Let $\widehat{\mathbf{T}}_2$ be a copy of $\mathbf{T}_2$ such that $\widehat{\mathbf{T}}_2$ is disjoint with $\mathbf{T}_1$, and let $\widehat{\Sigma}_2$ be the set of dependencies that results from Σ_2 by replacing every relation name in $\mathbf{T}_2$ by its copy in $\widehat{\mathbf{T}}_2$. Consider the mapping $\mathcal{M}' = (\mathbf{S}, \mathbf{T}_1 \cup \widehat{\mathbf{T}}_2, \Sigma_1 \cup \widehat{\Sigma}_2)$. Then $\mathcal{M}'$ is the union of $\mathcal{M}_1$ and $\mathcal{M}_2$. From this we obtain the following result.

Proposition 2. *Let $\mathcal{S}$ be the class of mappings specified by st-tgds, and $\mathcal{M}_1 = (\mathbf{S}, \mathbf{T}_1, \Sigma_1)$ and $\mathcal{M}_2 = (\mathbf{S}, \mathbf{T}_2, \Sigma_2)$ be mappings such that Σ_1 and Σ_2 are sets of st-tgds. Then the union $\mathcal{M}_1 \sqcup_{\mathcal{S}} \mathcal{M}_2$ always exists. Moreover, there exists an algorithm which, given $\mathcal{M}_1$ and $\mathcal{M}_2$, computes $\mathcal{M}_1 \sqcup_{\mathcal{S}} \mathcal{M}_2$ in polynomial time.*

Proposition 2 follows from a more general result on the existence of the union for a class of mappings that properly contains the class of mappings specified by st-tgds. Recall from Section 2.3 that we write Rec to denote the class of all total mappings that admit a maximum recovery. It was shown in [4] that every mapping specified by st-tgds is total and has a maximum recovery, and thus Rec contains the class of all mappings specified by st-tgds. The following is the general result for the union operator w.r.t. the class Rec.

Proposition 3. *For every pair of mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ in* Rec *having the same source schema, the union $\mathcal{M}_1 \sqcup_{\text{Rec}} \mathcal{M}_2$ exists.*

Proof (sketch). Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be disjoint schemas, $\mathcal{M}_1$ a mapping in Rec from $\mathbf{S}$ to $\mathbf{T}_1$, and $\mathcal{M}_2$ a mapping in Rec from $\mathbf{S}$ to $\mathbf{T}_2$. Consider the mapping $\mathcal{M}_1 \oplus \mathcal{M}_2$ from $\mathbf{S}$ to $\mathbf{T}_1 \cup \mathbf{T}_2$ defined as follows:

$$\mathcal{M}_1 \oplus \mathcal{M}_2 = \{(I, J_1 \cup J_2) \mid (I, J_1) \in \mathcal{M}_1 \text{ and } (I, J_2) \in \mathcal{M}_2\}.$$

It can be shown that $\mathcal{M}_1 \oplus \mathcal{M}_2$ is the union of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. Rec. If $\mathbf{T}_1$ and $\mathbf{T}_2$ are not disjoint, one can always construct a copy $\widehat{\mathbf{T}}_2$ of $\mathbf{T}_2$ that is disjoint with $\mathbf{T}_1$, and a mapping $\widehat{\mathcal{M}}_2$ from $\mathbf{S}$ to $\widehat{\mathbf{T}}_2$ such that $\mathcal{M}_2 \equiv_{inf} \widehat{\mathcal{M}}_2$, and then $\mathcal{M}_1 \oplus \widehat{\mathcal{M}}_2$ is the desired union. □

The proof of Proposition 2 follows from the proof of Proposition 3 plus the fact that if $\mathcal{M}_1$ and $\mathcal{M}_2$ are specified by st-tgds, then $\mathcal{M}_1 \oplus \mathcal{M}_2$ can also be specified by a set of st-tgds. The following example shows that computing the union is extremely easy for the case of mappings specified by st-tgds.

Example 1. Let mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ be defined by the following sets of st-tgds: $\mathcal{M}_1 = \{S(x,y) \rightarrow T(x,y)\}$; $M_2 = \{S(x,y) \rightarrow T(x,y),\ Q(x) \rightarrow T(x,x)\}$. The union $\mathcal{M}_1 \sqcup_{\text{REC}} \mathcal{M}_2$ is simply the mapping $\mathcal{M}$ containing all three dependencies, with appropriately renamed target relation symbols:

$$\mathcal{M} = \{S(x,y) \rightarrow T(x,y),\ S(x,y) \rightarrow T'(x,y),\ Q(x) \rightarrow T'(x,x)\}$$

Proposition 2 also allows us to prove positive algorithmic results regarding the union of schema mappings. The following result follows directly from Proposition 2 and the results in [1] regarding the order $\preceq_{inf}$.

Proposition 4. *Given mappings $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ specified by st-tgds, it is decidable (in* NEXPTIME*) whether $\mathcal{M}_3$ is the union of $\mathcal{M}_1$ and $\mathcal{M}_2$.*

A legitimate question, of course, is if the characterization in the proof of Proposition 3 also works outside Rec. We have to leave this as an open question for future research. The following proposition shows that this is a tricky problem which may even require some adaptation of the $\preceq_{inf}$ relation.

Proposition 5. *There exists a mapping $\mathcal{M}$ such that $\mathcal{M} \prec_{inf} \mathcal{M} \oplus \mathcal{M}$ (that is $\mathcal{M} \preceq_{inf} \mathcal{M} \oplus \mathcal{M}$ and $\mathcal{M} \oplus \mathcal{M} \not\preceq_{inf} \mathcal{M}$).*

In other words, the $\preceq_{inf}$ order displays an unexpected behaviour for mappings outside Rec: intuitively, one would expect that the amount of source information transferred remains unchanged if, for every source instance I, we combine all pairs of solutions of I. For mappings in Rec, this is of course the case. In contrast, by Proposition 5, there are mappings outside Rec such that the amount of source information transferred strictly increases by this simple syntactic trick.

5 Existence of the intersection

In this section we study the existence of the intersection. The main result is stated in the following theorem. Again, we use Rec to denote the class of total mappings that have a maximum recovery.

Theorem 3. *For every pair of mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ in* Rec *having the same source schema, the intersection $\mathcal{M}_1 \sqcap_{\text{REC}} \mathcal{M}_2$ exists.*

Proof (sketch). To describe the proof of the theorem we need to introduce some technical notions. Let $\mathbf{S}$ be a schema and consider a mapping $\mathcal{M}$ from $\mathbf{S}$ to $\mathbf{S}$ (that is $\mathcal{M} \subseteq \text{Inst}(\mathbf{S}) \times \text{Inst}(\mathbf{S})$). For a positive integer k, we define $\mathcal{M}^k$ recursively as follows:

$$\mathcal{M}^1 = \mathcal{M},$$
$$\mathcal{M}^{k+1} = \mathcal{M} \circ \mathcal{M}^k.$$

We shall also define $\mathcal{M}^+$ as the following mapping from $\mathbf{S}$ to $\mathbf{S}$:

$$\mathcal{M}^+ = \bigcup_{i=1}^{\infty} \mathcal{M}^i.$$

Notice that $\mathcal{M}^+$ is the *transitive closure* of $\mathcal{M}$ when it is viewed as a binary relation over $\text{Inst}(\mathbf{S})$.

Now consider mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ in the statement of the theorem. Given that $\mathcal{M}_1$ and $\mathcal{M}_2$ are mappings in Rec, we know that there exist mappings $\mathcal{M}_1'$ and $\mathcal{M}_2'$ such that $\mathcal{M}_1'$ is a maximum recovery of $\mathcal{M}_1$, and $\mathcal{M}_2'$ is a maximum recovery of $\mathcal{M}_2$. Now consider the mapping $\mathcal{M}$ given by

$$\mathcal{M} \;=\; \Big((\mathcal{M}_1 \circ \mathcal{M}_1') \cup (\mathcal{M}_2 \circ \mathcal{M}_2')\Big)^+.$$

It can be shown that $\mathcal{M}$ is the intersection $\mathcal{M}_1 \sqcap_{\text{Rec}} \mathcal{M}_2$. □

Since every mapping given by st-tgds is total and has a maximum recovery [4], from Theorem 3 we obtain that for mappings specified by st-tgds the intersection (w.r.t. the class Rec) always exists.

Notice that Theorem 3 is only about existence and says nothing about the language needed to express the intersection of mappings specified by st-tgds. The following result shows that as opposed to the case of the union operator, the intersection of mappings specified by st-tgds may not be expressible in First-Order logic (FO).

Theorem 4. *There exist mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ specified by st-tgds such that $\mathcal{M}_1 \sqcap_{\text{Rec}} \mathcal{M}_2$ cannot be specified by a set of* FO *sentences.*

Proof (sketch). Consider the schemas $\mathbf{S} = \{A(\cdot,\cdot), B(\cdot,\cdot)\}$, $\mathbf{T}_1 = \{T_1(\cdot,\cdot)\}$, and $\mathbf{T}_2 = \{T_2(\cdot,\cdot)\}$. In the proof we use mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ from $\mathbf{S}$ to $\mathbf{T}_1$ and from $\mathbf{S}$ to $\mathbf{T}_2$, respectively, specified by the following st-tgds:

$$\begin{aligned} \mathcal{M}_1 :&\;\; \exists u\, (A(x,u) \wedge B(u,y)) \rightarrow T_1(x,y) \\ \mathcal{M}_2 :&\;\; \exists u\, (B(x,u) \wedge A(u,y)) \rightarrow T_2(x,y) \end{aligned}$$

It can be shown by an argument based on Ehrenfeucht-Fraïssé games that $\mathcal{M}_1 \sqcap_{\text{Rec}} \mathcal{M}_2$ cannot be expressed by an FO sentence. □

Notice that the above theorem states that $\mathcal{M}_1 \sqcap_{\text{Rec}} \mathcal{M}_2$ is not expressible in FO. In principle, it might be the case that if we restrict ourselves to the intersection with respect to a smaller class of mappings, for example the class of mappings specified by st-tgds, then we could obtain better expressibility results. The following proposition shows a negative result in this respect. This is a corollary of the proof of Theorem 4.

Proposition 6. *Let $\mathcal{S}$ be the class of mappings specified by st-tgds. Then there are mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ in $\mathcal{S}$ such that $\mathcal{M}_1 \sqcap_{\mathcal{S}} \mathcal{M}_2$ does not exist.*

This raises the question as to which language is expressive enough to specify the intersection of two mappings specified by st-tgds. Below we provide an answer to this question.

Theorem 5. *Given two schema mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ given by st-tgds, the intersection $\mathcal{M}_1 \sqcap_{\text{REC}} \mathcal{M}_2$ is expressible by an Existential Second-Order logic (ESO) formula.*

