The Structure of Inverses in Schema Mappings

Ronald Fagin and Alan Nash
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120

Contact email: fagin@almaden.ibm.com

Abstract

A schema mapping is a specification that describes how data structured under one schema (the source schema) is to be transformed into data structured under a different schema (the target schema). The notion of an inverse of a schema mapping is subtle, because a schema mapping may associate many target instances with each source instance, and many source instances with each target instance. In PODS 2006, Fagin defined a notion of the inverse of a schema mapping. This notion is tailored to the types of schema mappings that commonly arise in practice (those specified by “source-to-target tuple-generating dependencies”, or s-t tgds). We resolve the key open problem of the complexity of deciding whether there is an inverse. We also explore a number of interesting questions, including: What is the structure of an inverse? When is the inverse unique? How many non-equivalent inverses can there be? When does an inverse have an inverse? How big must an inverse be? Surprisingly, these questions are all interrelated. We show that for schema mappings \( M \) specified by full s-t tgds (those with no existential quantifiers), if \( M \) has an inverse, then it has a polynomial-size inverse of a particularly nice form, and there is a polynomial-time algorithm for generating it. We introduce the notion of “essential conjunctions” (or “essential atoms” in the full case), and show that they play a crucial role in the study of inverses. We use them to give greatly simplified proofs of some known results about inverses. What emerges is a much deeper understanding about this fundamental and complex operator.

Categories and Subject Descriptors: H.2.5 [Heterogeneous Databases]: Data translation; H.2.4 [Systems]: Relational databases

General Terms: Algorithms, theory

Additional Keywords and Phrases: Schema mapping, data exchange, data integration, model management, inverse, chase, dependencies, essential conjunction, essential atom

1 Introduction

Schema mappings are high-level specifications that describe the relationship between two database schemas. A schema mapping is defined to be a triple \( M = (S, T, \Sigma) \), where \( S \) (the source schema) and \( T \) (the target schema) are sequences of distinct relation symbols with no relation symbols in common, and \( \Sigma \) is a set of database dependencies that specify the association between source instances and target instances. If \( I \) is a source instance (an instance of the schema \( S \)), and \( J \) is a target instance (an instance of the schema \( T \)), then we say that \( J \) is a solution for \( I \) if the pair \((I, J)\) together satisfies \( \Sigma \). We sometimes identify the schema mapping \( M \) with the set
of pairs \((I, J)\) that satisfy \(\Sigma\), and write \((I, J) \in \mathcal{M}\). The most widely studied case arises when \(\Sigma\) is a finite set of source-to-target tuple-generating dependencies (s-t tgds). We refer to a schema mapping specified by a finite set of s-t tgds as an s-t tgd mapping. These mappings have also been used in data integration scenarios under the name of GLAV (global-and-local-as-view) assertions [Len02]. Our main focus in this article is on inverses for s-t tgd mappings.

1.1 Motivation and History

Since schema mappings form the essential building blocks of such crucial data inter-operability tasks as data exchange and data integration (see the surveys [Kol05, Len02]), several different operators on schema mappings have been singled out as deserving study in their own right [Ber03]. The composition operator and the inverse operator have emerged as two of the most fundamental operators on schema mappings. The composition operator is important, because we wish to understand what schema mapping is obtained by first applying one schema mapping and then another schema mapping. The inverse operator is important, because we wish to know how to “undo” the effects of a schema mapping (intuitively, to go back and retrieve the original data). Both composition and inversion arise naturally in the study of schema evolution [BM07].

The composition operator has been investigated in depth [FKPT05, MH03, Mel04, NBM07]; however, progress on the study of the inverse operator was not made until recently. Even finding the exact semantics of this operator is a delicate task, since unlike the traditional use of the name “mapping”, a schema mapping is not simply a function that maps an instance of the source schema to an instance of the target schema. Instead, for each source instance, the schema mapping may associate many target instances. Furthermore, for each target instance, there may be many corresponding source instances.

Fagin [Fag07] gave a formal definition of what it means for a schema mapping \(\mathcal{M}'\) to be an inverse of a schema mapping \(\mathcal{M}\). The intuition behind his approach is that the result of applying first \(\mathcal{M}\) and then \(\mathcal{M}'\) (that is, the composition \(\mathcal{M} \circ \mathcal{M}'\)) should be the identity mapping. In Fagin’s (and our) case of special interest, where \(\mathcal{M}\) is an s-t tgd mapping, there is a complication with this intuition. Fagin showed that if \(\mathcal{M}\) is an s-t tgd mapping, then there is no schema mapping \(\mathcal{M}'\) (s-t tgd mapping or otherwise) such that \(\mathcal{M} \circ \mathcal{M}'\) equals the standard identity mapping (which, intuitively, is the mapping that consists precisely of the pairs \((I, J)\) where \(J = I\)). He showed that in a precise sense, the closest that \(\mathcal{M} \circ \mathcal{M}'\) can come to the standard identity mapping is for \(\mathcal{M} \circ \mathcal{M}'\) to be the copy mapping, which is specified by s-t tgds that “copy” the source instance to the target instance (we shall define the copy mapping formally later). Therefore, he defined \(\mathcal{M}'\) to be an inverse of \(\mathcal{M}\) if \(\mathcal{M} \circ \mathcal{M}'\) is the copy mapping. He showed how to construct an inverse of an s-t tgd mapping that is itself an s-t tgd mapping when such an inverse exists. He also developed a number of tools for the study of inverses of s-t tgd mappings. He showed that deciding invertibility of an s-t tgd mapping is coNP-hard, and left open the question as to whether it is even decidable. We give a matching coNP upper bound, which shows that deciding invertibility is coNP-complete. Since Fagin’s coNP-hardness lower bound holds even when the s-t tgds that specify the schema mapping are full (have no existential quantifiers), it follows that deciding invertibility in the full case is also coNP-complete.

Fagin, Kolaitis, Popa and Tan [FKPT08] observed that most schema mappings that arise in practice do not have an inverse. Therefore, they introduced and studied a principled relaxation of the notion of an inverse of a schema mapping, which they called a quasi-inverse. Intuitively, it is obtained from the notion of an inverse by not differentiating between instances \(I\) and \(I'\) that have the same set of solutions, and so are equivalent for data-exchange purposes. We shall not discuss the formal details of quasi-inverses here. Instead, we simply note that they showed that if \(\mathcal{M}\) is an s-t tgd mapping, then every inverse of \(\mathcal{M}\) is a quasi-inverse of \(\mathcal{M}\), but there are s-t tgd mappings with a quasi-inverse but no inverse.

Arenas, Pérez, and Riveros [APR08] defined another relaxation of the notion of inverse. If \(\mathcal{M}\) and \(\mathcal{M}'\) are schema mappings, they say that \(\mathcal{M}'\) is a recovery of \(\mathcal{M}\) if \((I, I) \in \mathcal{M} \circ \mathcal{M}'\) for every source instance \(I\). They say that \(\mathcal{M}'\) is a maximum recovery of \(\mathcal{M}\) if \(\mathcal{M}'\) is a recovery of \(\mathcal{M}\), and if \(\mathcal{M} \circ \mathcal{M}' \subseteq \mathcal{M} \circ \mathcal{M}''\) for every recovery \(\mathcal{M}''\) of \(\mathcal{M}\). Intuitively, \(\mathcal{M}'\) is a maximum recovery of \(\mathcal{M}\) if \(\mathcal{M} \circ \mathcal{M}'\) comes as close as possible to being the standard identity mapping. Arenas, Pérez, and Riveros showed that if \(\mathcal{M}\) is an s-t tgd mapping, then every inverse
of $\mathcal{M}$ is a maximum recovery of $\mathcal{M}$. They also showed the key result that every s-t tgd mapping has a maximum recovery (which contrasts with the fact that there are s-t tgd mappings with no inverse).

Although we did not explicitly say this earlier, all of the work we just described (on inverses, quasi-inverses, and maximum recoveries) were based on the assumption that we restrict our attention to source instances that are ground (contain no null values). Fagin, Kolaitis, Popa and Tan [FKPT09] defined new notions of inverse and of recovery, called extended inverses and extended recoveries (the latter gives a corresponding notion of a maximum extended recovery), that they argue are more appropriate notions to use when the source instances may contain null values. They showed that every s-t tgd mapping that is extended invertible is invertible, but not conversely. Intuitively, it is harder for a schema mapping to be extended invertible, since there are more instances (i.e., non-ground source instances) to consider. Fagin, Kolaitis, Popa and Tan showed that when source instances may contain nulls, then there are s-t tgd mappings that do not have a maximum recovery, under the definition of maximum recovery by Arenas, Pérez, and Riveros in [APR08], but that every s-t tgd mapping has a maximum extended recovery. In this article, we allow only ground source instances (in fact, as we shall explain shortly, many of the issues considered in this article are natural only under the assumption that source instances are restricted to being ground).

When it comes to “flavors of inverses” (that is, Fagin’s notion of inverse, Fagin, Kolaitis, Popa and Tan’s notions of quasi-inverse and of maximum extended recovery, and Arenas, Pérez, and Riveros’ notion of maximum recovery), Fagin’s notion of inverse is the gold standard. It is the hardest to attain, and it is automatically an inverse in all of the other flavors. This is why we feel that it is worthwhile to investigate the structure of inverses, which we do in this article.

1.2 The Language of Inverses

For s-t tgd mappings, Fagin [Fag07] focused only on inverses specified by tgds, and left open the problem of characterizing the language needed to express inverses of s-t tgd mappings. Fagin, Kolaitis, Popa and Tan [FKPT08] resolved this problem. Specifically, they gave an algorithm for constructing a canonical candidate inverse for an s-t tgd mapping. It is specified by using what they called s-t tgds with constants and inequalities. These are like s-t tgds, but there may also be formulas $\text{const}(x)$ (which say that $x$ is a constant) and inequalities in the premise. They showed that if an s-t tgd mapping is invertible, then its canonical candidate inverse is indeed an inverse.

We define normal inverses, that are specified by special cases of s-t tgds with constants and inequalities. The canonical candidate inverse is a normal inverse. Hence, if an s-t tgd mapping has an inverse, then it has a normal inverse. It is not hard to see that this would not be true if we were to allow non-ground source instances. This is why we consider only ground source instances, in the tradition of the earlier articles [APR08, Fag07, FKPT08]. Normal inverses are especially nice, in that if $I$ is a source instance, $\mathcal{M}$ is an s-t tgd mapping specified by $\Sigma$, and $\mathcal{M}'$ is a normal inverse of $\mathcal{M}$ that is specified by $\Sigma'$, then the result of chasing $I$ with $\Sigma$ and then chasing the result by $\Sigma'$ gives back exactly $I$ (this is not true of arbitrary inverses, even where the chase is well-defined).

We focus our study mainly on normal inverses.

1.3 Our Contributions

In addition to our result mentioned earlier where we resolve the complexity of the deciding if an s-t tgd mapping is invertible, we obtain a number of other new results about inverses, that we now discuss.

**Essential conjunctions and atoms.** We introduce the notions of essential conjunctions (and, in the full case, of essential atoms), which turn out to play a fundamental role for the study of inverses. Roughly speaking, an essential conjunction for a relational atom $A$ (with respect to an s-t tgd mapping) is a conjunction such that (a) the atoms in the conjunction arise in the chase of $A$, and (b) if all of these atoms arise together in a chase, then $A$ is present in the source. We show that an s-t tgd mapping is invertible if and only if each atom has an essential conjunction (in the full case, if and only if each atom has an essential atom). Further, we show how to construct a normal inverse directly from the essential conjunctions.
Unique inverses. For most notions of “inverse” that arise in mathematics, if there is an inverse, then it is unique. However, as we show, no schema mapping has a unique inverse. What about a unique normal inverse? This is possible, and we give a characterization of those s-t tgd mappings with a unique normal inverse.

In the full case (where the s-t tgd have no existential quantifiers) there is an especially interesting story. Let us say that a full s-t tgd mapping \( M = (S, T, \Sigma) \) is onto if every target instance is the result of chasing some source instance with \( \Sigma \). We show that if a full s-t tgd mapping is invertible and onto, then it has a unique normal inverse. What about the converse? We show that the converse fails. What if we enrich the language of possible inverses? Following [FKPT08], we define disjunctive tgd with inequalities by allowing inequalities in the premise and disjunctions in the conclusion (such mappings were shown to be necessary to express quasi-inverses of full s-t tgd mappings in [FKPT08]). We show that a full s-t tgd mapping \( M \) has a unique inverse specified by disjunctive tgd with inequalities if and only if \( M \) is invertible and onto. Furthermore, we show that \( M \) satisfies these conditions if and only if \( M \) is equivalent to a slight generalization of the copy mapping, called a p-copy mapping.

Inverse of an inverse. For most notions of “inverse” that arise in mathematics, if \( M' \) is an inverse of \( M \), then \( M \) is an inverse of \( M' \). This is because \( M' \) is a right-inverse of \( M \) (that is, \( M \circ M' \) is the identity) if and only if \( M \) is a left-inverse of \( M' \), and most notions of inverse that arise in mathematics are 2-sided (every right-inverse is a left-inverse, and vice-versa). An inverse of a schema mapping, as defined by Fagin in [Fag07], is only a right-inverse. In particular, it does not follow that if \( M' \) is an inverse of a schema mapping \( M \), then \( M \) is an inverse of \( M' \). In fact, it does not even follow that if \( M' \) is an inverse of a schema mapping, then \( M' \) is invertible. Surprisingly, it turns out to be rare that a normal inverse of an s-t tgd mapping is itself invertible. We show that \( M \) is a full s-t tgd mapping with an invertible normal inverse if and only if \( M \) is again equivalent to a p-copy mapping. By combining this result with our results about unique inverses, we obtain the unexpected result that several nice properties of a full s-t tgd mapping are equivalent: \( M \) has an invertible normal inverse if and only if \( M \) has a unique inverse specified by disjunctive tgd with inequalities if and only if \( M \) is invertible and onto if and only if \( M \) is equivalent to a p-copy mapping. We also show that these nice properties are not equivalent if we remove the restriction that \( M \) be full.\(^1\)

The size of an inverse, and the complexity of computing an inverse. How big does a normal inverse need to be? We show that there is a family of full, invertible s-t tgd mappings \( M \) such that the size of the smallest normal inverse of \( M \) is exponential in the size of \( M \). Therefore, we broaden the class of normal mappings by allowing not just inequalities but also Boolean combinations of equalities in the premises, and we call these mappings Boolean normal. Allowing Boolean normal mappings does not increase the expressive power of normal mappings, but allows a more compact representation. Indeed, we show that every invertible full s-t tgd mapping has a Boolean normal inverse of polynomial size. And in fact, we give a polynomial-time algorithm for generating this Boolean normal inverse.

Is there a relationship between the number of normal inverses and the size of the minimal Boolean normal inverse? We cannot bound the number of normal inverses in terms of the size of the minimal Boolean normal inverse, since there are examples with an infinite number of inequivalent normal inverses. However, we show that if there are only a small number of inequivalent normal inverses, then the minimal number of constraints in a Boolean normal inverse is small. Specifically, we show that if \( M \) is a full s-t tgd mapping, with \( k \) source relation symbols and with exactly \( m \geq 1 \) inequivalent normal inverses, then \( M \) has a Boolean normal inverse with at most \( k + \log_2(m) \) constraints.

Simpler proofs of known results. We give greatly simplified proofs of two results whose previous proofs were quite complex. First, we give a simple proof of the result in [FKPT08] that for invertible s-t tgd mappings \( M \), the canonical candidate inverse of \( M \) is indeed an inverse of \( M \). We now discuss the second result where we give a greatly simplified proof. Fagin [Fag07] introduced the unique-solutions property, which says that no two distinct source instances have the same set of solutions. He showed that the unique-solutions property is a necessary condition for a schema mapping to have an inverse. He gave a complicated proof that for LAV mappings (those specified by s-t tgd with a singleton premise), the unique-solutions property is not only a necessary condition but also a sufficient condition for invertibility. We give a simple proof of this result.

\(^1\)We do not define the property of being onto when \( M \) is not full. In fact, we remark that it is not completely clear what the “right” definition of onto should be in the nonfull case.
2 Preliminaries

In this section, we give a number of definitions. Most of these definitions are from [FKMP05], which we refer the reader to for an introduction to data exchange.

Schemas and Schema Mappings. A schema $R$ is a finite sequence $(R_1, \ldots, R_k)$ of relation symbols, each of a fixed arity. An instance $I$ over $R$ (which we may call an $R$-instance, or simply an instance, when $R$ is understood), is a sequence $(R_1^I, \ldots, R_k^I)$, where each $R_i^I$ is a finite relation of the same arity as $R_i$. We shall often use $R_i$ to denote both the relation symbol and the relation $R_i^I$ that interprets it.

A schema mapping is a triple $M = (S, T, \Sigma)$ consisting of a source schema $S$, a target schema $T$, and a set $\Sigma$ of constraints (defined shortly). We say that $M$ is specified by $\Sigma$. If $\Sigma$ is a finite set of $s$-$t$ tgds (defined shortly), then we may refer to $M$ as an $s$-$t$ tgd mapping. When $S$ and $T$ are clear from context, we will sometimes say $\Sigma$ when we should say $(S, T, \Sigma)$, and talk about a set of constraints, when we should talk about a schema mapping.

Instances and Formulas. From now on, we assume that $S$ and $T$ are two fixed schemas. We call $S$ the source schema and $T$ the target schema. We refer to $S$-instances as source instances, and $T$-instances as target instances. Let $C$ be a fixed countably infinite set of constants and let $N$ be a fixed countably infinite set of nulls that is disjoint from $C$. We assume that all source instances have individual values (that is, individual entries of tuples) from the set $C$ of constants only, while all target instances have individual values from $C \cup N$. We may sometimes refer to $S$-instances as ground instances to emphasize the fact that all individual values in such instances are constants. Intuitively, schema mappings of the form $M = (S, T, \Sigma)$ model the situation in which we perform data exchange from $S$ to $T$: the individual values of source instances are known, while incomplete information in the specification of data exchange may give rise to null values in the target instances. We write $\text{dom}(I)$ for the (active) domain of an instance $I$, that is, the set of all individual values that appear in $I$.

If $P$ is an $m$-ary relation symbol in $S$, and $x_1, \ldots, x_m$ are variables, not necessarily distinct, then $P(x_1, \ldots, x_m)$ is a relational atom, or simply atom (over $S$). We may refer to it as a $P$-atom. In the context of a schema mapping $M = (S, T, \Sigma)$, we may refer to a $P$-atom where $P$ is in $S$ as a source atom, and a $P$-atom where $P$ is in $T$ as a target atom. If $P$ is an $m$-ary relation symbol in $S$, and $c_1, \ldots, c_m$ are values (constants or nulls), not necessarily distinct, then $P(c_1, \ldots, c_m)$ is a fact (over $S$). We may refer to it as a $P$-fact. We sometimes identify an instance with its set of facts. If $I_1$ and $I_2$ are instances, and the set of facts of $I_1$ is a subset of the set of facts of $I_2$, then we may write $I_1 \subseteq I_2$, and say that $I_1$ is a subinstance of $I_2$.

There is a special unary relation symbol $\text{const}$ (that is not part of any schema). We refer to the formula $\text{const}(x)$ for a variable $x$ as a const formula; the intended interpretation of $\text{const}$ is that $\text{const}(x)$ should hold precisely if $x$ is a constant.

If $\delta$ is a conjunction of relational atoms (but no const formulas), then we define $I_\delta$ to be an instance obtained from $\delta$ as follows. For each variable $v$, assign a distinct fixed constant $c_v$, and let the facts of $I_\delta$ consist of the facts $P(c_{v_1}, \ldots, c_{v_k})$ where $P(v_1, \ldots, v_k)$ is an atom in $\delta$. For example, if $\delta = P(x, y) \land Q(y)$, then $I_\delta$ is the instance $\{P(c_x, c_y), Q(c_y)\}$. If $\delta$ is a conjunction of relational atoms and at least one const formula, then we define $I_\delta$ as follows. For each variable $v$ such that $\text{const}(v)$ is in $\delta$, assign a distinct fixed constant $c_v$, and for each remaining variable $v$ assign a distinct fixed null $n_v$. Define $I_\delta$ be the facts that result by taking each relational atom in $\delta$ and doing a replacement using the assignment we just described. For example, if $\delta = P(x, y) \land Q(y) \land \text{const}(x)$, then $I_\delta$ is the instance $\{P(c_x, n_y), Q(n_y)\}$. It is sometimes convenient to allow $\delta$ to contain also inequalities $x \neq y$, where $x$ and $y$ are distinct variables among the variables appearing in relational atoms of $\delta$. In that case, we simply ignore the inequalities in defining $I_\delta$. Note that if $\delta$ and $\delta'$ contain the same relational atoms, and $\delta$ has no const formula, while $\delta'$ contains the formulas $\text{const}(x)$ for every variable $x$ in the relational atoms of $\delta'$, then $I_\delta$ and $I_{\delta'}$ are the same, and contain only constants (no nulls). Throughout this paper, we reserve the symbol $\delta$ (possibly with subscripts or primes) to be a conjunction of relational atoms, const formulas $\text{const}(x)$ for variables $x$, and inequalities $x \neq y$ for distinct variables $x$ and $y$.

A renaming of variables is a one-to-one function that maps variables to variables. A weak renaming of variables is a function (not necessarily one-to-one) that maps variables to variables. We may sometimes refer to a renaming as a strict renaming, to distinguish it from a weak renaming.
Define a \textit{prime atom} to be one that contains precisely the variables $x_1, x_2, \ldots, x_k$ for some $k$, and where the initial appearance of $x_i$ precedes the initial appearance of $x_j$ if $i < j$. For example, $P(x_1, x_2, x_3, x_4)$ is a prime atom, but $Q(x_2, x_1)$ and $R(x_2, x_3)$ are not. Note that for every relational atom, there is a unique renaming of variables to obtain a prime atom.

\textbf{Constraints.} All sets of constraints we consider are finite, unless otherwise specified. We consider constraints of several forms. A \textit{source-to-target tuple-generating dependency (s-t tgd)} is a constraint of the form $\forall \bar{x} \forall \bar{y} (\alpha(\bar{x}, \bar{y}) \rightarrow \exists \bar{z} \beta(\bar{x}, \bar{z}))$, where $\alpha$ is a conjunction of source atoms and $\beta$ is a conjunction of target atoms. The variables in $\bar{x}$ are exactly the variables that appear in both $\alpha$ and $\beta$. Note in particular that this gives us the safety condition that every variable in $\beta$ that is not existentially quantified appears also in $\alpha$. We will generally suppress writing the $\forall \bar{x} \forall \bar{y}$ part. If $\bar{z}$ is empty, we say that $\varphi$ is \textit{full}. We use the standard notion of \textit{satisfaction} of constraints, denoted $\models$. For example, if $I$ is a source instance, $J$ is a target instance, and $\Sigma$ is a set of s-t tgds, then we may write $(I, J) \models \Sigma$ to mean that the pair $(I, J)$ satisfies the members of $\Sigma$.

\textbf{Homomorphisms.} Let $J$, $J'$ be two instances. A function $h$ that maps values to values is a \textit{homomorphism} from $J$ to $J'$ if for every constant $c$, we have that $h(c) = c$, and for every relation symbol $R$ and each tuple $(a_1, \ldots, a_n) \in R^J$, we have that $(h(a_1), \ldots, h(a_n)) \in R^{J'}$. We write $J \rightarrow J'$ if there is a homomorphism from $J$ to $J'$. The instances $J$ and $J'$ are said to be \textit{homomorphically equivalent} if there are homomorphisms from $J$ to $J'$ and from $J'$ to $J$. We then write $J \leftrightarrow J'$.

\textbf{Solutions and Universal Solutions.} Let $M = (S, T, \Sigma)$ be a schema mapping. Then $J$ is a \textit{solution} for $I$ (under $M$) if $(I, J) \models \Sigma$. We write $\text{Sol}(M, I)$ to denote the solutions for $I$ under $M$. We say that a solution $U$ for the source instance $I$ is an \textit{universal solution} [FKMP05] if $U \rightarrow J$ for every solution $J$ for $I$.

\textbf{Composition and Inverse.} We recall the concept of the \textit{composition} of two schema mappings, introduced in [FKPT05, Mel04], and the concept of an \textit{inverse} of a schema mapping, introduced in [Fag07].