Proof (sketch). Let $\mathbf{S}$ be the source schema of $\mathcal{M}_1$ and $\mathcal{M}_2$, and $\widehat{\mathbf{S}}$ a copy of schema $\mathbf{S}$. Moreover, let $\mathcal{M}'_1$ and $\mathcal{M}'_2$ be respective maximum recoveries of $\mathcal{M}_1$ and $\mathcal{M}_2$. One can show (see Appendix I) that the composition $\mathcal{M}_1 \circ \mathcal{M}'_1$ can be expressed as an FO formula:

$$\forall \boldsymbol{x}_1(\varphi_1(\boldsymbol{x}_1) \rightarrow \psi_1(\boldsymbol{x}_1)) \wedge \ldots \wedge \forall \boldsymbol{x}_n(\varphi_n(\boldsymbol{x}_n) \rightarrow \psi_n(\boldsymbol{x}_n))$$

where $\varphi_i(\boldsymbol{x}_i)$ and $\psi_i(\boldsymbol{x}_i)$ are FO formulas over $\mathbf{S}$ and $\widehat{\mathbf{S}}$, respectively. Similarly, $\mathcal{M}_2 \circ \mathcal{M}'_2$ can be expressed as $\forall \boldsymbol{y}_1(\alpha_1(\boldsymbol{y}_1) \rightarrow \beta_1(\boldsymbol{y}_1)) \wedge \ldots \wedge \forall \boldsymbol{y}_m(\alpha_m(\boldsymbol{y}_m) \rightarrow \beta_m(\boldsymbol{y}_m))$. The formula representing the intersection is based on the construction in the proof of Theorem 3, thus we need to show how to express the transitive closure of $\mathcal{M}_1 \circ \mathcal{M}'_1 \cup \mathcal{M}_2 \circ \mathcal{M}'_2$. For this we use an intermediate schema $\widetilde{\mathbf{S}}$ constructed as follows: for every n-ary relation R of $\mathbf{S}$, we include an $(n+1)$-ary relation $\widetilde{R}$ in $\widetilde{\mathbf{S}}$. The idea is that an atom $\widetilde{R}(\boldsymbol{a}, g)$ will represent the atom $R(\boldsymbol{a})$ in the generation g of the computation of the transitive closure. Now, to define the intersection, we use an ESO formula of the form

$$\exists \widetilde{\mathbf{S}}\ \exists s\ \exists \text{zero}\ (\Omega_s \wedge \Omega_{\widetilde{R}} \wedge \Omega^E_{\widetilde{R}})$$

where $\exists \widetilde{\mathbf{S}}$ denotes an existential quantification over all relation symbols in $\widetilde{\mathbf{S}}$, s is a function symbol, and zero a first order variable. The rest of the formulas is constructed as follows: Ω_s is the formula $\forall x \forall y\big((s(x) = s(y) \rightarrow x = y) \wedge \neg(s(x) = x) \wedge \neg(s(x) = \text{zero})\big)$ that defines a successor function, with zero as the first element; $\Omega_{\widetilde{R}}$ corresponds to the following FO formula (we assume $\mathbf{S} = \{R_1, \ldots, R_k\}$ and $\boldsymbol{z}_i$ is a tuple of variables of the same arity as R_i):

$$\begin{aligned}&\big(\forall \boldsymbol{z}_1(R_1(\boldsymbol{z}_1) \rightarrow \widetilde{R}_1(\boldsymbol{z}_1, \text{zero})) \wedge \cdots \wedge \forall \boldsymbol{z}_k(R_k(\boldsymbol{z}_k) \rightarrow \widetilde{R}_k(\boldsymbol{z}_k, \text{zero}))\big) \wedge \\ &\forall g \Bigg(\bigwedge_{i=1}^{n} \forall \boldsymbol{x}_i \big[\widetilde{\varphi}_i(\boldsymbol{x}_i, g) \rightarrow \widetilde{\psi}_i(\boldsymbol{x}_i, s(g))\big] \ \vee\ \bigwedge_{i=1}^{m} \forall \boldsymbol{y}_i \big[\widetilde{\alpha}_i(\boldsymbol{y}_i, g) \rightarrow \widetilde{\beta}_i(\boldsymbol{y}_i, s(g))\big] \Bigg)\end{aligned}$$

where $\widetilde{\varphi}_i(\boldsymbol{x}_i, g)$ is obtained from $\varphi_i(\boldsymbol{x}_i)$ by replacing every relational symbol $R(\boldsymbol{z})$ by $\widetilde{R}(\boldsymbol{z}, g)$, and $\widetilde{\psi}_i(\boldsymbol{x}_i, s(g))$ is obtained from $\psi_i(\boldsymbol{x}_i)$ by replacing every relational symbol $\widehat{R}(\boldsymbol{z})$ by $\widetilde{R}(\boldsymbol{z}, s(g))$, and similarly for $\widetilde{\alpha}_i$ and $\widetilde{\beta}_i$. The intuition is that the first line initializes the relations $\widetilde{R}_i$ at generation 0, and the second line mimics a formula representing $(\mathcal{M}_1 \circ \mathcal{M}'_1 \cup \mathcal{M}_2 \circ \mathcal{M}'_2)^+$ over schema $\widetilde{\mathbf{S}}$. Finally, $\Omega^E_{\widetilde{R}}$ just extracts the target relations at some generation g of the transitive closure:

$$\exists g\big(\forall \boldsymbol{z}_1(\widetilde{R}_1(\boldsymbol{z}_1, g) \rightarrow \widehat{R}_1(\boldsymbol{z}_1))\ \wedge \cdots \wedge\ \forall \boldsymbol{z}_k(\widetilde{R}_k(\boldsymbol{z}_k, g) \rightarrow \widehat{R}_k(\boldsymbol{z}_k))\big). \qquad \square$$

Example 2. Recall the mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ from Example 1. Composed with their maximum recoveries, they have the following form (see [4]):

$$\mathcal{M}_1 \circ \mathcal{M}'_1 = \{S(x_1, x_2) \to \widehat{S}(x_1, x_2)\} \text{ and}$$
$$\mathcal{M}_2 \circ \mathcal{M}'_2 = \{S(x_1, x_2) \to \widehat{S}(x_1, x_2) \vee (x_1 = x_2 \wedge \widehat{Q}(x_1)),$$
$$Q(x_1) \to \widehat{S}(x_1, x_1) \vee \widehat{Q}(x_1)\}.$$

The intersection $\mathcal{M}_1 \sqcap_{\text{Rec}} \mathcal{M}_2$ is expressed by the following ESO formula:

$$\begin{aligned}
&\exists \widetilde{S} \exists \widetilde{Q} \exists s \exists \,\text{zero}\, \big(\forall x \forall y \big((s(x) = s(y) \to x = y) \wedge s(x) \neq x \wedge s(x) \neq \text{zero}\big) \wedge \\
&\qquad \forall x_1 \forall x_2 \big(S(x_1, x_2) \to \widetilde{S}(x_1, x_2, \text{zero})\big) \wedge \forall x \big(Q(x) \to \widetilde{Q}(x, \text{zero})\big) \wedge \\
&\quad \forall g \Big(\forall x_1 \forall x_2 \big(\widetilde{S}(x_1, x_2, g) \to \widetilde{S}(x_1, x_2, s(g))\big) \vee \\
&\qquad \big[\forall x_1 \forall x_2 \big(\widetilde{S}(x_1, x_2, g) \to \widetilde{S}(x_1, x_2, s(g)) \vee (x_1 = x_2 \wedge \widetilde{Q}(x_1, s(g)))\big) \wedge \\
&\qquad \forall x_1 \forall x_2 \big(\widetilde{Q}(x_1, g) \to \widetilde{S}(x_1, x_1, s(g)) \vee \widetilde{Q}(x_1, s(g))\big)\big] \Big) \wedge \\
&\quad \exists g' \, \big(\forall x_1 \forall x_2 (\widetilde{S}(x_1, x_2, g') \to \widehat{S}(x_1, x_2)) \wedge \forall x (\widetilde{Q}(x, g') \to \widehat{Q}(x))\big)\big)
\end{aligned}$$

6 Conclusion

In this work, we have introduced two new operators *union* and *intersection* on schema mappings. We have proved that these operators allow us to define a lattice w.r.t. the order $\preceq_{inf}$ (up to $\equiv_{inf}$-equivalence) for the mappings in Rec (i.e., mappings having a maximum recovery). In particular, we have shown that the union and intersection always exist for mappings in Rec. When restricting us to the simple case of mappings specified by st-tgds it has turned out that the intersection operator is considerably more difficult to handle than the union operator. More specifically, the union of two mappings specified by st-tgds can again be specified by a set of st-tgds. In contrast, First-Order logic (FO) does in general not suffice to express the intersection of such mappings.

A lot of interesting research questions have been left for future work. First of all, while our definitions of $\sqcup$ and $\sqcap$ are applicable to arbitrary mappings, we have restricted ourselves to the mappings in Rec for investigating the questions of existence, expressiveness, and computability of $\sqcup$ and $\sqcap$. We would like to extend this study to arbitrary mappings. As has been illustrated in Proposition 5, such an extension may even require an adaptation of the $\preceq_{inf}$-relation.

Recall that in Theorem 5 we have shown that ESO is expressive enough to specify the intersection of two mappings given by sets of st-tgds. Further analysis is required to determine if a smaller fragment of ESO would also suffice. In addition it would be interesting to identify restrictions on the st-tgds such that (some fragment of) FO is expressive enough for expressing the intersection of mappings specified by such restricted st-tgds.

Finally, we would also like to extend our study to further set operators on schema mappings. Above all, we would like to study the complement of a mapping $\mathcal{M}$ (i.e., a mapping that transfers all the source information *not* transferred by $\mathcal{M}$) and, more generally, the set difference of two mappings $\mathcal{M}$ and $\mathcal{N}$ (i.e., a mapping that transfers all the source information that is transferred by $\mathcal{M}$ but not by $\mathcal{N}$). Overall, we think that the union and intersection operators can be crucial in not only defining operators such as difference and complement, but in laying the foundation to a framework of similar operators on schema mappings.

References

1. M. Arenas, J. Pérez, J. L. Reutter, and C. Riveros. Foundations of schema mapping management. In *Proc. PODS*, pages 227–238, 2010.
2. Marcelo Arenas, Ronald Fagin, and Alan Nash. Composition with target constraints. In *Proc. ICDT'10*, ACM International Conference Proceeding Series, pages 129–142. ACM, 2010.
3. Marcelo Arenas, Jorge Pérez, Juan L. Reutter, and Cristian Riveros. Composition and inversion of schema mappings. *SIGMOD Record*, 38(3):17–28, 2009.
4. Marcelo Arenas, Jorge Pérez, and Cristian Riveros. The recovery of a schema mapping: Bringing exchanged data back. *ACM Trans. Database Syst.*, 34(4), 2009.
5. P. Bernstein. Applying model management to classical meta data problems. In *Proc. CIDR*, 2003.
6. P. Bernstein and S. Melnik. Model management 2.0: manipulating richer mappings. In *Proc. SIGMOD*, pages 1–12, 2007.
7. R. Fagin, P. G. Kolaitis, L. Popa, and W.-C. Tan. Composing schema mappings: Second-order dependencies to the rescue. *TODS*, 30(4):994–1055, 2005.
8. Ronald Fagin. Inverting schema mappings. *ACM Trans. Database Syst.*, 32(4), 2007.
9. Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: semantics and query answering. *Theor. Comput. Sci.*, 336(1):89–124, 2005.
10. Ronald Fagin, Phokion G. Kolaitis, Lucian Popa, and Wang Chiew Tan. Reverse data exchange: Coping with nulls. *ACM Trans. Database Syst.*, 36(2):11, 2011.
11. L. M. Haas, M. A. Hernández, H. Ho, L. Popa, and M. Roth. Clio grows up: from research prototype to industrial tool. In *Proc. SIGMOD*, pages 805–810, 2005.
12. Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proc. PODS*, pages 233–246. ACM, 2002.
13. Leonid Libkin. *Elements of Finite Model Theory*. Springer, 2004.
14. Jayant Madhavan and Alon Y. Halevy. Composing mappings among data sources. In *Proc. VLDB'03*, pages 572–583, 2003.
15. Sergey Melnik. *Generic Model Management: concepts and Algorithms*, volume 2967 of *LNCS*. Springer, 2004.
16. Alan Nash, Philip A. Bernstein, and Sergey Melnik. Composition of mappings given by embedded dependencies. *ACM Trans. Database Syst.*, 32(1):4, 2007.
17. Jorge Pérez. *Schema Mapping Management in Data Exchange Systems*. PhD thesis, Escuela de Ingeniería, Pontificia Universidad Católica de Chile, 2011.

A Preliminary tools for maximum recoveries

In this section we reproduce some results presented in [4, 1, 17] as well as some new results about maximum recoveries and their relationship with the order $\preceq_{inf}$.

The first result is a characterization for maximum recoveries proved in [17].

Lemma 1 ([17]). *Let $\mathcal{M}$ be a total mapping that has a maximum recovery, and let $\mathcal{M}'$ be a maximum recovery of $\mathcal{M}$. Then for every pair of instances I, K, it holds that $\mathrm{Sol}_{\mathcal{M}}(K) \subseteq \mathrm{Sol}_{\mathcal{M}}(I)$ if and only if $(I, K) \in \mathcal{M} \circ \mathcal{M}'$*

The following result was provided in [4] as necessary and sufficient condition for a mapping to have a maximum recovery. To formulate it we need to introduce the notion of *witness solution*. We say that J is a witness solution of I under a mapping $\mathcal{M}$ if $J \in \mathrm{Sol}_{\mathcal{M}}(I)$ and for every other instance I' if $J \in \mathrm{Sol}_{\mathcal{M}}(I')$ then $\mathrm{Sol}_{\mathcal{M}}(I) \subseteq \mathrm{Sol}_{\mathcal{M}}(I')$.

Lemma 2 ([4]). *A mapping $\mathcal{M}$ has a maximum recovery if and only if every source instance has a witness solution under $\mathcal{M}$.*

The following is a characterization of the order $\preceq_{inf}$ for mappings that have maximum recoveries.