Let $M_{12} = (S_1, S_2, \Sigma_{12})$ and $M_{23} = (S_2, S_3, \Sigma_{23})$ be schema mappings. The \textit{composition} $M_{12} \circ M_{23}$ is a schema mapping $(S_1, S_3, \Sigma_{13})$ such that for every $S_1$-instance $I$ and every $S_3$-instance $J$, we have that $(I, J) \models \Sigma_{13}$ if and only if there is an $S_2$-instance $K$ such that $(I, K) \models \Sigma_{12}$ and $(K, J) \models \Sigma_{23}$. As noted in [FKPT05], the composition is unique. When the schemas are understood from the context, we will often write $\Sigma_{12} \circ \Sigma_{23}$ for the composition $M_{12} \circ M_{23}$.

Let $\tilde{S}$ be a replica of the source schema $S$; that is, for every relation symbol $R$ of $S$, the schema $\tilde{S}$ contains a relation symbol $\hat{R}$ that is not in $S$ and has the same arity as $R$. We also assume that $\hat{R}$ and $S$ are distinct when $R$ and $S$ are distinct. If $A$ is a relational atom $R(x_1, \ldots, x_k)$, then $\hat{A}$ is the relational atom $\hat{R}(x_1, \ldots, x_k)$. Similarly, if $F$ is a fact $R(c_1, \ldots, c_k)$, then $\hat{F}$ is the fact $\hat{R}(c_1, \ldots, c_k)$. If $I$ is an instance of $S_1$, define $\hat{I}$ to be the corresponding instance of $S_1$. Thus, $\hat{I}$ consists precisely of the facts $\hat{F}$ such that $F$ is a fact of $I$.

The \textit{copy mapping} is the schema mapping $\text{Id} = (S, \tilde{S}, \Sigma_{1\text{id}})$, where $\Sigma_{1\text{id}}$ consists of the s-t tgds $R(x_1, \ldots, x_k) \rightarrow \hat{R}(x_1, \ldots, x_k)$, where $R$ is $k$-ary, and where $\hat{R}$ ranges over the relation symbols in $\tilde{S}$. Thus, if $I_1$ is an $S$-instance and $I_2$ is an $\tilde{S}$-instance, then $(I_1, I_2) \models \Sigma_{1\text{id}}$ if and only if $\hat{I}_1 \subseteq I_2$.

Let $M_{12} = (S_1, S_2, \Sigma_{12})$ be a schema mapping. We say that a schema mapping $M_{21} = (S_2, \hat{S}_1, \Sigma_{21})$ is an \textit{inverse} of $M_{12}$ if $M_{12} \circ M_{21} = \text{Id}$, that is, the result of composing $M_{12}$ with $M_{21}$ is the copy mapping. So $M_{12}$ is an inverse of $M_{12}$ precisely if for every ground $S_1$-instance $I$ and every ground $\hat{S}_1$-instance $J$, we have that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ if and only if $\hat{I} \subseteq J$. If $M_{12}$ has an inverse, then we say that $M_{12}$ is \textit{invertible}.

\textbf{Chasing.} If $M_{12} = (S_1, S_2, \Sigma_{12})$ is an s-t tgd mapping, then \textit{chasing}, or applying the \textit{chase process} [MMS79] to the $S_1$-instance $I$ with $\Sigma_{12}$ produces an $S_2$-instance $U$ such that $U$ is a universal solution for $I$ under $M_{12}$ [FKMP05]. We may write $U = \text{chase}_{12}(I)$, and say that $U$ is \textit{the} result of the chase. For definiteness, we use the version of the chase as defined in [FKPT05], although it does not really matter, since whatever version of the chase we use, the results are all homomorphically equivalent. Similarly, we may write $\text{chase}_{21}(I)$ for the result of chasing an $S_2$-instance $I$ with $\Sigma_{21}$. We shall also extend this notation to cases where $\Sigma_{12}$ or $\Sigma_{21}$ are not simply sets of s-t tgds, but where we also allow \textit{const} formulas and inequalities in the premises.
3 Deciding Invertibility

In [Fag07] it is shown that deciding invertibility of s-t tgd mappings is coNP-hard, and it was left open as to whether it is even decidable. In this section, we prove a matching coNP upper bound, which shows that deciding invertibility of s-t tgd mappings is coNP-complete. Since the co-NP lower bound of [Fag07] holds also in the full case, this shows that deciding invertibility of full s-t tgd mappings is also coNP-complete.

It is not hard to verify (see [FKPT08]) that if \( M_{12} \) is an s-t tgd mapping, and \( I \) and \( I' \) are source instances where \( I \subseteq I' \), then \( \text{Sol}(M_{12}, I') \subseteq \text{Sol}(M_{12}, I) \). If the opposite implication also necessarily holds, then \( M_{12} \) is said to have the subset property [FKPT08]. Thus, an s-t tgd mapping \( M_{12} \) has the subset property if whenever \( I \) and \( I' \) are source instances where \( \text{Sol}(M_{12}, I') \subseteq \text{Sol}(M_{12}, I) \), then \( I \subseteq I' \). It was shown in [FKPT08] that the subset property (which they called the (=,=)-subset property) is a necessary and sufficient condition for invertibility of an s-t tgd mapping. We shall make use of the following property, which we call the “homomorphic version” of the subset property, and which we shall show is equivalent to the subset property: whenever \( I \) and \( I' \) are source instances where \( \text{chase}_{12}(I) \rightarrow \text{chase}_{12}(I') \), then \( I \subseteq I' \).

To prove our next result, we shall make use of the following simple lemma by Fagin, Kolaitis, Miller, and Popa [FKMP05].

**Lemma 3.1** [FKMP05] If \( M_{12} = (S_1, S_2, \Sigma_{12}) \) is an s-t tgd mapping, then the solutions for a ground instance \( I \) are exactly the target instances \( J \) such that \( \text{chase}_{12}(I) \rightarrow J \).

Recall that if \( A \) is the relational atom \( P(v_1, \ldots, v_k) \), then \( I_A \) is an instance that consists of the single fact \( P(c_{v_1}, \ldots, c_{v_k}) \). We shall make use of the following proposition.

**Proposition 3.2** For an s-t tgd mapping \( M_{12} = (S_1, S_2, \Sigma_{12}) \), the following are equivalent:

1. \( M_{12} \) is invertible.
2. Whenever \( I \) and \( I' \) are ground instances where \( \text{Sol}(M_{12}, I') \subseteq \text{Sol}(M_{12}, I) \), then \( I \subseteq I' \) (the subset property).
3. Whenever \( I \) and \( I' \) are ground instances where \( \text{chase}_{12}(I) \rightarrow \text{chase}_{12}(I') \), then \( I \subseteq I' \) (the homomorphic version of the subset property).
4. For every relational atom \( A \) and ground instance \( I \),
   \[
   \text{chase}_{12}(I_A) \rightarrow \text{chase}_{12}(I) \text{ implies } I_A \subseteq I.
   \]
5. For every relational atom \( A \) and ground instance \( I \) with at most \( n \) facts,
   \[
   \text{chase}_{12}(I_A) \rightarrow \text{chase}_{12}(I) \text{ implies } I_A \subseteq I
   \]
   where \( n \) is the number of facts in \( \text{chase}_{12}(I_A) \).

**Proof** The equivalence of (1) and (2) was shown in [FKPT08]. We now show that (2) and (3) are equivalent. Assume first that (2) holds. To prove that (3) holds, assume that \( \text{chase}_{12}(I) \rightarrow \text{chase}_{12}(I') \); we must show that \( I \subseteq I' \). By (2), it is sufficient to show that \( \text{Sol}(M_{12}, I') \subseteq \text{Sol}(M_{12}, I) \). Assume that \( J \in \text{Sol}(M_{12}, I') \); we must show that \( J \in \text{Sol}(M_{12}, I) \). Since \( J \in \text{Sol}(M_{12}, I') \), it follows from Lemma 3.1 that \( \text{chase}_{12}(I') \rightarrow J \). Since also \( \text{chase}_{12}(I) \rightarrow \text{chase}_{12}(I') \), it follows by transitivity of homomorphism that \( \text{chase}_{12}(I) \rightarrow J \). It then follows from Lemma 3.1 that \( J \in \text{Sol}(M_{12}, I) \), as desired.

Assume now that (3) holds. To prove that (2) holds, assume that \( \text{Sol}(M_{12}, I') \subseteq \text{Sol}(M_{12}, I) \); we must show that \( I \subseteq I' \). By (3), it is sufficient to show that \( \text{chase}_{12}(I) \rightarrow \text{chase}_{12}(I') \). Now \( \text{chase}_{12}(I') \in \text{Sol}(M_{12}, I') \). Therefore, since \( \text{Sol}(M_{12}, I') \subseteq \text{Sol}(M_{12}, I) \), it follows that \( \text{chase}_{12}(I') \in \text{Sol}(M_{12}, I) \). So by Lemma 3.1, it follows that \( \text{chase}_{12}(I) \rightarrow \text{chase}_{12}(I') \), as desired.
We now show that (3) and (4) are equivalent. It is clear that (3) implies (4), since (4) is a special case of (3) where \( I_A \) plays the role of \( I \), and \( I \) plays the role of \( I' \). To prove that (4) implies (3), we shall show that if (3) fails, then (4) fails. Assume that (3) fails. Therefore, there are ground instances \( I \) and \( I' \) such that \( \text{chase}_{12}(I) \rightarrow \text{chase}_{12}(I') \) and \( I \nsubseteq I' \). Since \( I \nsubseteq I' \), we can assume (by renaming constants if needed) that there is an atom \( A \) such that \( I_A \subseteq I \) but \( I_A \not\subseteq I' \). Since \( I_A \subseteq I \), we have \( \text{chase}_{12}(I_A) \rightarrow \text{chase}_{12}(I) \). Since also \( \text{chase}_{12}(I) \rightarrow \text{chase}_{12}(I') \), it follows by transitivity of homomorphism that \( \text{chase}_{12}(I_A) \rightarrow \text{chase}_{12}(I') \), witnessing that (4) fails (where \( I' \) plays the role of \( I \)).

Finally, we show that (4) and (5) are equivalent. It is clear that (4) implies (5). Assume now that (5) holds; we shall show that (4) holds. Assume that \( \text{chase}_{12}(I_A) \rightarrow \text{chase}_{12}(I) \). Then necessarily also \( \text{chase}_{12}(I_A) \rightarrow \text{chase}_{12}(I') \) for some \( I' \subseteq I \) with at most \( n \) facts, since \( \text{chase}_{12}(I_A) \) maps into at most \( n \) facts in \( \text{chase}_{12}(I) \). So by (5), we have that \( I_A \subseteq I \). Therefore, (4) holds. □

Proposition 3.2 gives us a very simple proof of the desired coNP upper bound on the problem of deciding invertibility of \( s \rightarrow t \) tgd mappings, as the next proof shows.

**Theorem 3.3** The problem of deciding if an \( s \rightarrow t \) tgd mapping is invertible is coNP-complete.

**Proof** The proof of coNP-hardness is in [Fag07]. We now show the coNP upper bound. We make use of the equivalence of (1) and (5) in Proposition 3.2. To check that \( M_{12} = (S_1, S_2, \Sigma_{12}) \) is not invertible, guess a relational atom \( A \), a ground instance \( I \) such that \( I_A \not\subseteq I \) where \( I \) has at most \( n \) facts, and a homomorphism \( h : \text{chase}_{12}(I_A) \rightarrow \text{chase}_{12}(I) \), where \( n \) is the number of facts in \( \text{chase}_{12}(I_A) \). □

Since Fagin’s coNP-hardness lower bound holds even when the schema mapping is full and LAV (that is, when each of the \( s \rightarrow t \) tgd that specify the schema mapping is full and has a singleton premise), we obtain the following corollary.

**Corollary 3.4** The following complexity results hold.

1. The problem of deciding if a full \( s \rightarrow t \) tgd mapping is invertible is coNP-complete.

2. The problem of deciding if a LAV \( s \rightarrow t \) tgd mapping is invertible is coNP-complete.

3. The problem of deciding if a full and LAV \( s \rightarrow t \) tgd mapping is invertible is coNP-complete.

## 4 Normal Mappings

In this section, we study a class of mappings (that we call *normal*), which are an especially attractive choice for inverses of \( s \rightarrow t \) tgd mappings. If an \( s \rightarrow t \) tgd mapping \( \mathcal{M} \) has an inverse, then it has a normal inverse, because (a) the canonical candidate inverse (defined later) is normal, and (b) if \( \mathcal{M} \) has an inverse, then the canonical candidate inverse of \( \mathcal{M} \) is indeed an inverse of \( \mathcal{M} \) [FKPT08]. Since we are interested in inverses \( \mathcal{M}_{21} = (S_2, S_1, \Sigma_{21}) \) of an \( s \rightarrow t \) tgd mapping \( \mathcal{M}_{12} = (S_1, S_2, \Sigma_{12}) \), the normal mappings of interest to us have source schema \( S_2 \) and target schema \( S_1 \).

**Definition 4.1** A constraint is normal if it is of the form \( \alpha \land \chi_A \land \eta \rightarrow A \), where \( \alpha \) is a conjunction of source atoms, \( A \) is a target atom, \( \chi_A \) is the conjunction of the formulas \( \text{const}(x) \) for every variable \( x \) of \( A \), and \( \eta \) is a conjunction (possibly empty) of inequalities of the form \( x \neq y \) for distinct variables \( x, y \) of \( A \). Further, there is the safety condition that every variable in \( A \) must appear in \( \alpha \). As usual, we have suppressed writing the leading universal quantifiers. A schema mapping is said to be normal if all of its constraints are normal. □
Notice that we require the \( \text{const} \) predicate on all variables in \( A \), but just allow inequalities on variables in \( A \). In particular, \( \chi_A \) is fully determined by \( A \) (which is why we write it with the subscript \( A \)), whereas \( \eta \) is not determined by \( A \) (and can even be empty). Note also that every normal constraint is full (has no existential quantifiers).

An example of a normal constraint is

\[
P(x_1, x_2, x_3, x_3) \land \text{const}(x_2) \land \text{const}(x_3) \land (x_2 \neq x_3) \rightarrow Q(x_2, x_3, x_2) \tag{1}
\]

If the formula \( \text{const}(x_1) \) were added to the premise of (1), then the resulting constraint would not be normal, since the variable \( x_1 \) does not appear in the conclusion of (1). If either of the formulas \( \text{const}(x_2) \) or \( \text{const}(x_3) \) were removed from (1), then the resulting constraint would not be normal, since the variables \( x_2 \) and \( x_3 \) appear in the conclusion of (1). If the inequality \( x_2 \neq x_3 \) were removed from (1), then the resulting constraint would still be normal, since inequalities are not required. If the inequality \( x_1 \neq x_2 \) were added to the premise of (1), then the resulting constraint would not be normal, since the variable \( x_1 \) does not appear in the conclusion of (1).

Let \( M_{12} = (S_1, S_2, \Sigma_{12}) \) and \( M_{21} = (S_2, \tilde{S}_1, \Sigma_{21}) \) be schema mappings. Let us say that \( \Sigma_{21} \) is too strong (for \( M_{12} \)) if there is a ground \( S_1 \)-instance \( I \) and a ground \( \tilde{S}_1 \)-instance \( J \) such that \( I \subseteq J \) but \( (I, J) \not\models \Sigma_{12} \circ \Sigma_{21} \). So \( \Sigma_{21} \) is not too strong precisely if whenever there is a ground \( S_1 \)-instance \( I \) and a ground \( \tilde{S}_1 \)-instance \( J \) such that \( I \subseteq J \), then \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \). If \( \Sigma_{21} \) is a set of s-t tgds, and \( \Sigma_{21} \) is arbitrary, then it follows from Proposition 5.2 of [Fag07] that \( \Sigma_{21} \) is not too strong precisely if \( (I, \tilde{I}) \models \Sigma_{12} \circ \Sigma_{21} \) for every ground \( S_1 \)-instance \( I \).

**Example 4.2** Let \( S_1 \) consist of the binary relation symbol \( P \) and the unary relation symbol \( T \). Let \( S_2 \) consist of the binary relation symbol \( P' \) and the unary relation symbol \( Q \) and \( T' \). Let \( \Sigma_{12} = \{ P(x, y) \rightarrow P'(x, y), P(x, x) \rightarrow Q(x), T(x) \rightarrow T'(x), T(x) \rightarrow P'(x, x) \} \), and let \( M_{12} = (S_1, S_2, \Sigma_{12}) \). It was shown in [FKPT08] that \( M_{12} \) is invertible, and that the inverse “requires inequalities” (details are in [FKPT08]). Let \( \Sigma_{21} \) consists only of the constraint \( P(x, y) \rightarrow P'(x, y) \). We now show that \( \Sigma_{21} \) is too strong for \( M_{12} \). Let \( I = \{ T(0) \} \), and let \( J = \tilde{I} \). So of course \( \tilde{I} \subseteq J \) (in fact, we have equality). However, we now show that \( (I, J) \not\models \Sigma_{12} \circ \Sigma_{21} \), which shows that \( \Sigma_{21} \) is too strong. Assume by way of contradiction that \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \). Then there is \( K \) such that \( (I, K) \models \Sigma_{12} \) and \( (K, J) \models \Sigma_{21} \). Since \( (I, K) \models \Sigma_{12} \), we know that \( K \) contains the fact \( P'(0, 0) \). Therefore, since \( (K, J) \models \Sigma_{21} \), we have that \( J \) contains \( \tilde{P}(0,0) \), which contradicts the fact that \( J = \{ \tilde{T}(0) \} \). This is the desired contradiction. So indeed, \( \Sigma_{21} \) is too strong for \( M_{12} \). Now let \( \Sigma_{21}' = \{ P'(x, y) \land (x \neq y) \rightarrow P(x, y) \} \) (so that \( \Sigma_{21}' \) is obtained from \( \Sigma_{21} \) by adding the inequality \( x \neq y \) to the premise). Then (as we now show), \( \Sigma_{21}' \) is not too strong for \( M_{12} \). In fact, if \( I \) is a ground \( S_1 \)-instance \( I \) and \( J \) is a ground \( \tilde{S}_1 \)-instance such that \( \tilde{I} \subseteq J \), and if \( U = \text{chase}_{S_1}(I) \), then \( (I, U) \models \Sigma_{12} \), and it is not hard to see that \( (U, J) \models \Sigma_{21} \). Hence, \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \), as desired. \( \square \)

Let us say that \( \Sigma_{21} \) is too weak (for \( M_{12} \)) if there is a ground \( S_1 \)-instance \( I \) and a ground \( \tilde{S}_1 \)-instance \( J \) such that \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \) but \( \tilde{I} \not\subseteq J \). So \( \Sigma_{21} \) is not too weak precisely if whenever there is a ground \( S_1 \)-instance \( I \) and a ground \( \tilde{S}_1 \)-instance \( J \) such that \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \), then \( \tilde{I} \subseteq J \).

**Example 4.3** Let \( M_{12} \) be as in Example 4.2. Let \( \Sigma_{21}' \) the empty set. Then \( (I, J) \models \Sigma_{12} \circ \Sigma_{21}' \) for every choice of \( I \) and \( J \), and it follows easily that \( \Sigma_{21}' \) is too weak for \( M_{12} \). As a more interesting example, let \( \Sigma_{21}' \) be as in Example 4.2. We now show that \( \Sigma_{21}' \) is too weak for \( M_{12} \). Let \( I = \{ T(0) \} \), let \( J \) be the empty set, and let \( K = \{ T'(0), P'(0, 0) \} \). It is easy to see that \( (I, K) \models \Sigma_{12} \), and \( (K, J) \models \Sigma_{21}' \). So \( (I, J) \models \Sigma_{12} \circ \Sigma_{21}' \). However, \( \tilde{I} \not\subseteq J \). So indeed, \( \Sigma_{21}' \) is too weak for \( M_{12} \). \( \square \)

The following simple proposition follows immediately from our definitions.

**Proposition 4.4** Let \( M_{12} = (S_1, S_2, \Sigma_{12}) \) and \( M_{21} = (S_2, \tilde{S}_1, \Sigma_{21}) \) be schema mappings. Then \( M_{21} \) is an inverse of \( M_{12} \) if and only if \( \Sigma_{21} \) is not too strong and not too weak for \( M_{12} \).
Example 4.5 Let $M_{12}$ be as in Examples 4.2 and 4.3. Let $\Sigma_{21} = \{P'(x, y) \land (x \neq y) \rightarrow \bar{P}(x, y), Q(x) \rightarrow \bar{P}(x, x), T'(x) \rightarrow \bar{T}(x)\}$, and let $M_{21} = (S_2, \hat{S}_1, \Sigma_{21})$. It can be shown that $M_{21}$ is an inverse of $M_{12}$, and so $\Sigma_{21}$ is not too strong and not too weak for $M_{12}$. □

If $\Sigma_{21}$ is not too strong, then for every ground $S_1$-instance $I$ and every ground $\hat{S}_1$-instance $J$ where $\hat{I} \subseteq J$, there is an instance $K$ “in the middle” such that $(I, K) \models \Sigma_{12}$ and $(K, J) \models \Sigma_{21}$. We may say that $K$ witnesses that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$. The next proposition says that if $M_{21}$ is a normal inverse of $M_{12}$, then any universal solution can play the role of this witness. This is a quite useful as a tool in proving properties of normal inverses.

Proposition 4.6 Assume $M_{21} = (S_2, \hat{S}_1, \Sigma_{21})$ is a normal inverse of the s-t tgd mapping $M_{12} = (S_1, S_2, \Sigma_{12})$. Let $I$ be a ground $S_1$-instance, and let $U$ be an arbitrary universal solution for $I$ with respect to $M_{12}$. Then $(U, \hat{I}) \models \Sigma_{21}$, and $U$ witnesses $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ when $\hat{I} \subseteq J$.

Proof Since $M_{21}$ is an inverse of $M_{12}$, we know that $(I, \hat{I}) \models \Sigma_{12} \circ \Sigma_{21}$, and therefore there exists some $K$ such that $(I, K) \models \Sigma_{12}$ and $(K, \hat{I}) \models \Sigma_{21}$. Let $U$ be an arbitrary universal solution for $I$ with respect to $M_{12}$. Then there is a homomorphism $h : U \rightarrow K$ that is the identity on all values that appear in $I$ (since $I$ is ground). Pick a constraint $\varphi \in \Sigma_{21}$; by our normality assumption, it must be of the form

$$\alpha(\bar{x}, \bar{y}) \land \chi_A(\bar{x}) \land \eta(\bar{x}) \rightarrow \bar{A}(\bar{x}).$$

Assume that $U$ satisfies the premise of $\varphi$ on $\bar{a}, \bar{b}$. Then $K \models \alpha(h(\bar{a}), h(\bar{b})).$ Since $U \models \chi_A(\bar{a})$, we have $h(\bar{a}) = \bar{a}$ and therefore $K \models \alpha(\bar{a}, h(\bar{b})).$ Also, since $U \models \eta(\bar{a})$, we have $K \models \eta(\bar{a}).$ So $K$ satisfies the premise of $\varphi$ on $\bar{a}, h(\bar{b})$. Hence, since $(K, \hat{I}) \models \Sigma_{21}$, we must have $\hat{I} \models \bar{A}(\bar{a})$. This shows that $(U, \hat{I}) \models \varphi$. Since $\varphi$ is an arbitrary member of $\Sigma_{21}$, it follows that $(U, \hat{I}) \models \Sigma_{21}$, as desired. Since $(U, \hat{I}) \models \Sigma_{21}$ and $\hat{I} \subseteq J$, it follows easily that $(U, J) \models \Sigma_{21}$. Since $(I, U) \models \Sigma_{12}$ and $(U, J) \models \Sigma_{21}$, we have that $U$ witnesses $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$, as desired. □

We now give an example that shows that if $M_{21}$ is not normal, then there may be no universal solution for $I$ that witnesses $(I, \hat{I}) \models \Sigma_{12} \circ \Sigma_{21}$.

Example 4.7 Let $S_1$ consist of the unary relation symbol $S$, and let $S_2$ consist of the binary relation symbol $T$. Let $\Sigma_{12}$ consist of the single s-t tgd $S(x) \rightarrow \exists y(T(x, y) \land T(y, x))$, and let $\Sigma_{21}$ consist of the single s-t tgd $T(x, y) \land T(y, x) \rightarrow S(x)$. Note that this latter s-t tgd is not a normal constraint, since it does not have the formula $\text{const}(x)$ in its premise. Let $M_{12} = (S_1, S_2, \Sigma_{12})$ and $M_{21} = (S_2, \hat{S}_1, \Sigma_{21})$.