Lemma 3 ([1]). *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be mappings that have a maximum recovery. Then $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$ if and only if for every pair of instances I, K, if $\mathrm{Sol}_{\mathcal{M}_1}(K) \subseteq \mathrm{Sol}_{\mathcal{M}_1}(I)$ then $\mathrm{Sol}_{\mathcal{M}_2}(K) \subseteq \mathrm{Sol}_{\mathcal{M}_2}(I)$.*

It turns out that the "only if" direction of Lemma 3 holds in general (not restricted to mappings that have a maximum recovery).

Lemma 4. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be mappings such that $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$. Then for every pair of instances I, K, if $\mathrm{Sol}_{\mathcal{M}_1}(K) \subseteq \mathrm{Sol}_{\mathcal{M}_1}(I)$, then $\mathrm{Sol}_{\mathcal{M}_2}(K) \subseteq \mathrm{Sol}_{\mathcal{M}_2}(I)$.*

Proof. Assume that $\mathrm{Sol}_{\mathcal{M}_1}(K) \subseteq \mathrm{Sol}_{\mathcal{M}_1}(I)$ and let $J \in \mathrm{Sol}_{\mathcal{M}_2}(K)$. Since $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$ we know that there exists a mapping $\mathcal{N}$ such that $\mathcal{M}_2 = \mathcal{M}_1 \circ \mathcal{N}$, and thus $(K, J) \in \mathcal{M}_1 \circ \mathcal{N}$ which implies that there exists an instance L such that $(K, L) \in \mathcal{M}_1$ and $(L, J) \in \mathcal{N}$. Given that $\mathrm{Sol}_{\mathcal{M}_1}(K) \subseteq \mathrm{Sol}_{\mathcal{M}_1}(I)$ we know that $(I, L) \in \mathcal{M}_1$ which implies that $(I, J) \in \mathcal{M}_1 \circ \mathcal{N}$ and thus $(I, J) \in \mathcal{M}_2$. Thus we have that if $J \in \mathrm{Sol}_{\mathcal{M}_2}(K)$ then $J \in \mathrm{Sol}_{\mathcal{M}_2}(I)$ which implies that $\mathrm{Sol}_{\mathcal{M}_2}(K) \subseteq \mathrm{Sol}_{\mathcal{M}_2}(I)$. $\square$

We use the above results to prove the following.

Lemma 5. *Let $\mathcal{M}_1$ be a total mapping that has a maximum recovery, and assume that $\mathcal{M}_1'$ is the maximum recovery of $\mathcal{M}_1$. Then we have that $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$ if and only if $\mathcal{M}_2 = \mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{M}_2$.*

Proof. If $\mathcal{M}_2 = \mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{M}_2$, then by definition we have $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$. Towards the opposite direction, assume that $\mathcal{M}_2$ is such that $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$. We first prove that $\mathcal{M}_2 \subseteq \mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{M}_2$. Let $(I, J) \in \mathcal{M}_2$. Given that $\mathcal{M}_1$ is a total mapping and $\mathcal{M}_1'$ is a recovery of $\mathcal{M}_1$, we have that $(I, I) \in \mathcal{M}_1 \circ \mathcal{M}_1'$ and then $(I, J) \in \mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{M}_2$.

We now prove that $\mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{M}_2 \subseteq \mathcal{M}_2$. Let $(I, J) \in \mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{M}_2$. We need to prove that $(I, J) \in \mathcal{M}_2$. Now, since $(I, J) \in \mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{M}_2$ we know that there exists an instance K such that $(I, K) \in \mathcal{M}_1 \circ \mathcal{M}_1'$ and $(K, J) \in \mathcal{M}_2$. By Lemma 1 we know that $\mathrm{Sol}_{\mathcal{M}_1}(K) \subseteq \mathrm{Sol}_{\mathcal{M}_1}(I)$. Moreover, given that $\mathcal{M}_2 \preceq_{inf} \mathcal{M}_1$, by Lemma 4 we have that $\mathrm{Sol}_{\mathcal{M}_2}(K) \subseteq \mathrm{Sol}_{\mathcal{M}_2}(I)$. Thus, since $J \in \mathrm{Sol}_{\mathcal{M}_2}(K)$, we have that $J \in \mathrm{Sol}_{\mathcal{M}_2}(I)$ and then $(I, J) \in \mathcal{M}_2$ which was to be shown. $\square$

B Proof of Proposition 1

Proof. We only prove the claim for the union operator. The intersection is treated analogously. Suppose that in the setting of this proposition, $\mathcal{N} = \mathcal{M}_1 \sqcup_{\mathcal{C}} \mathcal{M}_2$ exists. Then we have $\mathcal{M}_i' \equiv_{inf} \mathcal{M}_i \preceq_{inf} \mathcal{N}$ for $i \in \{1, 2\}$. Now suppose that there exists some $\mathcal{N}'$ with $\mathcal{M}_i' \preceq_{inf} \mathcal{N}'$ for $i \in \{1, 2\}$. Then also $\mathcal{M}_i \preceq_{inf} \mathcal{N}'$ for $i \in \{1, 2\}$ holds and, therefore, $\mathcal{N} \preceq_{inf} \mathcal{N}'$. Thus, also $\mathcal{N} = \mathcal{M}_1' \sqcup_{\mathcal{C}} \mathcal{M}_2'$ holds. $\square$

C Proof of Propositions 2 and 3

We begin by proving Proposition 3. In the proof we denote by Rec the class of total mappings that have maximum recovery. Moreover, given disjoint schemas $\mathbf{T}_1$ and $\mathbf{T}_2$ we consider the mapping $\mathcal{M}_1 \oplus \mathcal{M}_2$ from $\mathbf{S}$ to $\mathbf{T}_1 \cup \mathbf{T}_2$ defined as:

$$\mathcal{M}_1 \oplus \mathcal{M}_2 = \{(I, J_1 \cup J_2) \mid (I, J_1) \in \mathcal{M}_1 \text{ and } (I, J_2) \in \mathcal{M}_2\}.$$

Proof (of Proposition 3). Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be disjoint schemas, $\mathcal{M}_1$ a mapping in Rec from $\mathbf{S}$ to $\mathbf{T}_1$, and $\mathcal{M}_2$ a mapping in Rec from $\mathbf{S}$ to $\mathbf{T}_2$. We prove next that $\mathcal{M}_1 \oplus \mathcal{M}_2$ is the union $\mathcal{M}_1 \sqcup_{\mathrm{Rec}} \mathcal{M}_2$. We show first that $\mathcal{M}_1 \oplus \mathcal{M}_2$ is in Rec. Notice that $\mathcal{M}_1 \oplus \mathcal{M}_2$ is a total mapping (given that $\mathcal{M}_1$ and $\mathcal{M}_2$ is total), thus we only need to prove that $\mathcal{M}_1 \oplus \mathcal{M}_2$ has a maximum recovery. For this we use Lemma 2 that characterizes mappings that have maximum recovery by using the notion of witness solution. Let I be an instance of $\mathbf{S}$, and let J_1 be the witness solution of I under $\mathcal{M}_1$ and J_2 the witness solution of I under $\mathcal{M}_2$ (we know that they exist since $\mathcal{M}_1$ and $\mathcal{M}_2$ are in Rec). We claim that $J_1 \cup J_2$ is a witness solution of I under $\mathcal{M}_1 \oplus \mathcal{M}_2$. Thus, let I' be an instance such that $J_1 \cup J_2 \in \mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I')$. We need to prove that $\mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I) \subseteq \mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I')$. Given that $J_1 \cup J_2 \in \mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I')$, and by the construction of $\mathcal{M}_1 \oplus \mathcal{M}_2$, we know that $J_1 \in \mathrm{Sol}_{\mathcal{M}_1}(I')$ and $J_2 \in \mathrm{Sol}_{\mathcal{M}_2}(I')$, and therefor, $\mathrm{Sol}_{\mathcal{M}_1}(I) \subseteq \mathrm{Sol}_{\mathcal{M}_1}(I')$ and $\mathrm{Sol}_{\mathcal{M}_2}(I) \subseteq \mathrm{Sol}_{\mathcal{M}_2}(I')$. Now, let $K \in \mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I)$. By the construction of $\mathcal{M}_1 \oplus \mathcal{M}_2$ we know that $K = K_1 \cup K_2$ with $K_1 \in \mathrm{Sol}_{\mathcal{M}_1}(I)$ and $K_2 \in \mathrm{Sol}_{\mathcal{M}_2}(I)$. Given that $\mathrm{Sol}_{\mathcal{M}_1}(I) \subseteq \mathrm{Sol}_{\mathcal{M}_1}(I')$ we obtain that $K_1 \in \mathrm{Sol}_{\mathcal{M}_1}(I')$, and by a similar argument we have that $K_2 \in \mathrm{Sol}_{\mathcal{M}_2}(I')$ which implies that $K_1 \cup K_2 \in$

$\mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I')$. This completes the proof that $\mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I) \subseteq \mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I')$, and then $J_1 \cup J_2$ is a witness solution of I. From Lemma 2, we conclude that $\mathcal{M}_1 \oplus \mathcal{M}_2$ has a maximum recovery and then $\mathcal{M}_1 \oplus \mathcal{M}_2$ is in Rec.

We prove now that $\mathcal{M}_1 \preceq_{\mathit{inf}} \mathcal{M}_1 \oplus \mathcal{M}_2$. Consider the mapping $\mathcal{N}_1$ from $\mathbf{T}_1 \cup \mathbf{T}_2$ to $\mathbf{T}_1$ defined as follows:

$$\mathcal{N}_1 = \{(J_1 \cup J_2, J_1) \mid J_1 \in \mathrm{Inst}(\mathbf{T}_1) \text{ and } J_2 \in \mathrm{Inst}(\mathbf{T}_2)\}.$$

We prove now that $(\mathcal{M}_1 \oplus \mathcal{M}_2) \circ \mathcal{N}_1 = \mathcal{M}_1$. In order to prove that $\mathcal{M}_1 \subseteq (\mathcal{M}_1 \oplus \mathcal{M}_2) \circ \mathcal{N}_1$ let $(I, J_1) \in \mathcal{M}_1$. Given that $\mathcal{M}_2$ is a total mapping, then we know that there exists an instance J_2 such that $(I, J_2) \in \mathcal{M}_2$. Then we have that $(I, J_1 \cup J_2) \in \mathcal{M}_1 \oplus \mathcal{M}_2$. Moreover, $(J_1 \cup J_2, J_1) \in \mathcal{N}_1$ and thus $(I, J_1) \in (\mathcal{M}_1 \oplus \mathcal{M}_2) \circ \mathcal{N}_1$. This proves that $\mathcal{M}_1 \subseteq (\mathcal{M}_1 \oplus \mathcal{M}_2)$. Assume now that $(I, J_1) \in (\mathcal{M}_1 \oplus \mathcal{M}_2) \circ \mathcal{N}_1$. Then we know that there exists an instance L such that $(I, L) \in \mathcal{M}_1 \oplus \mathcal{M}_2$ and $(L, J_1) \in \mathcal{N}_1$. By the definition of $\mathcal{N}_1$ we know that there exists an instance $J_2 \in \mathrm{Inst}(\mathbf{T}_2)$ such that $L = J_1 \cup J_2$. Thus, since $(I, J_1 \cup J_2) \in \mathcal{M}_1 \oplus \mathcal{M}_2$ we have that $(I, J_1) \in \mathcal{M}_1$ and $(I, J_2) \in \mathcal{M}_2$. We have proved that if $(I, J_1) \in (\mathcal{M}_1 \oplus \mathcal{M}_2) \circ \mathcal{N}_1$ then $(I, J_1) \in \mathcal{M}_1$ which proves that $(\mathcal{M}_1 \oplus \mathcal{M}_2) \circ \mathcal{N}_1 \subseteq \mathcal{M}_1$. Thus, since $(\mathcal{M}_1 \oplus \mathcal{M}_2) \circ \mathcal{N}_1 = \mathcal{M}_1$ we have that $\mathcal{M}_1 \preceq_{\mathit{inf}} \mathcal{M}_1 \oplus \mathcal{M}_2$. By using a symmetrical argument one can prove that $\mathcal{M}_2 \preceq_{\mathit{inf}} \mathcal{M}_1 \oplus \mathcal{M}_2$.