Let $I$ be an arbitrary nonempty ground instance, and let $U$ be an arbitrary universal solution for $I$ with respect to $\Sigma_{12}$. Assume that $S(c)$ is a fact of $I$. Then $U$ must contain facts $T(c, n)$ and $T(n, c)$ for some $n$. If $(U, J) \models \Sigma_{21}$, then necessarily $J$ contains the fact $\hat{S}(n)$, which is not a fact of $\hat{I}$. So $U$ does not witness $(I, \hat{I}) \models \Sigma_{12} \circ \Sigma_{21}$. However, if we take $K$ to be an instance that contains precisely the facts $T(c, c)$ such that $S(c)$ is a fact of $I$, then it is easy to see that $K$ witnesses $(I, \hat{I}) \models \Sigma_{12} \circ \Sigma_{21}$. It is then straightforward to see that $M_{21}$ is an inverse of $M_{12}$. Thus, $M_{21}$ is a (non-normal) inverse of $M_{12}$ where no universal solution witnesses $(I, \hat{I}) \models \Sigma_{12} \circ \Sigma_{21}$.

Let $M_{21}' = (S_2, \hat{S}_1, \Sigma_{21}')$, where $\Sigma_{21}'$ consists of the single normal constraint $T(x, y) \land T(y, x) \land \text{const}(x) \rightarrow \hat{S}(x)$. Then $M_{21}'$ is a normal inverse of $M_{12}$, and as Proposition 4.6 tells us, every universal solution for $I$ witnesses $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$. □

It is easy to see that the chase process can be applied for normal mappings. The final theorem of this section gives an elegant characterization of normal inverses of s-t tgd mappings. Specifically, it says that $M_{21}$ is an inverse of $M_{12}$ if and only if $\text{chase}_{21}(\text{chase}_{12}(I)) = \hat{I}$ for every ground instance $I$. To prove this theorem, we need two lemmas, that characterize when a normal mapping is not too strong for an s-t tgd mapping, and when a normal mapping is not too weak for an s-t tgd mapping.

---

2If $\bar{a} = (a_1, \ldots, a_r)$, then by $h(\bar{a})$, we mean $(h(a_1), \ldots, h(a_r))$. 

---

10
Lemma 4.8 Let $\mathcal{M}_{12} = (S_1, S_2, \Sigma_{12})$ be an s-t tgd mapping and $\mathcal{M}_{21} = (S_2, \widehat{S}_1, \Sigma_{21})$ be a normal mapping. Then $\Sigma_{21}$ is not too strong if and only if $\text{chase}_{21}(\text{chase}_{12}(I)) \subseteq \hat{I}$ for every ground instance $I$.

Proof Assume first that $\text{chase}_{21}(\text{chase}_{12}(I)) \subseteq \hat{I}$ for every ground instance $I$. We must show that whenever there are ground instances $I$ and $J$ such that $\hat{I} \subseteq J$, then $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$. Let $I$ and $J$ be ground instances such that $\hat{I} \subseteq J$. Let $U = \text{chase}_{12}(I)$, and let $U' = \text{chase}_{21}(U)$. So $(I, U) \models \Sigma_{12}$ and $(U, U') \models \Sigma_{21}$. Also, by assumption, $U' \subseteq \hat{I}$. Since also $\hat{I} \subseteq J$, it follows that $U' \subseteq J$. Since $(U, U') \models \Sigma_{21}$ and $U' \subseteq J$, we see that $(I, J) \models \Sigma_{21}$. Since $(I, U) \models \Sigma_{12}$ and $(U, J) \models \Sigma_{21}$, it follows that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$, as desired.

Assume now that $\Sigma_{21}$ is not too strong. So $(I, \hat{I}) \models \Sigma_{12} \circ \Sigma_{21}$. Let $U = \text{chase}_{12}(I)$, and let $U' = \text{chase}_{21}(U)$. We must show that $U' \subseteq \hat{I}$. The argument in the proof of Proposition 4.6 shows that $(U, \hat{I}) \models \Sigma_{21}$. Since $\mathcal{M}_{21}$ is a normal mapping, it is easy to see that the result of chasing an arbitrary instance with $\Sigma_{21}$ is a ground instance. In particular, $U'$ is a ground instance. Since $U'$ is the result of chasing $U$ with $\Sigma_{21}$, and $U'$ is ground, a standard property of the chase tells us that for every instance $J$ such that $(U, J) \models \Sigma_{21}$, necessarily $U' \subseteq J$. If we take $J$ to be $\hat{I}$, then we see that $U' \subseteq \hat{I}$, as desired. \qed

Lemma 4.9 Let $\mathcal{M}_{12} = (S_1, S_2, \Sigma_{12})$ be an s-t tgd mapping and $\mathcal{M}_{21} = (S_2, \widehat{S}_1, \Sigma_{21})$ be a normal mapping. Then $\Sigma_{21}$ is not too weak if and only if $\text{chase}_{21}(\text{chase}_{12}(I)) \subseteq \hat{I}$ for every ground instance $I$.

Proof Assume first that $\Sigma_{21}$ is not too weak. So whenever there are ground instances $I$ and $J$ such that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$, then $\hat{I} \subseteq J$. Let $I$ be $\text{chase}_{21}(\text{chase}_{12}(I))$. Then $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$. So $\hat{I} \subseteq J$, that is, $\hat{I} \subseteq \text{chase}_{21}(\text{chase}_{12}(I))$, as desired.

Assume now that $\hat{I} \subseteq \text{chase}_{21}(\text{chase}_{12}(I))$ for every ground instance $I$. Let $I$ and $J$ be ground instances such that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$; we must show that $\hat{I} \subseteq J$. Since $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$, there is $K$ such that $(I, K) \models \Sigma_{12}$ and $(K, J) \models \Sigma_{21}$. Since $(I, K) \models \Sigma_{12}$, a standard property of the chase tells us that $\text{chase}_{12}(I) \rightarrow K$. Similarly, $\text{chase}_{21}(K) \rightarrow J$. Since $\text{chase}_{21}(I) \rightarrow K$, we see (by chasing both sides with $\Sigma_{21}$) that $\text{chase}_{21}(\text{chase}_{12}(I)) \rightarrow \text{chase}_{21}(K)$. Hence, since also $\text{chase}_{21}(K) \rightarrow J$, we have $\text{chase}_{21}(\text{chase}_{12}(I)) \rightarrow J$, by transitivity of homomorphism. Since also $\hat{I} \subseteq \text{chase}_{21}(\text{chase}_{12}(I))$, it follows that $\hat{I} \rightarrow J$. Therefore, since $\hat{I}$ and $J$ are both ground, we have $\hat{I} \subseteq J$, as desired. \qed

We now discuss a nice property of normal inverses. Corollary 7.4 of [Fag07] says that if $\mathcal{M}_{12} = (S_1, S_2, \Sigma_{12})$ and $\mathcal{M}_{21} = (S_2, \widehat{S}_1, \Sigma_{21})$ are both full s-t tgd mappings, then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ if and only if $\text{chase}_{21}(\text{chase}_{12}(I)) = \hat{I}$ for every ground instance $I$. The next theorem says that this strong property (that $\text{chase}_{21}(\text{chase}_{12}(I)) = \hat{I}$ for every ground instance $I$) holds for normal inverses $\mathcal{M}_{21}$, even when $\mathcal{M}_{12}$ is not full.

Theorem 4.10 Let $\mathcal{M}_{12} = (S_1, S_2, \Sigma_{12})$ be an s-t tgd mapping and $\mathcal{M}_{21} = (S_2, \widehat{S}_1, \Sigma_{21})$ a normal mapping. Then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ if and only if $\text{chase}_{21}(\text{chase}_{12}(I)) = \hat{I}$ for every ground instance $I$.

Proof This follows easily from Proposition 4.4, along with Lemmas 4.8 and 4.9. \qed

Theorem 4.10 fails if we drop the assumption that $\mathcal{M}_{21}$ be normal. In particular, it is straightforward to verify that if $\mathcal{M}_{12}$ and $\mathcal{M}_{21}$ are as in Example 4.7, and $I$ is an arbitrary nonempty ground instance, then $\text{chase}_{21}(\text{chase}_{12}(I)) \not\subseteq \hat{I}$. It is more challenging to find an example where $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ but $\hat{I} \not\subseteq \text{chase}_{21}(\text{chase}_{12}(I))$. Such an example was given in [FKPT08], as we now describe.

Example 4.11 We now give an example from [FKPT08] where $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ but where there is a ground instance $I$ such that $\hat{I} \not\subseteq \text{chase}_{21}(\text{chase}_{12}(I))$. Let $S_1$ consist of the unary relation symbol $P$, and let $S_2$ consist of the binary relation symbol $Q$. Let $\Sigma_{12}$ consist of $P(x) \rightarrow \exists y Q(x, y)$, and let $\Sigma_{21}$ consist
of the constraints \( Q(x, y) \rightarrow P(y) \) and \( Q(x, y) \land \text{const}(y) \rightarrow P(x) \). Let \( M_{12} = (S_1, S_2, \Sigma_{12}) \) and \( M_{21} = (S_2, S_1, \Sigma_{21}) \). Note that \( M_{21} \) is not normal, because the second member of \( \Sigma_{21} \) has \( \text{const}(y) \) rather than \( \text{const}(x) \) in its premise.

We now show that \( M_{21} \) is an inverse of \( M_{12} \). To do this, we need to show that if \( I \) and \( J \) are ground instances, then \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \) if and only if \( \hat{I} \subseteq J \).

First, let \( I \) be a ground instance that consists of \( n \) facts \( P(c_1), \ldots, P(c_n) \), and let \( K \) be \( \{Q(c_i, c_i) : 1 \leq i \leq n\} \). It is easy to see that \( (I, K) \models \Sigma_{12} \) and \( (K, \hat{I}) \models \Sigma_{21} \). Hence, \( (I, \hat{I}) \models \Sigma_{12} \circ \Sigma_{21} \), which implies that if \( \hat{I} \subseteq J \), then \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \). Next, assume that \( I \) and \( J \) are ground instances such that \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \); we shall show that \( \hat{I} \subseteq J \). Since \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \), there is \( K \) such that \( (I, K) \models \Sigma_{12} \) and \( (K, J) \models \Sigma_{21} \). Suppose \( I \) consists of \( n \) facts \( P(c_1), \ldots, P(c_n) \). Since \( (I, K) \models \Sigma_{12} \), we know that \( K \) contains \( \{Q(c_i, y_i) \mid 1 \leq i \leq n\} \), for some choices of \( y_1, \ldots, y_n \). There are two cases:

- **Case 1.** Some \( y_i \) is not a constant. Because of the first constraint in \( \Sigma_{21} \), we see that \( J \) contains \( P(y_i) \), and so \( J \) is not ground. Hence, this case is not possible.
- **Case 2.** Every \( y_i \) is a constant. Because of the second constraint in \( \Sigma_{21} \), we see that \( J \) contains \( P(c_i) \), for \( 1 \leq i \leq n \), and so \( I \subseteq J \), as desired.

This concludes the proof that \( M_{21} \) is an inverse of \( M_{12} \).

Now let \( I = \{P(0)\} \). We have that \( \text{chase}_{12}(I) = \{Q(0, n)\} \) for a null \( n \). Then \( \text{chase}_{21}(\text{chase}_{12}(I)) = \{\hat{P}(n)\} \). So \( \hat{I} \not\subseteq \text{chase}_{21}(\text{chase}_{12}(I)) \). □

### 5 Essential Conjunctions and Essential Atoms

In this section, we introduce the notions of **essential conjunctions** (and, in the full case, of **essential atoms**), which turn out to play a fundamental role for the study of inverses. Roughly speaking, an essential conjunction for a relational atom \( A \) (with respect to an s-t tgd mapping \( M_{12} = (S_1, S_2, \Sigma_{12}) \)) is a conjunction such that (a) the atoms in the conjunction arise in the chase of \( A \) with \( \Sigma_{12} \), and (b) all of these atoms can arise together in a chase with \( \Sigma_{12} \) only if \( A \) is present in the source. We show that an s-t tgd mapping is invertible if and only if each atom has an essential conjunction (in the full case, if and only if each atom has an essential atom). Further, we show how to construct a normal inverse directly from the essential conjunctions. It is convenient to consider separately the two parts (a) and (b) above. When a conjunction satisfies (a), that is, roughly speaking, the atoms in the conjunction arise in the chase of \( A \), then we say that the conjunction is **relevant**. When a conjunction satisfies (b), that is, roughly speaking, all of the atoms in the conjunction can arise together in a chase with \( \Sigma_{12} \) only if \( A \) is present in the source, then we say that the conjunction is **demanding**.

We now say more precisely what we mean for a conjunction \( \delta \) to be relevant for a source atom \( A \). The definition is given in terms of the existence of a certain homomorphism. In addition to having \( \delta \) contain target atoms, we also allow \( \delta \) to contain \( \text{const} \) formulas \( \text{const}(x) \) for some or all of the variables \( x \) in \( A \), where the \( \text{const} \) formulas \( \text{const}(x) \) in \( \delta \) tell us what variables must be mapped onto themselves in the homomorphism. This way, instead of simply requiring that the target atoms in \( \delta \) be in the chase of \( A \), we can be a little more general, and instead require only that a homomorphic image of the target atoms of \( \delta \) be in the chase of \( A \). This added level of generality is needed in our characterizations we give later of when normal mappings are inverses. We follow the simplifying convention established earlier that if \( \delta \) contains no \( \text{const} \) formulas, this is treated the same as having it contain all of the \( \text{const} \) formulas \( \text{const}(x) \) for every variable \( x \) of \( A \), in that the homomorphism must map each variable onto itself. In addition to target atoms and \( \text{const} \) formulas, we also allow \( \delta \) to contain inequalities, for notational simplicity later, even though we ignore the inequalities in our definitions. Since the chase works on instances, not on atoms, we make use of \( I_A \) and \( I_A \), as defined earlier, rather than \( \delta \) and \( A \) directly. We now give the formal definition of a relevant conjunction \( \delta \) for a source atom \( A \).

---

3Note that our choice of \( \Sigma_{21} \), although clearly peculiar (given what \( \Sigma_{12} \) is), is what we intended: it is not a typographical error!
Definition 5.1 Let $\Sigma_{12}$ be a finite set of s-t tgds. Assume that $A$ is a relational atom, and $\delta$ is a conjunction $\alpha \land \chi \land \eta$, where $\alpha$ is a conjunction of relational atoms, $\chi$ is a conjunction (possibly empty) of const formulas $\text{const}(x)$ for variables $x$ in $A$, and $\eta$ is a conjunction (possibly empty) of inequalities of the form $x \neq y$ for distinct variables $x, y$ in $A$. Let us say that $\delta$ is relevant for $A$ (with respect to $\Sigma_{12}$) if $I_\delta \rightarrow \text{chase}_{12}(I_A)$. Note that the inequalities play no role, but are allowed for notational convenience. □

Full case: We now examine the notion of “relevant” in the case when $\Sigma_{12}$ is full. The const formulas play no role in inverses of full s-t tgds mappings, as shown in [FKPT08] (see Proposition 7.11 for a precise statement of what we mean by “play no role”). Therefore, it is natural to assume in the full case that the relevant conjunction $\delta$ has no const formulas. Then $I_\delta$ contains only constants (no nulls). So we have that $I_\delta \rightarrow \text{chase}_{12}(I_A)$ holds if and only if $I_\delta \subseteq \text{chase}_{12}(I_A)$. Hence, $\delta$ is relevant for $A$ if and only if $I_\delta \subseteq \text{chase}_{12}(I_A)$. This corresponds very well with the intuition we gave earlier that $\delta$ is relevant for $A$ if and only if “the atoms in $\delta$ arise in the chase of $A$”.

We now consider examples of the notion of being relevant.

Example 5.2 We begin with examples in the full case. Let $\Sigma_{12}$ consist of the full s-t tgds $P(x, y) \rightarrow (R(x, y) \land S(x, x) \land T(y))$ and $Q(x) \rightarrow S(x, x)$. Let $A$ be $P(x, y)$. Then $I_A$ is $\{P(c_x, c_y)\}$, where $c_x$ and $c_y$ are constants. Also, we have that $\text{chase}_{12}(I_A)$ is $\{R(c_x, c_y), S(c_x, c_x), T(c_y)\}$. Let $I_\delta$ be $R(x, y) \land S(x, x)$, let $I_\delta$ be $S(x, y)$, and let $I_{\delta_3}$ be $S(x, x)$. Then we have that $I_{\delta_3} = \{R(c_x, c_y), S(c_x, c_x)\}$, that $I_{\delta_3} = \{S(c_x, c_y)\}$, and that $I_{\delta_4}$ is $\{S(c_x, c_x)\}$. So $I_\delta$ and $I_{\delta_3}$ are both relevant for $A$, whereas $I_{\delta_4}$ is not relevant for $A$. □

Example 5.3 Assume that $\Sigma_{12}$ consists of the s-t tgds $P(x) \rightarrow \exists z(R(x, z) \land S(x, x))$ and $Q(x) \rightarrow S(x, x)$. Let $A$ be $P(x)$. Then $I_A$ is $\{P(c_x)\}$, where $c_x$ is a constant, and $\text{chase}_{12}(I_A)$ is $\{R(c_x, n), S(x, c_x)\}$ for a null $n$. Let $I_\delta$ be $R(x, y) \land \text{const}(x)$. Then $I_{\delta_1}$ is $\{R(c_x, n_y)\}$ for a null $n_y$. So $I_{\delta_1} \rightarrow \text{chase}_{12}(I_A)$ under the homomorphism $h_1$ where $h_1(c_x) = c_x$ and $h_1(n_y) = n$. Therefore, $\delta_1$ is relevant for $A$. Now let $I_{\delta_2}$ be $R(x, y)$. Then $I_{\delta_2}$ is $\{R(c_x, c_y)\}$ for constants $c_x, c_y$. So $\delta_2$ is not relevant for $A$. Since if there were a homomorphism $h_2 : I_{\delta_2} \rightarrow \text{chase}_{12}(I_A)$, we would have $h_2(c_y) = n$, which is not possible. Finally, let $I_{\delta_3}$ be $S(x, y) \land \text{const}(x)$. Then $I_{\delta_3}$ is $\{S(c_x, n_y)\}$ for a null $n_y$. So $I_{\delta_3} \rightarrow \text{chase}_{12}(I_A)$ under the homomorphism $h_3$ where $h_3(c_x) = c_x$ and $h_3(n_y) = c_x$. Therefore, $\delta_3$ is relevant for $A$. □

We now consider the notion of a demanding conjunction. Roughly speaking, the conjunction $\delta$ is demanding for the source atom $A$ if “all of the atoms in $\delta$ can arise together in a chase only if $A$ is present in the source”. We now give the formal definition of a demanding conjunction $\delta$ for a source atom $A$.

Definition 5.4 Let $\Sigma_{12}, A$, and $\delta$ be as in Definition 5.1. Let us say that $\delta$ is demanding for $A$ (with respect to $\Sigma_{12}$) if for every ground instance $I$ such that $I_\delta \rightarrow \text{chase}_{12}(I_A)$, necessarily $I_A \subseteq I$. □

Full case: We now examine the notion of “demanding” when $\Sigma_{12}$ is full. As in our discussion of the full case for the notion of “relevant”, we assume that $\delta$ has no const formulas. As before, $I_\delta$ then contains only constants (no nulls). So we have that $I_\delta \rightarrow \text{chase}_{12}(I_A)$ holds if and only if $I_\delta \subseteq \text{chase}_{12}(I_A)$. Hence, $\delta$ is demanding for $A$ if and only if whenever $I_\delta \subseteq \text{chase}_{12}(I_A)$, then $I_A \subseteq I$. This corresponds very well to the intuition we gave earlier that $\delta$ is demanding for $A$ if “all of the atoms in $\delta$ can arise together in a chase only if $A$ is present in the source”. We now consider examples of the notion of being demanding.

Example 5.5 We begin with examples in the full case. Let $\Sigma_{12}, A, \delta_1, \delta_2$, and $\delta_3$ be as in Example 5.2. We now show that $\delta_1$ is demanding for $A$. As before, $I_{\delta_1}$ is $\{R(c_x, c_y), S(c_x, c_x)\}$ and $I_A$ is $\{P(c_x, c_y)\}$, where $c_x$ and $c_y$ are constants. Assume that $I_{\delta_1} \subseteq \text{chase}_{12}(I_A)$, that is, $\{R(c_x, c_y), S(c_x, c_x)\} \subseteq \text{chase}_{12}(I_A)$. In particular, $\{R(c_x, c_y)\} \subseteq \text{chase}_{12}(I_A)$. By looking at $\Sigma_{12}$, we see that the only way this can happen is if $\{P(c_x, c_y)\} \subseteq I$. That is, $I_A \subseteq I$, as desired.

It is easy to see that there is no instance $I$ such that $I_{\delta_2} \rightarrow \text{chase}_{12}(I_A)$. Therefore, it is automatically true that $\delta_2$ is demanding for $A$. However, we now show that $\delta_3$ is not demanding for $A$. Let $I$ be $\{Q(c_x)\}$. It is easy to see that $I_{\delta_3}$ and $\text{chase}_{12}(I_A)$ are each $\{S(c_x, c_x)\}$ for the constant $c_x$, so $I_{\delta_3} = \text{chase}_{12}(I)$. Therefore, $I_{\delta_3} \subseteq \text{chase}_{12}(I)$. However, $I_A \nsubseteq I$. So $\delta_3$ is not demanding for $A$. □
Example 5.6 Let \( \Sigma_{12}, A, \delta_1, \delta_2, \) and \( \delta_3 \) be as in Example 5.3. We now show that \( \delta_1 \) is demanding for \( A \). Let \( I \) be a ground instance such that \( I_{\delta_1} \rightarrow \text{chase}_{12}(I) \); we must show that \( I_A \subseteq I \). Assume that the distinct \( P \)-facts in \( I \) are \( P(c_1), \ldots, P(c_k) \) (also, \( I \) may contain some \( Q \)-facts). Then the \( R \)-facts in \( \text{chase}_{12}(I) \) are precisely \( R(c_1, n_1), \ldots, R(c_k, n_k) \) for distinct nulls \( n_1, \ldots, n_k \). Let \( h_1 \) be a homomorphism such that \( h_1 : I_{\delta_1} \rightarrow \text{chase}_{12}(I) \) (such a homomorphism exists by assumption). Since \( I_{\delta_1} \) is \( \{R(c_x, n_y)\} \) for a null \( n_y \), and since \( h_1(c_x) = c_x \), it follows easily that one of \( c_1, \ldots, c_k \) is \( c_x \). So \( P(c_x) \) is a fact of \( I \). But \( I_A \) is \( \{P(c_x)\} \). Hence, \( I_A \subseteq I \), as desired. So indeed, \( \delta_1 \) is demanding for \( A \).

By a similar argument to that given in Example 5.3, we see that there is no instance \( I \) such that \( I_{\delta_2} \rightarrow \text{chase}_{12}(I) \). Therefore, it is automatically true that \( \delta_2 \) is demanding for \( A \).

Finally, we show that \( \delta_3 \) is not demanding for \( A \). As in Example 5.3, we have that \( I_{\delta_3} \) is \( \{S(c_x, n_y)\} \) for a null \( n_y \). Let \( I \) be \( \{Q(c_x, c_x)\} \). So \( \text{chase}_{12}(I) \) is \( \{S(c_x, c_x)\} \). Therefore, \( I_{\delta_3} \rightarrow \text{chase}_{12}(I) \). However, \( I_A \not\subseteq I \). So \( \delta_3 \) is not demanding for \( A \). \( \square \)

We can now put together the notions of “relevant” and “demanding” to obtain the notion we really want, that of an essential conjunction.

Definition 5.7 Let \( \Sigma_{12}, A, \) and \( \delta \) be as in Definition 5.1. We say that \( \delta \) is essential for \( A \) (with respect to \( \Sigma_{12} \)) if \( \delta \) is both relevant for \( A \) and demanding for \( A \) (with respect to \( \Sigma_{12} \)). \( \square \)

Example 5.8 Let \( \Sigma_{12}, A, \delta_1, \delta_2, \) and \( \delta_3 \) be as in Examples 5.2 and 5.5. We see from these examples that \( \delta_1 \) is both relevant and demanding for \( A \), that \( \delta_2 \) is demanding but not relevant for \( A \), and that \( \delta_3 \) is relevant but not demanding for \( A \). So \( \delta_1 \) is essential for \( A \), but neither \( \delta_2 \) nor \( \delta_3 \) are essential for \( A \). \( \square \)

Example 5.9 Let \( \Sigma_{12}, A, \delta_1, \delta_2, \) and \( \delta_3 \) be as in Examples 5.3 and 5.6. We see from these examples that \( \delta_1 \) is both relevant and demanding for \( A \), that \( \delta_2 \) is demanding but not relevant for \( A \), and that \( \delta_3 \) is relevant but not demanding for \( A \). So \( \delta_1 \) is essential for \( A \), but neither \( \delta_2 \) nor \( \delta_3 \) are essential for \( A \). \( \square \)

The s-t tgd mapping \( M_{12} = (S_1, S_2, \Sigma_{12}) \) is said to have the constant-propagation property [Fag07] if for every ground instance \( I \), every member of the active domain of \( I \) is a member of the active domain of \( \text{chase}_{12}(I) \) (that is, \( \text{dom}(I) \subseteq \text{dom}(\text{chase}_{12}(I)) \)). Later, we shall make use of the following proposition from [Fag07].