To complete the proof of Proposition 3 we only need to show that if $\mathcal{N}$ is a mapping in Rec such that $\mathcal{M}_1 \preceq_{\mathit{inf}} \mathcal{N}$ and $\mathcal{M}_2 \preceq_{\mathit{inf}} \mathcal{N}$ then $\mathcal{M}_1 \oplus \mathcal{M}_2 \preceq_{\mathit{inf}} \mathcal{N}$. Given that $\mathcal{M}_1 \oplus \mathcal{M}_2$ and $\mathcal{N}$ are in Rec we can use the characterization in Lemma 3. Thus, in order to prove that $\mathcal{M}_1 \oplus \mathcal{M}_2 \preceq_{\mathit{inf}} \mathcal{N}$ we need to show that for every pair of instances I and K, if $\mathrm{Sol}_{\mathcal{N}}(I) \subseteq \mathrm{Sol}_{\mathcal{N}}(K)$ then $\mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I) \subseteq \mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(K)$. Thus, assume that $\mathrm{Sol}_{\mathcal{N}}(I) \subseteq \mathrm{Sol}_{\mathcal{N}}(K)$ and let $J \in \mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I)$. By definition of $\mathcal{M}_1 \oplus \mathcal{M}_2$ we know that $J = J_1 \cup J_2$ with $J_1 \in \mathrm{Sol}_{\mathcal{M}_1}(I)$ and $J_2 \in \mathrm{Sol}_{\mathcal{M}_2}(I)$. Now notice that $\mathcal{M}_1$ is also in Rec, then since $\mathcal{M}_1 \preceq_{\mathit{inf}} \mathcal{N}$ and $\mathrm{Sol}_{\mathcal{N}}(I) \subseteq \mathrm{Sol}_{\mathcal{N}}(K)$, we have that $\mathrm{Sol}_{\mathcal{M}_1}(I) \subseteq \mathrm{Sol}_{\mathcal{M}_1}(K)$. Similarly we can conclude that $\mathrm{Sol}_{\mathcal{M}_2}(I) \subseteq \mathrm{Sol}_{\mathcal{M}_2}(K)$. Thus, given that $J_1 \in \mathrm{Sol}_{\mathcal{M}_1}(I)$ and $J_2 \in \mathrm{Sol}_{\mathcal{M}_2}(I)$ we have that $J_1 \in \mathrm{Sol}_{\mathcal{M}_1}(K)$ and $J_2 \in \mathrm{Sol}_{\mathcal{M}_2}(K)$, and then $J = J_1 \cup J_2 \in \mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(K)$. We have shown that $\mathrm{Sol}_{\mathcal{N}}(I) \subseteq \mathrm{Sol}_{\mathcal{N}}(K)$ implies $\mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(I) \subseteq \mathrm{Sol}_{\mathcal{M}_1 \oplus \mathcal{M}_2}(K)$. Finally, by applying Lemma 3 we obtain $\mathcal{M}_1 \oplus \mathcal{M}_2 \preceq_{\mathit{inf}} \mathcal{N}$.

Assume now that $\mathcal{M}_1$ is a mapping in Rec from $\mathbf{S}$ to $\mathbf{T}_1$ and $\mathcal{M}_2$ is a mapping in Rec from $\mathbf{S}$ to $\mathbf{T}_2$ such that $\mathbf{T}_1$ and $\mathbf{T}_2$ are not necessarily disjoint. Then consider a copy $\widehat{\mathbf{T}}_2$ of schema $\mathbf{T}_2$ disjoint of $\mathbf{T}_1$, and the mapping

$$\widehat{\mathcal{M}_2} = \{(I, \hat{J}) \mid (I, J) \in \mathcal{M}_2 \text{ and } \hat{J} \text{ the copy of } J \text{ over } \widehat{\mathbf{T}}_2\}.$$

It is straightforward to show that $\widehat{\mathcal{M}_2} \equiv_{\mathit{inf}} \mathcal{M}_2$ and thus, given that $\mathcal{M}_1 \oplus \widehat{\mathcal{M}_2}$ is the union $\mathcal{M}_1 \sqcup_{\text{Rec}} \widehat{\mathcal{M}_2}$, we have that $\mathcal{M}_1 \oplus \widehat{\mathcal{M}_2}$ is also the union $\mathcal{M}_1 \sqcup_{\text{Rec}} \mathcal{M}_2$. $\square$

We can now prove Proposition 2.

Proof (of Proposition 2). Let $\mathcal{M}_1 = (\mathbf{S}, \mathbf{T}_1, \Sigma_1)$ and $\mathcal{M}_2 = (\mathbf{S}, \mathbf{T}_2, \Sigma_2)$ be two mappings specified by st-tgds. Let $\widehat{\mathbf{T}}_2$ be a copy of $\mathbf{T}_2$ such that $\widehat{\mathbf{T}}_2$ is disjoint with $\mathbf{T}_1$, and $\widehat{\Sigma}_2$ be the set that results from Σ_2 by replacing every relation name in $\mathbf{T}_2$ by its copy in $\widehat{\mathbf{T}}_2$. Now consider the mapping $\mathcal{M}' = (\mathbf{S}, \mathbf{T}_1 \cup \widehat{\mathbf{T}}_2, \Sigma_1 \cup \widehat{\Sigma}_2)$. We show next that $\mathcal{M}' = \mathcal{M}_1 \oplus \widehat{\mathcal{M}}_2$ where $\widehat{\mathcal{M}}_2$ is the mapping from $\mathbf{S}$ to $\widehat{\mathbf{T}}_2$ specified by $\widehat{\Sigma}_2$.

Given an instance J over schema $\mathbf{T}_1 \cup \widehat{\mathbf{T}}_2$ we denote by $J_{\mathbf{T}_1}$ the restriction of J to schema $\mathbf{T}_1$ and by $J_{\widehat{\mathbf{T}}_2}$ the restriction of J to schema $\widehat{\mathbf{T}}_2$. Now, assume that $(I, J) \in \mathcal{M}'$. Then $(I, J) \models \Sigma_1$ and $(I, J) \models \widehat{\Sigma}_2$. Moreover, since $\mathbf{T}_1$ and $\widehat{\mathbf{T}}_2$ are disjoint, we have that $(I, J_{\mathbf{T}_1}) \models \Sigma_1$ and $(I, J_{\widehat{\mathbf{T}}_2}) \models \widehat{\Sigma}_2$. Thus, $(I, J_{\mathbf{T}_1}) \in \mathcal{M}_1$, $(I, J_{\widehat{\mathbf{T}}_2}) \in \widehat{\mathcal{M}}_2$ and $J = J_{\mathbf{T}_1} \cup J_{\widehat{\mathbf{T}}_2}$, which implies that $(I, J) \in \mathcal{M}_1 \oplus \widehat{\mathcal{M}}_2$. The other direction is similar.

It is straightforward that mapping $\mathcal{M}'$ can be constructed from $\mathcal{M}_1$ and $\mathcal{M}_2$ in linear time. □

D Proof of Proposition 5

Proof. Consider $\mathcal{M}$ defined as $\{(I_1, J_1), (I_2, J_1), (I_2, J_2), (I_3, J_2)\}$. $\mathcal{M}$ has no maximum recovery, since there is no witness solution for I_2. We will make a transition from $\mathcal{M} \oplus \mathcal{M}$ to $\mathcal{M} \oplus \widehat{\mathcal{M}}$, where $\widehat{\mathcal{M}}$ is obtained from $\mathcal{M}$ by renaming apart the target relations. As mentioned in the proof of Proposition 3, $\widehat{\mathcal{M}} \equiv_{inf} \mathcal{M}$ holds.

We observe that $\mathcal{M} \circ \mathcal{N} = \mathcal{M} \oplus \widehat{\mathcal{M}}$ cannot be satisfied: Indeed, $\mathcal{M} \oplus \widehat{\mathcal{M}} = \{(I_1, J_1 \cup \hat{J}_1), (I_2, J_1 \cup \hat{J}_1), (I_2, J_1 \cup \hat{J}_2), (I_2, J_2 \cup \hat{J}_1), (I_2, J_2 \cup \hat{J}_2), (I_3, J_2 \cup \hat{J}_2)\}$, and a simple enumeration of all 16 possible images for J_1 resp. J_2 (2^4 possible sets of target instances) shows that no mapping $\mathcal{N}$ can accommodate for the equality $\mathcal{M} \circ \mathcal{N} = \mathcal{M} \oplus \widehat{\mathcal{M}}$.

In the other direction, the mapping $\mathcal{N}' = \{(J_1 \cup \hat{J}_1, J_1), (J_2 \cup \hat{J}_2, J_2)\}$ satisfies the equality $\mathcal{M} = (\mathcal{M} \oplus \widehat{\mathcal{M}}) \circ \mathcal{N}'$. □

E Proof of Proposition 4

In [1] it was shown that given mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ specified by st-tgds, checking whether $\mathcal{M}_1 \preceq_{inf} \mathcal{M}_2$ can be done in NEXPTIME. Thus given three mappings $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ specified by st-tgds, in order to check whether $\mathcal{M}_3$ is the union of $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. the class of mappings specified by st-tgds, one can do the following:

1. First construct mapping $\mathcal{M}'$ described in the proof of Proposition 2.
2. Then check that $\mathcal{M}_3 \preceq_{inf} \mathcal{M}'$ and that $\mathcal{M}' \preceq_{inf} \mathcal{M}_3$.

Step (1) can be done in linear time, and (2) can be done in NEXPTIME [1].

F Proof of Theorem 3

Proof. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be mappings with source schema $\mathbf{S}$, and let $\mathcal{M}_1'$ and $\mathcal{M}_2'$ be the maximum recoveries of $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. To simplify the notation, let $\mathcal{N} = (\mathcal{M}_1 \circ \mathcal{M}_1') \cup (\mathcal{M}_2 \circ \mathcal{M}_2')$. We prove next that $\mathcal{N}^+$ satisfies the following properties:

(1) $\mathcal{N}^+ \preceq_{\mathit{inf}} \mathcal{M}_1$ and $\mathcal{N}^+ \preceq_{\mathit{inf}} \mathcal{M}_2$, and
(2) if $\mathcal{M}$ is an arbitrary mapping such that $\mathcal{M} \preceq_{\mathit{inf}} \mathcal{M}_1$ and $\mathcal{M} \preceq_{\mathit{inf}} \mathcal{M}_2$, then $\mathcal{M} \preceq_{\mathit{inf}} \mathcal{N}^+$.

Notice that (1) and (2) implies that $\mathcal{N}^+$ is the intersection $\mathcal{M}_1 \sqcap_{\text{ALL}} \mathcal{M}_2$, where ALL is the class of all mappings.

We first prove (1). By using Lemma 5, in order to prove (1) it is enough to show that

$$\mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{N}^+ = \mathcal{N}^+ \text{ and } \mathcal{M}_2 \circ \mathcal{M}_2' \circ \mathcal{N}^+ = \mathcal{N}^+.$$

We begin by proving $\mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{N}^+ = \mathcal{N}^+$. Consider first the inclusion $\mathcal{N}^+ \subseteq \mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{N}^+$ and let $(I, J) \in \mathcal{N}^+$. Then, since $\mathcal{M}_1$ is total and $\mathcal{M}_1'$ is a recovery of $\mathcal{M}_1$, we have that $(I, I) \in \mathcal{M}_1 \circ \mathcal{M}_1'$, and thus $(I, J) \in \mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{N}^+$ which was to be shown. In order to prove the inclusion $\mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{N}^+ \subseteq \mathcal{N}^+$, let $(I, J) \in \mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{N}^+$. Then there exists an instance K such that $(I, K) \in \mathcal{M}_1 \circ \mathcal{M}_1'$ and $(K, J) \in \mathcal{N}^+$. Given that $\mathcal{M}_1 \circ \mathcal{M}_1' \subseteq \mathcal{N} \subseteq \mathcal{N}^+$ we have that $(I, K) \in \mathcal{N}^+$. Finally, since $(I, K) \in \mathcal{N}^+$, $(K, J) \in \mathcal{N}^+$ and $\mathcal{N}^+$ is transitive, we obtain that $(I, J) \in \mathcal{N}^+$ which was to be shown. The equality $\mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{N}^+ = \mathcal{N}^+$ is proved similarly.