Proposition 5.10 [Fag07] Every invertible s-t tgd mapping has the constant-propagation property.

The next proposition gives a similar propagation property.

Proposition 5.11 Assume that \( A \) is a source atom, and \( \delta \) is an essential conjunction for \( A \) with respect to the set \( \Sigma_{12} \) of s-t tgds. Then every variable in \( A \) appears in \( \delta \).

Proof Assume that \( A = P(v_1, \ldots, v_k) \), where \( v_1, \ldots, v_k \) are variables, not necessarily distinct. Assume that the variable \( v_i \) does not appear in \( \delta \); we shall derive a contradiction.

Let \( d \) be a new constant, and let \( I \) be obtained from \( I_A \) by replacing every occurrence of \( c_{v_i} \) in \( I_A \) by \( d \). Since \( \delta \) is relevant for \( A \), we know that there is a homomorphism \( h : I_\delta \rightarrow \text{chase}_{12}(I_A) \). So for the same homomorphism \( h \), we have \( h : I_\delta \rightarrow \text{chase}_{12}(I) \), since \( c_{v_i} \) does not appear in \( I_\delta \). Since \( I_\delta \rightarrow \text{chase}_{12}(I) \), even though \( I_A \not\subseteq I \), it follows that \( \delta \) is not demanding for \( A \), which contradicts the assumption that \( \delta \) is essential for \( A \). \( \square \)

Recall that a weak renaming is a function that maps variables to variables (the word “weak” refers to the fact that the function is not necessarily one-to-one). If \( \varphi \) is a formula, and \( f \) is a weak renaming, let \( \varphi^f \) be the result of replacing every variable \( x \) in \( \varphi \) by \( f(x) \). We may refer to \( \varphi^f \) as a weak renaming of \( \varphi \). If \( \varphi \) is a normal constraint with premise \( \delta \), then we say that \( f \) is consistent with the inequalities of \( \delta \) if \( f(x) \) and \( f(y) \) are distinct for each inequality \( x \neq y \) in \( \delta \). The intuition is that if \( f \) is consistent with the inequalities of the normal constraint \( \varphi \), then
\( \varphi^f \) is a special case of \( \varphi \). For example, assume that \( \varphi \) is the normal constraint (1) of Section 4. Let \( f \) be the weak renaming that is the identity except that \( f(x_1) = x_2 \). Then \( \varphi^f \) is

\[
P(x_2, x_2, x_3, x_3) \land \text{const}(x_2) \land \text{const}(x_3) \land (x_2 \neq x_3) \rightarrow Q(x_2, x_3, x_2)
\]  

Note that (2) can be viewed as a special case of (1) where \( x_1 \) and \( x_2 \) are equal. In fact, the formulas \( \varphi^f \) where \( f \) is consistent with the inequalities \( \varphi \) can be thought of as consisting of all of the special cases of \( \varphi \).

The next theorem relates the notion of “not too strong” with the notion of “demanding”, and relates the notion of “not too weak” with the notion of “relevant”. Shortly, we shall discuss the intuition behind this rather technical theorem.

**Theorem 5.12** Let \( \mathcal{M}_{12} = (\mathcal{S}_1, \mathcal{S}_2, \Sigma_{12}) \) be an s-ttg mapping and \( \mathcal{M}_{21} = (\mathcal{S}_2, \hat{\mathcal{S}}_1, \Sigma_{21}) \) be a normal mapping. Then

1. \( \Sigma_{21} \) is not too strong if and only if every constraint in \( \Sigma_{21} \) is of the form \( \delta \rightarrow \hat{A} \), where \( \delta^f \) is demanding for \( A^f \) for every weak renaming \( f \) consistent with the inequalities of \( \delta \).

2. \( \Sigma_{21} \) is not too weak if and only if for each source atom \( A \), there is a relevant conjunction \( \delta \) for \( A \) such that \( \delta \rightarrow \hat{A} \) is a weak renaming of a constraint in \( \Sigma_{21} \).

**Proof** (1) Assume first that there is a constraint \( \delta \rightarrow \hat{A} \) of \( \Sigma_{21} \) and a weak renaming \( f \) consistent with the inequalities of \( \delta \) such that \( \delta^f \) is not demanding for \( A^f \). Let \( \delta' \) be \( \delta^f \), and let \( A' \) be \( A^f \). Then \( \delta' \rightarrow \hat{A}' \) is a normal constraint that is a logical consequence of \( \Sigma_{21} \). Since \( \delta' \) is not demanding for \( A' \), there is an instance \( I \) such that \( I_{\delta'} \rightarrow \text{chase}_{12}(I) \), yet \( I_{A'} \not\subseteq I \). Since \( I_{\delta'} \rightarrow \text{chase}_{12}(I) \), it follows that \( \text{chase}_{12}(I) \) produces a fact not in \( I \). By renaming constants in \( I \) if needed, this tells us that there is a weak renaming \( f \) such that \( I_{(\delta^f)} \rightarrow \text{chase}_{12}(I) \) and \( I_{(A^f)} \not\subseteq I \). Hence, \( \delta^f \) is not demandig for \( A^f \).

Conversely, assume that \( \Sigma_{21} \) is too strong. Then, by Lemma 4.8, there is a ground instance \( I \) such that \( \text{chase}_{21}(\text{chase}_{12}(I)) \not\subseteq \hat{I} \). It follows that there must be a constraint of the form \( \delta \rightarrow \hat{A} \) in \( \Sigma_{21} \) such that the result of chasing \( \text{chase}_{12}(I) \) with \( \delta \rightarrow \hat{A} \) produces a fact not in \( I \). By renaming constants in \( I \) if needed, this tells us that there is a weak renaming \( f \) such that \( I_{(\delta^f)} \rightarrow \text{chase}_{12}(I) \) and \( I_{(A^f)} \not\subseteq I \). Hence, \( \delta^f \) is not demanding for \( A^f \).

(2) Assume first that \( \Sigma_{21} \) is not too weak. Pick a source atom \( A \). By Lemma 4.9, we know that \( \text{chase}_{12}(I_A) \subseteq \text{chase}_{21}(\text{chase}_{12}(I_A)) \). So there must be a constraint in \( \Sigma_{21} \) that fires on \( \text{chase}_{12}(I_A) \) to introduce \( I_A \). Hence, there must be a weak renaming \( \delta \rightarrow \hat{A} \) of a constraint in \( \Sigma_{21} \) such that \( I_A \rightarrow \text{chase}_{12}(I_A) \). So \( \delta \) is relevant for \( A \).

Conversely, assume that for each source atom \( A \), there is a relevant conjunction \( \delta \) for \( A \) such that \( \delta \rightarrow \hat{A} \) is a weak renaming of a constraint \( \varphi \in \Sigma_{21} \). Then \( I_{\delta} \rightarrow \text{chase}_{12}(I_A) \) because \( \delta \) is relevant for \( A \), and therefore we have \( I_A \subseteq \text{chase}_{21}(\text{chase}_{12}(I_A)) \) because \( \varphi \) fires on \( \text{chase}_{12}(I_A) \) to introduce \( I_A \). Since \( I_A \subseteq \text{chase}_{21}(\text{chase}_{12}(I_A)) \) for each source atom \( A \), it follows easily that \( \hat{I} \subseteq \text{chase}_{21}(\text{chase}_{12}(I)) \) for every ground instance \( I \). By Lemma 4.9, \( \Sigma_{21} \) is not too weak. \( \square \)

Let us discuss the intuition behind Theorem 5.12. We consider first part (1) of Theorem 5.12. For \( \Sigma_{21} \) to be not too strong, the members of \( \Sigma_{21} \) must be restricted. Let \( \varphi \) be a member of \( \Sigma_{21} \). As we commented earlier, the formulas \( \varphi^f \), where \( f \) is a weak renaming consistent with the inequalities of \( \delta \), can be viewed as all of the special cases of \( \varphi \). So Part (1) of Theorem 5.12 says that \( \varphi \) is not too strong if and only if for every special case \( \varphi^f \) of every member \( \varphi \) of \( \Sigma_{21} \), if \( \varphi^f \) is \( \delta^f \rightarrow \hat{A}^f \), then \( \delta^f \) is demanding for \( A^f \). By the intuition we gave earlier for the notion of “demanding”, \( \delta^f \) being demanding for \( A^f \) means that, roughly speaking, “all of the atoms in \( \delta^f \) can arise together in a chase only if \( A^f \) is present in the source”.

Consider now part (2) of Theorem 5.12. For \( \Sigma_{21} \) to be not too weak, \( \Sigma_{21} \) must contain (or imply) certain constraints. If we again think of \( \varphi^f \) as a special case of a normal constraint \( \varphi \), then Part (2) of Theorem 5.12 says that \( \Sigma_{21} \) is not too weak if and only if for each source atom \( A \), there is a relevant conjunction \( \delta \) for \( A \) such that...
\( \delta \rightarrow \hat{A} \) is a special case of a constraint in \( \Sigma_{21} \). By the intuition we gave earlier for the notion of “relevant”, \( \delta \) being relevant for \( A \) means that, roughly speaking, “the atoms in \( \delta \) arise in the chase of \( A \).”

The next theorem characterizes normal inverses of s-t tgd mappings in terms of the notions of demanding, relevant, and essential.

**Theorem 5.13** Let \( \mathcal{M}_{12} = (S_1, S_2, \Sigma_{12}) \) be an s-t tgd mapping and \( \mathcal{M}_{21} = (S_2, \hat{S}_1, \Sigma_{21}) \) be a normal mapping. Then \( \mathcal{M}_{21} \) is an inverse of \( \mathcal{M}_{12} \) if and only if

1. Every constraint in \( \Sigma_{21} \) is of the form \( \delta \rightarrow \hat{A} \), where \( \delta^f \) is demanding for \( A^f \) for every weak renaming \( f \) consistent with the inequalities of \( \delta \).

2. For each source atom \( A \), there is a relevant conjunction \( \delta \) for \( A \) such that \( \delta \rightarrow \hat{A} \) is a weak renaming of a constraint in \( \Sigma_{21} \). (By Part (1), this relevant conjunction is essential.)

**Proof** This follows immediately from Proposition 4.4 and Theorem 5.12. \( \square \)

**Definition 5.14** Let \( \mathcal{M}_{12} = (S_1, S_2, \Sigma_{12}) \) be an s-t tgd mapping. Let \( e \) be a partial function whose domain consists of all prime source atoms that have an essential conjunction. If the prime source atom \( A \) has an essential conjunction, then \( e(A) \) is an essential conjunction for \( A \) that is a conjunction of relational atoms and the formulas \( \text{const}(x) \) for each variable \( x \) of \( A \). (Note that each essential conjunction for \( A \) must contain all of the variables of \( A \), by Proposition 5.11.) If \( A \) has no essential conjunction, then \( e(A) \) is undefined. Let \( \Sigma_{21}^e \) consist of all formulas \( e(A) \land \eta_A \rightarrow \hat{A} \), where \( A \) is a prime source atom and where \( e(A) \) is defined, and \( \eta_A \) consists of all inequalities of the form \( x \neq y \) where \( x \) and \( y \) are distinct variables of \( A \). \( \square \)

The next theorem shows how we can construct an inverse out of essential conjunctions.

**Theorem 5.15** Let \( \mathcal{M}_{12} \) be an s-t tgd mapping. The following are equivalent.

1. \( \mathcal{M}_{12} \) is invertible.

2. For every source atom \( A \), there is an essential conjunction for \( A \).

3. \( \mathcal{M}_{21}^e \) is an inverse of \( \mathcal{M}_{12} \), for every partial function \( e \) as in Definition 5.14.

4. \( \mathcal{M}_{21}^e \) is an inverse of \( \mathcal{M}_{12} \), for some partial function \( e \) as in Definition 5.14.