We now prove (2). Let $\mathcal{M}$ be a mapping such that $\mathcal{M} \preceq_{\mathit{inf}} \mathcal{M}_1$ and $\mathcal{M} \preceq_{\mathit{inf}} \mathcal{M}_2$. We next prove that $\mathcal{N}^+ \circ \mathcal{M} = \mathcal{M}$ which implies that $\mathcal{M} \preceq_{\mathit{inf}} \mathcal{N}^+$. In order to prove that $\mathcal{N}^+ \circ \mathcal{M} = \mathcal{M}$ it is enough to show that for every positive integer k it holds that $\mathcal{N}^k \circ \mathcal{M} = \mathcal{M}$. We prove this by induction in k. For the base case, we need to prove that $\mathcal{N} \circ \mathcal{M} = \mathcal{M}$. Given that $\mathcal{M} \preceq_{\mathit{inf}} \mathcal{M}_1$ by Lemma 5, we know that $\mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{M} = \mathcal{M}$. Similarly we have that $\mathcal{M}_2 \circ \mathcal{M}_2' \circ \mathcal{M} = \mathcal{M}$. This implies that

$$\begin{aligned}
\mathcal{N} \circ \mathcal{M} &= \Big((\mathcal{M}_1 \circ \mathcal{M}_1') \cup (\mathcal{M}_2 \circ \mathcal{M}_2') \Big) \circ \mathcal{M} \\
&= \Big(\mathcal{M}_1 \circ \mathcal{M}_1' \circ \mathcal{M} \Big) \cup \Big(\mathcal{M}_2 \circ \mathcal{M}_2' \circ \mathcal{M} \Big) \\
&= \mathcal{M} \cup \mathcal{M} \\
&= \mathcal{M}
\end{aligned}$$

and thus, the base case holds. Now as induction hypothesis, assume that $\mathcal{N}^k \circ \mathcal{M} = \mathcal{M}$, and consider $\mathcal{N}^{k+1} \circ \mathcal{M}$. By definition we have that $\mathcal{N}^{k+1} \circ \mathcal{M} = \mathcal{N}^k \circ \mathcal{N} \circ \mathcal{M}$. We also know that $\mathcal{N} \circ \mathcal{M} = \mathcal{M}$ and thus $\mathcal{N}^{k+1} \circ \mathcal{M} = \mathcal{N}^k \circ \mathcal{M}$. Finally, by induction hypothesis we know that $\mathcal{N}^k \circ \mathcal{M} = \mathcal{M}$, and thus we obtain that $\mathcal{N}^{k+1} \circ \mathcal{M} = \mathcal{M}$ which completes the proof of (2).

Notice that we have shown that $\mathcal{N}^+$ is the intersection $\mathcal{M}_1 \sqcap_{\text{ALL}} \mathcal{M}_2$. Thus, in order to prove that $\mathcal{N}^+$ is the intersection $\mathcal{M}_1 \sqcap_{\text{REC}} \mathcal{M}_2$ we only need to prove that $\mathcal{N}^+$ is in REC. We prove next that $\mathcal{N}^+$ is its own maximum recovery. First, given that $\mathcal{M}_1$ is a total mapping and $\mathcal{M}'_1$ is a maximum recovery of $\mathcal{M}_1$, then for every instance I of $\mathbf{S}$, we have that $(I, I) \in \mathcal{M}_1 \circ \mathcal{M}'_1$ and similarly $(I, I) \in \mathcal{M}_2 \circ \mathcal{M}'_2$. This implies that for every I in $\mathbf{S}$ we have $(I, I) \in \mathcal{N}$ and thus $(I, I) \in \mathcal{N}^+$. Moreover, $\mathcal{N}^+ \circ \mathcal{N}^+ = \mathcal{N}^+$ (since $\mathcal{N}^+$ is a transitive relation), and thus $(I, I) \in \mathcal{N}^+ \circ \mathcal{N}^+$. This implies that $\mathcal{N}^+$ is a recovery of $\mathcal{N}^+$.

To complete the proof we use a characterization provided in [4]. It was proved in [4] (Proposition 3.8[3]) that for a total mapping $\mathcal{M}$, the mapping $\mathcal{M}'$ is a maximum recovery of $\mathcal{M}$ if and only if $\mathcal{M}'$ is a recovery of $\mathcal{M}$ and $\mathcal{M} \circ \mathcal{M}' \circ \mathcal{M} = \mathcal{M}$. We already know that $\mathcal{N}^+$ is a recovery of $\mathcal{N}^+$. Moreover, since $\mathcal{N}^+$ is a transitive relation, we have that $\mathcal{N}^+ \circ \mathcal{N}^+ \circ \mathcal{N}^+ = \mathcal{N}^+$ and thus $\mathcal{N}^+$ is a maximum recovery of $\mathcal{N}^+$. Thus we have proved that $\mathcal{N}^+$ is in REC which completes the proof of the Theorem. □

G Proof of Theorem 4

Recall that we are assuming that instances are constructed from a countably infinite set $\mathbf{D}$. We also assume that all formulas used to specify mappings are *domain independent*. We also need the following technical definition. We say that a mapping $\mathcal{M}$ is *closed under isomorphisms*, if for every $(I, J) \in \mathcal{M}$ and injective function $f : \text{dom}(I) \cup \text{dom}(J) \to \mathbf{D}$, it holds that $(f(I), f(J)) \in \mathcal{M}$.

Before formally proving Theorem 4 we introduce some terminology and prove some intermediate lemmas regarding $\mathcal{M}_1$ and $\mathcal{M}_2$.

Definition 3. *Let $\mathcal{M}$ be a mapping from $\mathbf{S}$ to $\mathbf{T}$.*

- *We say that $\mathcal{M}$ is* invariant under source-isomorphisms *if for every pair of isomorphic instances I_1 and I_2 of schema $\mathbf{S}$, it holds that $\text{Sol}_{\mathcal{M}}(I_1) = \text{Sol}_{\mathcal{M}}(I_2)$.*
- *We say that $\mathcal{M}$ is* invariant under target-isomorphisms *if for every instance I of schema $\mathbf{S}$ and pair of isomorphic instances J_1 and J_2 of schema $\mathbf{T}$, it holds that $J_1 \in \text{Sol}_{\mathcal{M}}(I)$ if and only if $J_2 \in \text{Sol}_{\mathcal{M}}(I)$.*

The following simple result shows the relationship between invariance under source and target isomorphisms.

Lemma 6. *Let $\mathcal{M}$ be a mapping from $\mathbf{S}$ to $\mathbf{T}$ and assume that $\mathcal{M}$ is closed under isomorphisms. Then $\mathcal{M}$ is invariant under source isomorphism if and only if $\mathcal{M}$ is invariant under target isomorphisms.*

[3] Proposition 3.8 uses the notion of *reduced recovery* to provide the characterization. The notion of reduced recovery coincides with the notion of recovery for total mappings which is the case that we are considering in this proof

Proof. Assume first that $\mathcal{M}$ is invariant under source isomorphisms, and let $J_1 \in \mathrm{Sol}_{\mathcal{M}}(I)$. Consider a isomorphic copy J_2 of J_1 and an isomorphism f such that $J_2 = f(J_1)$. Let f' be the extension of f over $\mathrm{dom}(I)$ such that for every $a \in \mathrm{dom}(I) \smallsetminus \mathrm{dom}(J_1)$, the value $f'(a)$ is a fresh value (not in $\mathrm{dom}(I) \cup \mathrm{dom}(J_1) \cup \mathrm{dom}(J_2)$). Then $(f'(I), f'(J_1))$ is isomorphic to (I, J_1) and then $(f'(I), f'(J_1)) \in \mathcal{M}$ (since $\mathcal{M}$ is closed under isomorphisms). Moreover $f'(J_1) = f(J_1) = J_2$. Thus we have that $J_2 \in \mathrm{Sol}_{\mathcal{M}}(f'(I))$. Finally, since I is isomorphic to $f'(I)$, and given that $\mathcal{M}$ is closed under source isomorphisms, we have that $J_2 \in \mathrm{Sol}_{\mathcal{M}}(I)$. Thus we have shown that if $J_1 \in \mathrm{Sol}_{\mathcal{M}}(I)$ and J_2 is an isomorphic copy of J_1, then $J_2 \in \mathrm{Sol}_{\mathcal{M}}(I)$. A symmetric argument shows that if $J_2 \in \mathrm{Sol}_{\mathcal{M}}(I)$ then $J_1 \in \mathrm{Sol}_{\mathcal{M}}(I)$ implying that $\mathcal{M}$ is invariant under target isomorphisms.

For the other direction, assume that $\mathcal{M}$ is invariant under target isomorphisms, and let I_1 and I_2 be two isomorphic source instances of $\mathbf{S}$ and f the isomorphism between I_1 and I_2, that is, $I_2 = f(I_1)$. Now assume that $J \in \mathrm{Sol}_{\mathcal{M}}(I_1)$ and consider f' as the extension of f over $\mathrm{dom}(J)$ such that for every $a \in \mathrm{dom}(J) \smallsetminus \mathrm{dom}(I_1)$, the value $f'(a)$ is a fresh value (not in $\mathrm{dom}(I_1) \cup \mathrm{dom}(I_2) \cup \mathrm{dom}(J)$). Then $(f'(I_1), f'(J))$ is isomorphic to (I_1, J) and thus $f'(J) \in \mathrm{Sol}_{\mathcal{M}}(f'(I_1))$. Moreover $f'(I_1) = f(I_1) = I_2$. Thus we have that $f'(J) \in \mathrm{Sol}_{\mathcal{M}}(I_2)$, or equivalently $(I_2, f'(J)) \in \mathcal{M}$. Finally, since $f'(J)$ is isomorphic to J, and given that $\mathcal{M}$ is closed under target isomorphisms, we have that $(I_2, J) \in \mathcal{M}$, and thus $J \in \mathrm{Sol}_{\mathcal{M}}(I_2)$. Thus we have shown that $\mathrm{Sol}_{\mathcal{M}}(I_1) \subseteq \mathrm{Sol}_{\mathcal{M}}(I_2)$. A symmetric argument can be used to show that $\mathrm{Sol}_{\mathcal{M}}(I_2) \subseteq \mathrm{Sol}_{\mathcal{M}}(I_1)$, and then $\mathrm{Sol}_{\mathcal{M}}(I_1) = \mathrm{Sol}_{\mathcal{M}}(I_2)$ which proves that $\mathcal{M}$ is invariant under source isomorphisms. □

Consider the schemas $\mathbf{S} = \{A(\cdot,\cdot), B(\cdot,\cdot)\}$, $\mathbf{T}_1 = \{T_1(\cdot,\cdot)\}$, and $\mathbf{T}_2 = \{T_2(\cdot,\cdot)\}$. In the rest of the proof, we will use mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ from $\mathbf{S}$ to $\mathbf{T}_1$ and from $\mathbf{S}$ to $\mathbf{T}_2$, respectively, specified by the following st-tgds:

$$\mathcal{M}_1 : \ \exists u\ (A(x,u) \wedge B(u,y)) \to T_1(x,y) \tag{1}$$

$$\mathcal{M}_2 : \ \exists u\ (B(x,u) \wedge A(u,y)) \to T_2(x,y) \tag{2}$$

The following lemma shows that any mapping which is less informative than both $\mathcal{M}_1$ and $\mathcal{M}_2$ defined above, must be invariant under source isomorphisms.

Lemma 7. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be the mappings* (1) *and* (2), *respectively. If $\mathcal{M}$ is a mapping which is closed under isomorphisms and such that $\mathcal{M} \preceq_{inf} \mathcal{M}_1$ and $\mathcal{M} \preceq_{inf} \mathcal{M}_2$, then $\mathcal{M}$ is invariant under source isomorphisms.*

Proof. Let $\mathcal{M}$ be a mapping such that $\mathcal{M} \preceq_{inf} \mathcal{M}_1$ and $\mathcal{M} \preceq_{inf} \mathcal{M}_2$, and consider two isomorphic instances I_1 and I_2 of schema $\mathbf{S}$. We need to show that $\mathrm{Sol}_{\mathcal{M}}(I_1) = \mathrm{Sol}_{\mathcal{M}}(I_2)$. In the proof we assume that f is the isomorphism between I_1 and I_2, that is, $I_1 = f(I_2)$.

Consider first an instance I_1' constructed from I_1 as follows. Let $g : \mathrm{dom}(I_1) \to \mathbf{D}$ be an injective function such that for every element $a \in \mathrm{dom}(I_1)$, the value $g(a)$ is a *fresh* element not present in $\mathrm{dom}(I_1) \cup \mathrm{dom}(I_2)$. Then we construct I_1' as the set of facts:

$$I_1' = \{A(a, g(b)) \mid A(a,b) \in I_1\} \cup \{B(g(a), b) \mid B(a,b) \in I_1\}.$$

That is I_1' is obtained from I_1 by replacing elements in the second component of relation A and in the first component of relation B, by fresh elements. It is easy to see that $\mathrm{Sol}_{\mathcal{M}_1}(I_1) = \mathrm{Sol}_{\mathcal{M}_1}(I_1')$. Just notice that buy the construction of I_1 we have that $I \models \exists u\ (A(a,u) \wedge B(u,b))$ for arbitrary elements a, b, if and only if $I_1 \models \exists u\ (A(a,u) \wedge B(u,b))$. Consider now an instance I_1'' constructed from I_1' as follows. Let $h : \mathrm{dom}(I_1') \to \mathbf{D}$ be an injective function such that for every element in $a \in \mathrm{dom}(I_1)$, the value $h(a)$ is a *fresh* element not present in $\mathrm{dom}(I_1) \cup \mathrm{dom}(I_2) \cup \mathrm{dom}(I_1')$. Then we construct I_1'' as the set of facts:

$$I_1'' = \{A(h(a), b) \mid A(a,b) \in I_1'\} \ \cup\ \{B(a, h(b)) \mid B(a,b) \in I_1'\}.$$

It is easy to prove that $\mathrm{Sol}_{\mathcal{M}_2}(I_1') = \mathrm{Sol}_{\mathcal{M}_2}(I_1'')$. Just notice that by the construction of I_1'' we have that $I_1' \models \exists u\ (B(a,u) \wedge A(u,b))$ for arbitrary elements a, b, if and only if $I_1'' \models \exists u\ (B(a,u) \wedge A(u,b))$. Finally, notice that I_1'' can be written in terms of I_1 as the set of facts:

$$I_1'' = \{A(h(a), g(b)) \mid A(a,b) \in I_1\} \ \cup\ \{B(g(a), h(b)) \mid B(a,b) \in I_1\}.$$

Similarly as for I_1, we construct two instances I_2' and I_2'' from I_2 as follows. The instance I_2' is constructed as the set of facts:

$$I_2' = \{A\big(a, g(f^{-1}(b))\big) \mid A(a,b) \in I_2\} \ \cup\ \{B\big(g(f^{-1}(a)), b\big) \mid B(a,b) \in I_2\}.$$

Notice that, similarly as in the construction of I_1', instance I_2' is obtained from I_2 by replacing every element a occurring in the second component of relation A or in the first component of relation B, by a fresh element, in this case given by $g(f^{-1}(a))$. It is also easy to see that $\mathrm{Sol}_{\mathcal{M}_1}(I_2) = \mathrm{Sol}_{\mathcal{M}_1}(I_2')$. Now to construct I_2'' we consider the following set of facts:

$$I_2'' = \{A\big(h(f^{-1}(a)), b\big) \mid A(a,b) \in I_2'\} \ \cup\ \{B\big(a, h(f^{-1}(b))\big) \mid B(a,b) \in I_2'\}.$$

It is easy to prove that $\mathrm{Sol}_{\mathcal{M}_2}(I_2') = \mathrm{Sol}_{\mathcal{M}_2}(I_2'')$. Moreover, I_2'' satisfies an additional property, namely, that $I_2'' = I_1''$. To see that, notice I_2'' can be written in terms of I_2 as the set of facts:

$$I_2'' = \{A\big(h(f^{-1}(a)), g(f^{-1}(b)\big) \mid A(a,b) \in I_2\} \cup \\ \{B\big(g(f^{-1}(a)), h(f^{-1}(b))\big) \mid B(a,b) \in I_2\}.$$

Now since $I_1 = f(I_2)$ and f is an isomorphism we have that $I_2 = f^{-1}(I_1)$, and then $A(a,b) \in I_2$ if and only if $A(f^{-1}(a), f^{-1}(b)) \in I_1$, and $B(a,b) \in I_2$ if and only if $B(f^{-1}(a), f^{-1}(b)) \in I_1$. This implies that we can write I_2'' as

$$I_2'' = \{A\big(h(f^{-1}(a)), g(f^{-1}(b)\big) \mid A\big(f^{-1}(a), f^{-1}(b)\big) \in I_1\} \cup \\ \{B\big(g(f^{-1}(a)), h(f^{-1}(b))\big) \mid B\big(f^{-1}(a), f^{-1}(b)\big) \in I_1\}.$$

Finally, since f is an isomorphism, we have that f^{-1} is just a renaming of the elements in I_1 and thus we can write I_2'' as the set of facts

$$I_2'' = \{A(h(a), g(b)) \mid A(a,b) \in I_1\} \ \cup\ \{B(g(a), h(b)) \mid B(a,b) \in I_1\},$$

which implies that $I''_1 = I''_2$.

So far we have constructed instances I'_1, I''_1, I'_2 and I''_2 that satisfies the following

$$\begin{align} \mathrm{Sol}_{\mathcal{M}_1}(I_1) &= \mathrm{Sol}_{\mathcal{M}_1}(I'_1) \tag{3}\\ \mathrm{Sol}_{\mathcal{M}_2}(I'_1) &= \mathrm{Sol}_{\mathcal{M}_2}(I''_1) \tag{4}\\ \mathrm{Sol}_{\mathcal{M}_1}(I_2) &= \mathrm{Sol}_{\mathcal{M}_1}(I'_2) \tag{5}\\ \mathrm{Sol}_{\mathcal{M}_2}(I'_2) &= \mathrm{Sol}_{\mathcal{M}_1}(I''_2) \tag{6} \end{align}$$

plus the equality $I''_1 = I''_2$.

Now consider mapping $\mathcal{M}$. Since $\mathcal{M} \preceq_{inf} \mathcal{M}_1$ we know that there exists a mapping $\mathcal{N}_1$ such that $\mathcal{M}_1 \circ \mathcal{N}_1 = \mathcal{M}$. Then given that $\mathrm{Sol}_{\mathcal{M}_1}(I_1) = \mathrm{Sol}_{\mathcal{M}_1}(I'_1)$ we know that $\mathrm{Sol}_{\mathcal{M}_1 \circ \mathcal{N}_1}(I_1) = \mathrm{Sol}_{\mathcal{M}_1 \circ \mathcal{N}_1}(I'_1)$, and thus, $\mathrm{Sol}_{\mathcal{M}}(I_1) = \mathrm{Sol}_{\mathcal{M}}(I'_1)$. Moreover, since $\mathcal{M} \preceq_{inf} \mathcal{M}_2$ we know that there exists a mapping $\mathcal{N}_2$ such that $\mathcal{M}_2 \circ \mathcal{N}_2 = \mathcal{M}$. Then, from $\mathrm{Sol}_{\mathcal{M}_2}(I'_1) = \mathrm{Sol}_{\mathcal{M}_2}(I''_1)$ we obtain that $\mathrm{Sol}_{\mathcal{M}}(I'_1) = \mathrm{Sol}_{\mathcal{M}}(I''_1)$. Thus we have that $\mathrm{Sol}_{\mathcal{M}}(I_1) = \mathrm{Sol}_{\mathcal{M}}(I'_1) = \mathrm{Sol}_{\mathcal{M}}(I''_1)$. By a symmetric argument it can be shown that $\mathrm{Sol}_{\mathcal{M}}(I_2) = \mathrm{Sol}_{\mathcal{M}}(I'_2) = \mathrm{Sol}_{\mathcal{M}}(I''_2)$. Finally, since $\mathrm{Sol}_{\mathcal{M}}(I_1) = \mathrm{Sol}_{\mathcal{M}}(I''_1)$ and $\mathrm{Sol}_{\mathcal{M}}(I_2) = \mathrm{Sol}_{\mathcal{M}}(I''_2)$, and given that $I''_1 = I''_2$, we obtain that $\mathrm{Sol}_{\mathcal{M}}(I_1) = \mathrm{Sol}_{\mathcal{M}}(I_2)$ which was to be shown. □

From this last result and Lemma 6 we obtain the following corollary.

Corollary 1. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be the mappings* (1) *and* (2)*, respectively. If $\mathcal{M}$ is a mapping which is closed under isomorphisms and such that $\mathcal{M} \preceq_{inf} \mathcal{M}_1$ and $\mathcal{M} \preceq_{inf} \mathcal{M}_2$, then $\mathcal{M}$ is invariant under target isomorphisms.*

We almost have all the necessary ingredients to prove Theorem 4. We first construct a family of mappings $\{\mathcal{C}_1, \mathcal{C}_2, \dots\}$ that we later use in the proof. Let $\mathbf{S} = \{A(\cdot,\cdot), B(\cdot,\cdot)\}$. Define for every $n \geq 1$ the sentence φ_n over $\mathbf{S}$ as:

$$\varphi_n \;:\; \exists x_1 \exists y_1 \cdots \exists x_n \exists y_n \left[\left(\bigwedge_{i=1}^{n-1} A(x_i, y_i) \wedge B(y_i, x_{i+1}) \right) \wedge A(x_n, y_n) \wedge B(y_n, x_1) \right]$$

That is, φ_n states that there exists an *AB-cycle* of length n. We consider now for every $n \geq 1$ a schema $\mathbf{T}_n = \{P_1(\cdot), \dots, P_n(\cdot)\}$, and the mapping $\mathcal{C}_n = (\mathbf{S}, \mathbf{T}_n, \Sigma_n)$ where

$$\begin{aligned} \Sigma_n = \{ \; & \varphi_1 \to \exists u \; P_1(u) \\ & \quad \vdots \\ & \varphi_n \to \exists u \; P_n(u) \; \}. \end{aligned} \tag{7}$$

Lemma 8. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be the mappings* (1) *and* (2)*, respectively. For every $n \geq 1$ it holds that $\mathcal{C}_n \preceq_{inf} \mathcal{M}_1$ and $\mathcal{C}_n \preceq_{inf} \mathcal{M}_2$.*

Proof. We consider first mapping $\mathcal{M}_1$. We need to prove that for every $n \geq 1$ there exists a mapping $\mathcal{N}_1$ such that $\mathcal{M}_1 \circ \mathcal{N}_1 = \mathcal{C}_n$ Consider for every $n \geq 1$ the formula α_n defined as:

$$\alpha_n \;:\; \exists x_1 \cdots \exists x_n \left[\left(\bigwedge_{i=1}^{n-1} T_1(x_i, x_{i+1}) \right) \wedge T_1(x_n, x_1) \right]$$

Then for the mapping $\mathcal{N}_1 = (\mathbf{T}, \mathbf{T}_n, \Gamma_n)$ with $\Gamma_n = \{\alpha_1 \to \exists u P_1(u)\ , \ldots,\ \alpha_n \to \exists u P_n(u)\}$, it is not difficult to prove that $\mathcal{M}_1 \circ \mathcal{N}_1 = \mathcal{C}_n$. To prove that $\mathcal{C}_n \preceq_{\mathit{inf}} \mathcal{M}_2$ we use a similar argument. $\square$

We now have all the ingredients to prove Theorem 4.

Proof (of Theorem 4). Consider mappings $\mathcal{M}_1$ and $\mathcal{M}_2$ specified by (1) and (2), respectively. Let $\mathcal{M}$ be a mapping from $\mathbf{S}$ to an arbitrary schema $\mathbf{T}$ such that $\mathcal{M}$ is the intersection mapping for $\mathcal{M}_1$ and $\mathcal{M}_2$ w.r.t. Rec, that is $\mathcal{M} \equiv_{\mathit{inf}} \mathcal{M}_1 \sqcap_{\text{REC}} \mathcal{M}_2$ and, in order to obtain a contradiction, assume that $\mathcal{M}$ is expressible in FO. Thus there exists a sentence Φ in FO over schema $\mathbf{S} \cup \mathbf{T}$ such that $(I, J) \in \mathcal{M}$ if and only if $(I, J) \models \Phi$. Notice that since $\mathcal{M}$ is expressible by an FO sentence, then $\mathcal{M}$ is closed under isomorphisms.

In the proof we use the notion of Ehrenfeucht-Fraïssé game which characterizes the expressibility of FO sentences [13]. Recall that $\mathbf{S} = \{A(\cdot,\cdot), B(\cdot,\cdot)\}$. Then for every n consider the instance I_{C_n} over $\mathbf{S}$ constructed as follows. Let $a_1, b_1, \ldots, a_n, b_n$ be $2n$ different elements, then I_{C_n} is the set of facts:

$$I_{C_n} = \{A(a_i, b_i) \mid 1 \leq i \leq n\} \cup \{B(b_i, a_{i+1}) \mid 1 \leq i \leq n-1\} \cup \{B(b_n, a_1)\}.$$

That is, I_{C_n} is an AB-cycle of length n. Now by the properties of Ehrenfeucht-Fraïssé games, we know that for every k there exists a value N such that the *duplicator* has a winning strategy in the k-round Ehrenfeucht-Fraïssé game over the instances I_{C_N} and $I_{C_{N+1}}$ [13]. Moreover, if K is an instance such that $\text{dom}(K) \cap \text{dom}(I_{C_N}) = \text{dom}(K) \cap \text{dom}(I_{C_{N+1}}) = \emptyset$ and the *duplicator* has a winning strategy in the k-round Ehrenfeucht-Fraïssé game over the instances I_{C_N} and $I_{C_{N+1}}$, then it is easy to see that the *duplicator* also has a winning strategy in the k-round Ehrenfeucht-Fraïssé game over the instances (I_{C_N}, K) and $(I_{C_{N+1}}, K)$ (if the *spoiler* plays a value in $\text{dom}(K)$ in some of the instances, then the *duplicator* just play the same value in the other instance, and for all values outside $\text{dom}(K)$ the *duplicator* just repeat the strategy for the game over instances I_{C_N} and $I_{C_{N+1}}$).

Now let k be the quantifier-rank of Φ, and let N be such that the *duplicator* has a winning strategy in the k-round Ehrenfeucht-Fraïssé game over the instances I_{C_N} and $I_{C_{N+1}}$. We show next that $\text{Sol}_{\mathcal{M}}(I_{C_N}) = \text{Sol}_{\mathcal{M}}(I_{C_{N+1}})$. Assume that $J \in \text{Sol}_{\mathcal{M}}(I_{C_N})$, that is $(I_{C_N}, J) \in \mathcal{M}$. Given that $\mathcal{M} \preceq_{\mathit{inf}} \mathcal{M}_1$ and $\mathcal{M} \preceq_{\mathit{inf}} \mathcal{M}_2$ and since $\mathcal{M}$ is closed under isomorphisms, by Corollary 1 we know that $\mathcal{M}$ is invariant under target isomorphisms. Let $f : \text{dom}(J) \to \mathbf{D}$ be a function that assigns a fresh element (not in $\text{dom}(I_{C_N}) \cup \text{dom}(I_{C_{N+1}}) \cup \text{dom}(J)$) to every element in $\text{dom}(J)$. Since $\mathcal{M}$ is invariant under target isomorphisms and $(I_{C_N}, J) \in \mathcal{M}$, we know that $(I_{C_N}, f(J)) \in \mathcal{M}$, and thus $(I_{C_N}, f(J)) \models \Phi$. Now since $\text{dom}(f(J)) \cap \text{dom}(I_{C_N}) = \text{dom}(f(J)) \cap \text{dom}(I_{C_{N+1}}) = \emptyset$ we know that the *duplicator* has a winning strategy in the k-round Ehrenfeucht-Fraïssé game over the instances $(I_{C_N}, f(J))$ and $(I_{C_{N+1}}, f(J))$. By the properties of Ehrenfeucht-Fraïssé games [13], this implies that Φ cannot distinguish between instances $(I_{C_N}, f(J))$ and $(I_{C_{N+1}}, f(J))$, and then since $(I_{C_N}, f(J)) \models \Phi$ we obtain that $(I_{C_{N+1}}, f(J)) \models \Phi$, implying that $(I_{C_{N+1}}, f(J)) \in \mathcal{M}$. Finally, since

$\mathcal{M}$ is invariant under target isomorphisms, we obtain that $(I_{C_{N+1}}, J) \in \mathcal{M}$ and then $J \in \mathrm{Sol}_{\mathcal{M}}(I_{C_{N+1}})$. This shows that $\mathrm{Sol}_{\mathcal{M}}(I_{C_N}) \subseteq \mathrm{Sol}_{\mathcal{M}}(I_{C_{N+1}})$. We can use a similar argument to prove that $\mathrm{Sol}_{\mathcal{M}}(I_{C_{N+1}}) \subseteq \mathrm{Sol}_{\mathcal{M}}(I_{C_N})$, and then $\mathrm{Sol}_{\mathcal{M}}(I_{C_N}) = \mathrm{Sol}_{\mathcal{M}}(I_{C_{N+1}})$. We show next that this leads to our desired contradiction.

Consider for every n the mapping $\mathcal{C}_n$ of Lemma 8. Notice that $\mathcal{C}_n$ is in REC for every n- Moreover, we know that $\mathcal{C}_n \preceq_{inf} \mathcal{M}_1$ and $\mathcal{C}_n \preceq_{inf} \mathcal{M}_2$, and then $\mathcal{C}_n \preceq_{inf} \mathcal{M}$ (since $\mathcal{M}$ is the intersection $\mathcal{M}_1 \sqcap_{\mathrm{REC}} \mathcal{M}_2$). In particular, $\mathcal{C}_{N+1} \preceq_{inf} \mathcal{M}$, which implies that there exists a mapping $\mathcal{N}$ such that $\mathcal{M} \circ \mathcal{N} = \mathcal{C}_{N+1}$. Thus, since $\mathrm{Sol}_{\mathcal{M}}(I_{C_N}) = \mathrm{Sol}_{\mathcal{M}}(I_{C_{N+1}})$, we have that $\mathrm{Sol}_{\mathcal{M}\circ\mathcal{N}}(I_{C_N}) = \mathrm{Sol}_{\mathcal{M}\circ\mathcal{N}}(I_{C_{N+1}})$ which implies that $\mathrm{Sol}_{\mathcal{C}_{N+1}}(I_{C_N}) = \mathrm{Sol}_{\mathcal{C}_{N+1}}(I_{C_{N+1}})$. Now consider the instance $J = \{T_{N+1}(a)\}$ of schema $\mathbf{T}_{N+1}$, with a an arbitrary element. Then by the definition of mapping $\mathcal{C}_{N+1}$ we know that $J \in \mathrm{Sol}_{\mathcal{C}_{N+1}}(I_{C_{N+1}})$ but $J \notin \mathrm{Sol}_{\mathcal{C}_{N+1}}(I_{C_N})$ thus contradicting the fact that $\mathrm{Sol}_{\mathcal{C}_{N+1}}(I_{C_N}) = \mathrm{Sol}_{\mathcal{C}_{N+1}}(I_{C_{N+1}})$. This completes the proof of the Theorem. $\square$

H Proof Proposition 6

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be the mappings used in the proof of Theorem 4. Recall also the mappings $\mathcal{C}_n$ used in the proof of the same theorem. Notice that for every n the mapping $\mathcal{C}_n$ is specified by st-tgds. Moreover $\mathcal{C}_n \preceq_{inf} \mathcal{M}_1$ and $\mathcal{C}_n \preceq_{inf} \mathcal{M}_2$. It was shown in the proof of Theorem 4 that there is no mapping $\mathcal{N}$ expressible in FO such that $\mathcal{C}_n \preceq_{inf} \mathcal{N}$ for every n. Thus, in particular, there is no mapping $\mathcal{N}$ specified by st-tgds such that $\mathcal{C}_n \preceq_{inf} \mathcal{N}$ for every n. This is enough to conclude that $\mathcal{M}_1 \sqcap_{\mathcal{S}} \mathcal{M}_2$ does not exist.

I Proof of Theorem 5

We will need a lemma about FO expressibility of FO-TO-CQ mappings composed with their maximum recoveries:

Lemma 9. *Let $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ be an st-mapping where Σ is a set of* FO-TO-CQ *dependencies. There exists a set Σ^* of* FO-TO-FO *dependencies from $\mathbf{S}$ to $\widehat{\mathbf{S}}$ of the form $\alpha(\boldsymbol{x}) \to \widehat{\alpha}(\boldsymbol{x})$ such that, for every maximum recovery $\mathcal{M}'$ of $\mathcal{M}$ it holds that $(I_1, I_2) \in \mathcal{M} \circ \mathcal{M}'$ iff $(I_1, \hat{I}_2) \models \Sigma^*$.*

Proof. Let $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ be an st-mapping where Σ is a set of FO-TO-CQ dependencies and let $\mathcal{M}' = (\mathbf{T}, \mathbf{S}, \Sigma')$ be the output of the algorithm MAXIMUM-RECOVERY($\mathcal{M}$). Since for every maximum recovery $\mathcal{M}''$ of $\mathcal{M}$ we know that the composition $\mathcal{M} \circ \mathcal{M}''$ equals $\mathcal{M} \circ \mathcal{M}'$, it is enough to show that the statement of the proposition holds for $\mathcal{M}\circ\mathcal{M}'$. For every $\sigma \in \Sigma$ of the form $\varphi(\boldsymbol{x}) \to \exists \boldsymbol{y} \psi(\boldsymbol{x}, \boldsymbol{y})$, let $\mathcal{C}_{\psi(\boldsymbol{x},\boldsymbol{y})}$ be the set generated in Step 2 of the algorithm. Recall that the set Σ' is built by considering the dependencies $\psi(\boldsymbol{x}, \boldsymbol{y}) \wedge C(\boldsymbol{x}) \to \alpha(\boldsymbol{x})$ where $\alpha(\boldsymbol{x})$ is the disjunction of the formulas in $\mathcal{C}_{\psi(\boldsymbol{x},\boldsymbol{y})}$. Now let Σ^* be a set of dependencies constructed as follows. For every dependency $\sigma \in \Sigma$ of the form $\varphi(\boldsymbol{x}) \to \exists \boldsymbol{y} \psi(\boldsymbol{x}, \boldsymbol{y})$

add to Σ^* the dependency $\alpha(\boldsymbol{x}) \to \widehat{\alpha}(\boldsymbol{x})$, where $\alpha(\boldsymbol{x})$ is the disjunction of all the formulas in $\mathcal{C}_{\psi(\boldsymbol{x},\boldsymbol{y})}$. We claim that for every pair of instances $I_1, I_2 \in \text{Inst}(\mathbf{S})$ it holds that $(I_1, I_2) \in \mathcal{M} \circ \mathcal{M}'$ if and only if $(I_1, \widehat{I}_2) \models \Sigma^*$.

We show first that if $(I_1, \widehat{I}_2) \models \Sigma^*$ then $(I_1, I_2) \in \mathcal{M} \circ \mathcal{M}'$. Let $(I_1, \widehat{I}_2) \in \Sigma^*$. In order to show that $(I_1, I_2) \in \mathcal{M} \circ \mathcal{M}'$ we have to prove that there exists an instance J such that $(I_1, J) \in \mathcal{M}$ and $(J, I_2) \in \mathcal{M}'$. We claim now that $\mathit{chase}_\Sigma(I_1)$ is such an instance. Let σ' be a formula in Σ' of the form $\exists \boldsymbol{y} \psi(\boldsymbol{x}, \boldsymbol{y}) \wedge C(\boldsymbol{x}) \to \alpha(\boldsymbol{x})$. We show now that $(\mathit{chase}_\Sigma(I_1), I_2) \models \sigma'$. Assume that there exists a tuple $\boldsymbol{a}$, such that $\text{chase}_\Sigma(I_1) \models \exists \boldsymbol{y} \psi(\boldsymbol{a}, \boldsymbol{y}) \wedge C(\boldsymbol{a})$ holds. We have to show that $I_2 \models \alpha(\boldsymbol{a})$. In the proof of Theorem 7.3 [4] it is proved that $(\mathit{chase}_\Sigma(I_1), I_1) \models \Sigma'$. Therefore, we know that $(\mathit{chase}_\Sigma(I_1), I_1) \models \sigma'$, and hence, $I_1 \models \alpha(\boldsymbol{a})$ must hold. Notice that $\alpha(\boldsymbol{x}) \to \widehat{\alpha}(\boldsymbol{x})$ is a dependency in Σ^* and then, since $(I_1, \hat{I}_2) \models \Sigma^*$, we have that $(I_1, \hat{I}_2) \models \alpha(\boldsymbol{x}) \to \widehat{\alpha}(\boldsymbol{x})$ and, consequently, $I_2 \models \widehat{\alpha}(\boldsymbol{a})$, which was to be shown. We have proved that if σ' is a dependency in Σ' then $(\mathit{chase}_\Sigma(I_1), I_2) \models \sigma'$, and thus, we conclude that $(\text{chase}_\Sigma(I_1), I_2) \in \mathcal{M}'$. Finally, since $(I_1, \text{chase}_\Sigma(I_1)) \in \mathcal{M}$ and $(\text{chase}_\Sigma(I_1), I_2) \in \mathcal{M}'$ we obtain that $(I_1, I_2) \in \mathcal{M} \circ \mathcal{M}'$.

We show now that if $(I_1, I_2) \in \mathcal{M} \circ \mathcal{M}'$ then $(I_1, \widehat{I}_2) \models \Sigma^*$. Let $(I_1, I_2) \in \mathcal{M} \circ \mathcal{M}'$, we have to show that for every σ in Σ^*, it holds that $(I_1, \widehat{I}_2) \models \sigma$. Notice that since $(I_1, I_2) \in \mathcal{M} \circ \mathcal{M}'$, there exists an instance, say J^*, such taht $(I_1, J^*) \in \mathcal{M}$ and $(J^*, I_2) \in \mathcal{M}'$. Now, let $\alpha(\boldsymbol{x}) \to \widehat{\alpha}(\boldsymbol{x})$ be a dependency in Σ^*, and assume that $I_1 \models \alpha(\boldsymbol{a})$ for some tuple $\boldsymbol{a}$ of elements in $\text{dom}(I_1)$. We have to show that $\widehat{I}_2 \models \widehat{\alpha}(\boldsymbol{a})$. We know that $\alpha(\boldsymbol{a})$ is the disjunction of the formulas in $\mathcal{C}_{\phi(\boldsymbol{x},\boldsymbol{y})}$, where $\exists \boldsymbol{y} \psi(\boldsymbol{x}, \boldsymbol{y})$ is the consequent of a dependency in Σ. Since $I_1 \models \alpha(\boldsymbol{a})$ we know that there exists a disjunct $\beta(\boldsymbol{x}) \in \mathcal{C}_{\psi(\boldsymbol{x},\boldsymbol{y})}$ such that $I_1 \models \beta(\boldsymbol{a})$. From Claim A.2 in the Electronic Appendix A.1 to [4] we know that $\beta(\boldsymbol{x}) \to \exists \boldsymbol{y} \psi(\boldsymbol{x}, \boldsymbol{y})$ is a logical consequence of Σ, we obtain that $J^* \models \exists \boldsymbol{y} \psi(\boldsymbol{a}, \boldsymbol{y})$. Moreover, since $\boldsymbol{a}$ is a tuple of elements in $\text{dom}(I_1)$ it holds that $J^* \models \exists \boldsymbol{y} \psi(\boldsymbol{a}, \boldsymbol{y}) \wedge C(\boldsymbol{a})$. Notice that formula $\boldsymbol{y} \psi(\boldsymbol{a}, \boldsymbol{y}) \wedge C(\boldsymbol{a}) \to \alpha(\boldsymbol{x})$ is in Σ'. Then since $(J, I_2) \in \mathcal{M}'$ and $\boldsymbol{y} \psi(\boldsymbol{a}, \boldsymbol{y}) \wedge C(\boldsymbol{a}) \to \alpha(\boldsymbol{x})$, we obtain that $I_2 \models \alpha(\boldsymbol{a})$. Thus we have that $\widehat{I}_2 \models \widehat{\alpha}(\boldsymbol{a})$ which was to be shown. We have shown that, if σ is a dependency in Σ^* then $(I_1, \widehat{I}_2) \models \sigma$ and therefore, $(I_1, I_2) \models \Sigma^*$ thus completing the proof of the lemma. □

Now we can prove the Theorem:

Proof of Theorem 5.

Let $\mathbf{S}$ be the source schema of $\mathcal{M}_1$ and $\mathcal{M}_2$, and $\widehat{\mathbf{S}}$ a copy of schema $\mathbf{S}$. Moreover, let $\mathcal{M}_1'$ and $\mathcal{M}_2'$ be respective maximum recoveries of $\mathcal{M}_1$ and $\mathcal{M}_2$. One can show (See Appendix I) that the composition $\mathcal{M}_1 \circ \mathcal{M}_1'$ can be expressed as an FO formula:

$$\forall \boldsymbol{x}_1 (\varphi_1(\boldsymbol{x}_1) \to \psi_1(\boldsymbol{x}_1)) \wedge \ldots \wedge \forall \boldsymbol{x}_n (\varphi_n(\boldsymbol{x}_n) \to \psi_n(\boldsymbol{x}_n))$$

were $\varphi_i(\boldsymbol{x}_i)$ and $\psi_i(\boldsymbol{x}_i)$ are FO formulas over $\mathbf{S}$ and $\widehat{\mathbf{S}}$, respectively. Similarly, $\mathcal{M}_2 \circ \mathcal{M}_2'$ can be expressed as $\forall \boldsymbol{y}_1 (\alpha_1(\boldsymbol{y}_1) \to \beta_1(\boldsymbol{y}_1)) \wedge \ldots \wedge \forall \boldsymbol{y}_m (\alpha_m(\boldsymbol{y}_m) \to \beta_m(\boldsymbol{y}_m))$. The formula representing the intersection is based on the construction in the proof of Theorem 3, thus we need to show how to express the transitive closure of $\mathcal{M}_1 \circ \mathcal{M}'_1 \cup \mathcal{M}_2 \circ \mathcal{M}'_2$. For this we use an intermediate schema $\widetilde{\mathbf{S}}$ constructed as follows: for every n-ary relation R of $\mathbf{S}$, we include an $(n+1)$-ary relation $\widetilde{R}$ in $\widetilde{\mathbf{S}}$. The idea is that an atom $\widetilde{R}(\boldsymbol{a}, g)$ will represent the atom $R(\boldsymbol{a})$ in the generation g of the computation of the transitive closure. Now, to define the intersection, we use an ESO formula of the form

$$\exists \widetilde{\mathbf{S}}\ \exists s\ \exists \text{zero}\ (\Omega_s \wedge \Omega_{\widetilde{R}} \wedge \Omega^E_{\widetilde{R}})$$

where $\exists\widetilde{\mathbf{S}}$ denotes an existential quantification over all relation symbols in $\widetilde{\mathbf{S}}$, s is a function symbol, and zero a first order variable. The rest of the formulas are constructed as follows: Ω_s is the formula $\forall x \forall y \big((s(x) = s(y) \to x = y) \wedge \neg(s(x) = x) \wedge \neg(s(x) = \text{zero})\big)$ that defines a successor function, with zero as the first element; $\Omega_{\widetilde{R}}$ corresponds to the following FO formula (we assume $\mathbf{S} = \{R_1, \dots, R_k\}$ and $\boldsymbol{z}_i$ is a tuple of variables of the same arity as R_i):

$$\big(\forall \boldsymbol{z}_1 (R_1(\boldsymbol{z}_1) \to \widetilde{R}_1(\boldsymbol{z}_1, \text{zero})) \wedge \cdots \wedge \forall \boldsymbol{z}_k (R_k(\boldsymbol{z_k}) \to \widetilde{R}_k(\boldsymbol{z}_k, \text{zero}))\big) \wedge$$
$$\forall g \Bigg(\bigwedge_{i=1}^{n} \forall \boldsymbol{x}_i \big[\widetilde{\varphi}_i(\boldsymbol{x}_i, g) \to \widetilde{\psi}_i(\boldsymbol{x_i}, s(g))\big] \ \vee\ \bigwedge_{i=1}^{m} \forall \boldsymbol{y}_i \big[\widetilde{\alpha}_i(\boldsymbol{y}_i, g) \to \widetilde{\beta}_i(\boldsymbol{y_i}, s(g))\big] \Bigg)$$

where $\widetilde{\varphi}_i(\boldsymbol{x}_i, g)$ is obtained from $\varphi_i(\boldsymbol{x}_i)$ by replacing every relational symbol $R(\boldsymbol{z})$ by $\widetilde{R}(\boldsymbol{z}, g)$, and $\widetilde{\psi}_i(\boldsymbol{x_i}, s(g))$ is obtained from $\psi_i(\boldsymbol{x_i})$ by replacing every relational symbol $\widehat{R}(\boldsymbol{z})$ by $\widetilde{R}(\boldsymbol{z}, s(g))$, and similarly for $\widetilde{\alpha}_i$ and $\widetilde{\beta}_i$. The intuition is that the first line initializes the relations $\widetilde{R}_i$ at generation 0, and the second line mimics a formula representing $(\mathcal{M}_1 \circ \mathcal{M}'_1 \cup \mathcal{M}_2 \circ \mathcal{M}'_2)^+$ over schema $\widetilde{\mathbf{S}}$. Finally, $\Omega^E_{\widetilde{R}}$ just extracts the target relations at some generation g of the transitive closure:

$$\exists g \big(\forall \boldsymbol{z}_1 (\widetilde{R}_1(\boldsymbol{z}_1, g) \to \widehat{R}_1(\boldsymbol{z}_1))\ \wedge \cdots \wedge\ \forall \boldsymbol{z}_k (\widetilde{R}_k(\boldsymbol{z}_k, g) \to \widehat{R}_k(\boldsymbol{z}_k))\big).$$

Before we prove the correctness of this ESO encoding, let us fix some terminology concerning the intermediate schema $\widetilde{\mathbf{S}}$:

- We say that an instance K of schema $\mathbf{S}$ resp. $\widehat{\mathbf{S}}$ is *encoded in* $\widetilde{\mathbf{S}}$ *at depth* g, if each atom $R(\boldsymbol{x})$ of K is transformed into an atom $\widetilde{R}(\boldsymbol{x}, g)$ in the instance of $\widetilde{\mathbf{S}}$.
- Conversely, given an instance $\widetilde{K}$ of $\widetilde{\mathbf{S}}$, we say that an instance K of the schema $\mathbf{S}$ resp. $\widehat{\mathbf{S}}$ is *extracted from* $\widetilde{K}$ *at depth* g, if for every atom $\widetilde{R}(\boldsymbol{x}, g) \in \widetilde{K}$, there is an atom $R(\boldsymbol{x})$ resp. $\widehat{R}(\boldsymbol{x})$ in K. (As noticed above, the formula $\Omega^E_{\widetilde{R}}$ specifies the extraction of an instance of schema $\widehat{\mathbf{S}}$, at some depth g).

The "If" direction. Let $(I, \hat{J}) \in (\mathcal{M}_1 \circ \mathcal{M}'_1 \cup \mathcal{M}_2 \circ \mathcal{M}'_2)^+$. This means that there exists some natural k, such that $(I, \hat{J}) \in (\mathcal{M}_1 \circ \mathcal{M}'_1 \cup \mathcal{M}_2 \circ \mathcal{M}'_2)^k$. Consider an instance $\widetilde{K}$ of schema $\widetilde{\mathbf{S}}$, encoding I at depth 0, and $\hat{J}$ at depth $\underbrace{s...s}_{k}(0)$. Moreover, at all other depths, from 1 to ω, we encode an instance such that the following condition is satisfied: *For each pair of instances I_1, I_2 encoded in $\widetilde{K}$ at depths g and $s(g)$ respectively, $(I_g, I_{s(g)}) \in \mathcal{M}_1 \cup \mathcal{M}_2$ holds.* It is now easy to see, that $(I, \hat{J}) \models \Omega$, since it is possible to choose the relations $\widetilde{R}_i$ as described above, and set g equal k (in unary representation).

The "Only If" direction. Let instance $(I, \hat{J})$ satisfy Ω. This means that all conjuncts in it are satisfied. Let us take $\Omega_{\widetilde{R}}$. Assuming that I is not empty, any set of relations $\widetilde{R}_1, ...\widetilde{R}_k$ (which we denote as an instance $\widetilde{K}$ of schema $\widetilde{\mathbf{S}}$) must have I encoded at depth 0. Note also, that by $\Omega_{\widetilde{R}}$, $\widetilde{K}$ must contain facts also at depths $s(0), ...$ Moreover, by construction of $\Omega_{\widetilde{R}}$, we know that any pair of instances extracted at the depths g and $s(g)$ respectively, belongs to $\mathcal{M}_1 \circ \mathcal{M}'_1 \cup \mathcal{M}_2 \circ \mathcal{M}'_2$. Finally, to satisfy the formula $\Omega^E_{\widetilde{R}}$, $\hat{J}$ must contain an instance extracted from $\widetilde{K}$ at some depth g. Thus, $(I, \hat{J}) \in (\mathcal{M}_1 \circ \mathcal{M}'_1 \cup \mathcal{M}_2 \circ \mathcal{M}'_2)^+$ follows. $\square$