We defer the proof of Theorem 5.15 to Section 6. Note that the inverses \( \mathcal{M}_{21}^e \) in (3) and (4) of Theorem 5.15 are normal. The equivalence of (1) and (2) in Theorem 5.15 gives a clean necessary and sufficient condition for the existence of an inverse. Further, the equivalence of (1) and (2) shows the fundamental importance of essential conjunctions for the study of inverses (normal or otherwise).

The next two lemmas will be useful later.

**Lemma 5.16** Let \( A \) and \( A' \) be source atoms. Assume that \( \delta \) is relevant for \( A \) and demanding for \( A' \). Then \( A \) and \( A' \) are the same atom.

**Proof** Since \( \delta \) is relevant for \( A \), we know that \( I_{\delta} \rightarrow \text{chase}_{12}(I_A) \). Therefore, since \( \delta \) is demanding for \( A' \), we have \( I_{A'} \subseteq I_A \). Since \( I_A \) and \( I_{A'} \) are both singletons, we have \( I_{A'} = I_A \), so \( A \) and \( A' \) are the same atom, as desired. \( \square \)

**Lemma 5.17** Let \( \mathcal{M}_{12} \) be a full s-t tgd mapping, and \( \mathcal{M}_{21} = (S_2, \hat{S}_1, \Sigma_{21}) \) a normal inverse for \( \mathcal{M}_{12} \). Let \( A \) be a source atom and \( B \) a target atom where \( \hat{I}_A \subseteq \text{chase}_{21}(I_B) \). Then \( B \) is demanding for \( A \) with respect to \( \Sigma_{12} \).

**Proof** Assume that \( I_B \subseteq \text{chase}_{12}(I) \); we must show that \( I_A \subseteq I \). Let \( U = \text{chase}_{12}(I) \). We know from Proposition 4.6 that \( (U, \bar{I}) \models \Sigma_{21} \). Since \( \Sigma_{21} \) is full, this implies further that \( \text{chase}_{21}(U) \subseteq \bar{I} \). Since \( I_B \subseteq \text{chase}_{12}(I) \) and \( \hat{I}_A \subseteq \text{chase}_{21}(I_B) \), it follows that \( \hat{I}_A \subseteq \text{chase}_{21}(\text{chase}_{12}(I)) = \text{chase}_{21}(U) \). Since also \( \text{chase}_{21}(U) \subseteq \bar{I} \), we have that \( \hat{I}_A \subseteq \bar{I} \), and so \( I_A \subseteq I \), as desired. \( \square \)

To prevent confusion, the reader should note that the inclusion \( \hat{I}_A \subseteq \text{chase}_{21}(I_B) \) in Lemma 5.17 uses \( \text{chase}_{21} \), not \( \text{chase}_{12} \). If we were to instead consider the inclusion \( I_A \subseteq \text{chase}_{12}(I_B) \) (where we now take \( B \) to be a source atom and \( A \) a target atom), then this inclusion would say that \( A \) is relevant for \( B \).
5.1 The Full Case

If $\Sigma_{12}$ is full, then we are interested in the situation where $\delta$ has no const formulas. Then $I_\delta$ contains only constants (no nulls). If $\delta$ is simply a relational atom, and if $\delta$ is demanding, then we call $\delta$ a demanding atom. Similarly, we define a relevant atom and an essential atom. The reason we are interested in the demanding atoms (and the essential atoms) in the full case is because of the following proposition.

**Proposition 5.18** Let $M_{12} = (S_1, S_2, \Sigma_{12})$ be a full s-t tgd mapping. Assume that $A$ is a source atom. Assume that the conjunction $\delta$ has no const formulas. If $\delta$ is demanding for $A$, then $\delta$ contains a demanding atom for $A$. If $\delta$ is essential for $A$, then $\delta$ contains an essential atom for $A$.

**Proof** Assume by way of contradiction that $\delta$ is demanding for $A$, but that no atom $B$ of $\delta$ is demanding for $A$. So for every atom $B$ of $\delta$, there is a ground instance $J_B$ such that $I_B \rightarrow \text{chase}_{12}(J_B)$ and $I_A \not\subseteq J_B$. Since every member of $I_B$ is a constant, and since $I_B \rightarrow \text{chase}_{12}(J_B)$, it follows that $I_B \subseteq \text{chase}_{12}(J_B)$. Let $I$ be the union, over all atoms $B$ in $\delta$, of $I_B$. Therefore, since $I_B \subseteq \text{chase}_{12}(I)$ for every atom $B$ of $\delta$, it follows that $I_\delta \subseteq \text{chase}_{12}(\delta)$. Let $I$ be the union, over all atoms $B$ in $\delta$, of $I_B$. Therefore, $I_\delta \subseteq \text{chase}_{12}(\delta)$.

Now assume that $\delta$ is essential for $A$. We have shown that $\delta$ contains a demanding atom $B$ for $A$. Since $\delta$ is relevant for $A$, we have $I_\delta \rightarrow \text{chase}_{12}(I_A)$. But $I_\delta$ contains only constants (no nulls), and so $I_\delta \subseteq \text{chase}_{12}(I_A)$. As we noted, $I_\delta$ is the union, over all atoms $B$ in $\delta$, of $I_B$. Therefore, $I_B \subseteq I_\delta \subseteq \text{chase}_{12}(I_A)$, and so $I_B \rightarrow \text{chase}_{12}(I_A)$. Therefore, $B$ is relevant for $A$. Since $B$ is both relevant and demanding for $A$, we see that $B$ is essential for $A$, as desired. □

The next proposition says that, in the full case, we can strengthen the equivalence of (1) and (2) in Theorem 5.15.

**Theorem 5.19** Let $M_{12}$ be a full s-t tgd mapping. The following are equivalent.

1. $M_{12}$ is invertible.
2. For every source atom $A$, there is an essential atom for $A$.

**Proof** Proposition 6.2 will tell us that in the full case, the essential conjunction in (2) of Theorem 5.15 can be taken to have no const formulas. The result then follows from Proposition 5.18. □

As with Theorem 5.15, the equivalence of (1) and (2) in Theorem 5.19 gives a clean necessary and sufficient condition for the existence of an inverse, this time in the full case. Further, the equivalence of (1) and (2) shows the fundamental importance of essential atoms for the study of inverses (normal or otherwise) in the full case.

In the full case, we can strengthen Proposition 5.11 as follows.

**Proposition 5.20** Assume that $A$ is a source atom, and $B$ is an essential atom for $A$ with respect to the set $\Sigma_{12}$ of full s-t tgds. Then the variables in $B$ are exactly the variables in $A$.

**Proof** By Proposition 5.11, we know that every variable in $A$ appears in $B$. Since $B$ is relevant for $A$, and $\Sigma_{12}$ is full, we have $I_B \subseteq \text{chase}_{12}(I_A)$. Therefore, every variable in $B$ appears in $A$. □
6 The Canonical Candidate Inverse

In [FKPT08], the “canonical candidate inverse” was defined for each s-t tgd mapping, and it was shown (with quite a complicated proof) that if an s-t tgd mapping $M$ has an inverse, then the canonical candidate inverse of $M$ is also an inverse of $M$. Since the canonical candidate inverse is a normal inverse, in this section we take advantage of the machinery we have developed to give a much simpler proof of this result (and of the fact that the canonical candidate inverse is the weakest possible normal inverse).

**Definition 6.1** Let $M_{12} = (S_1, S_2, \Sigma_{12})$ be an s-t tgd mapping. For each source atom $A$, let $I_A$ be, as before, the instance containing the fact obtained by replacing each variable $v$ in $A$ by a distinct constant $c_v$. Let $V_A$ be the result of chasing $I_A$ with $\Sigma_{12}$. Let $\nu_A$ be the conjunction of relational atoms obtained by replacing every constant $c_v$ of $V_A$ by the variable $v$, and replacing every null $n$ of $V_A$ by a new variable $v_n$ (that does not appear in $A$). Let $\chi_A$ be the conjunction of the formulas $\text{const}(x)$ for each variable $x$ in $A$, and let $\omega_A$ be the conjunction of $\nu_A$ and $\chi_A$. Let $\eta_A$ be the conjunction of all inequalities of the form $x \neq y$ where $x$ and $y$ are distinct variables in $A$.

**Proposition 6.2** Let $M_{12} = (S_1, S_2, \Sigma_{12})$ be an invertible s-t tgd mapping. Let $A$ be a source atom. Then $\omega_A$, as defined in Definition 6.1, is an essential conjunction for $A$. If $M_{12}$ is full, then $\nu_A$, as defined in Definition 6.1, is an essential conjunction for $A$.

**Proof** It is clear that $\omega_A$ is relevant for $A$ (as is $\nu_A$, in the full case).

We now show that $\omega_A$ is demanding for $A$. Assume that $I_{\omega_A} \rightarrow \text{chase}_{12}(I)$ for some ground instance $I$; we must show that $I_A \subseteq I$. Now $I_{\omega_A} = \text{chase}_{12}(I_A)$. So $\text{chase}_{12}(I_A) \rightarrow \text{chase}_{12}(I)$. By the implication $(1) \Rightarrow (4)$ of Proposition 3.2, it follows that $I_A \subseteq I$, as desired. A similar argument shows that $\nu_A$ is demanding for $A$ in the full case.

The canonical candidate inverse [FKPT08] of an invertible s-t tgd mapping $M_{12} = (S_1, S_2, \Sigma_{12})$ is the normal mapping $M_{21}^{c_1} = (S_2, S_1, \Sigma_{21})$ where $\Sigma_{21}$ contains, for every prime source atom $A$, the constraint $\nu_A \land \chi_A \land \eta_A \rightarrow A$. By Propositions 5.11 and 6.2, every variable in $A$ appears in $\nu_A$, and so these constraints are well-defined, and specify normal mappings.

We can now prove Theorem 5.15.

**Proof of Theorem 5.15:** The implications $(3) \Rightarrow (4)$, and $(4) \Rightarrow (1)$, are immediate. The implication $(1) \Rightarrow (2)$ follows by Proposition 6.2. So we need only show the implication $(2) \Rightarrow (3)$. Assume that $(2)$ holds. Therefore, $e(A)$ is an essential conjunction for $A$, for each atomic formula $A$. We now make use of Theorem 5.13, where the role of $\Sigma_{21}$ is played by $\Sigma_{21}^\delta$. Let $e(A) \land \eta_A \rightarrow A$ be an arbitrary member of $\Sigma_{21}^\delta$. Let $\delta = e(A) \land \eta_A$. Then $\delta$ is essential for $A$, and hence demanding for $A$ and relevant for $A$. Let $f$ be a weak renaming consistent with $\eta_A$. By construction of $\eta_A$, it follows that $f$ is a one-to-one on the variables of $A$. Therefore, since $\delta$ is demanding for $A$, also $\delta^f$ is demanding for $A^f$ (the possibility that $f$ is not necessarily one-to-one when we consider also the variables not in $A$ cannot hurt). So part $(1)$ of Theorem 5.13 holds when the role of $\Sigma_{21}$ is played by $\Sigma_{21}^\delta$. Also, part $(2)$ of Theorem 5.13 holds, where the weak renaming is a renaming obtained by renaming the variables in a prime source atom. Therefore, $M_{21}^{\delta^f}$ is an inverse of $M_{12}$, and so $(3)$ of Theorem 5.15 holds, as desired.

It is shown in [FKPT08] that if $M_{12}$ is an invertible s-t tgd mapping, then the canonical candidate inverse of $M_{12}$ is indeed an inverse of $M_{12}$. The proof in [FKPT08] is quite complicated. We will now give a proof, based on the following proposition, that is much simpler (given our machinery).

**Theorem 6.3** [FKPT08] Assume that $M_{12} = (S_1, S_2, \Sigma_{12})$ is an invertible s-t tgd mapping. Then the canonical candidate inverse of $M_{12}$ is indeed an inverse of $M_{12}$.

**Proof** Let $e$ be the function that assigns to each prime source atom $A$ the formula $\omega_A$. By Proposition 6.2, we know that $e(A)$ is an essential conjunction for $A$. So by the implication $(1) \Rightarrow (3)$ of Theorem 5.15, we know that $M_{21}^{e(A)}$ is an inverse of $M$. But $M_{21}^{e(A)}$ is the canonical candidate inverse of $M_{12}$, and so the canonical candidate inverse of $M_{12}$ is an inverse of $M_{12}$, as desired.
Our final proposition (which we shall make use of later) in this section says that the canonical candidate inverse of an invertible s-t tgd mapping \( M_{12} \) is the weakest normal inverse of \( M_{12} \). This proposition follows from results in [FKPT08] (although the notion of a normal inverse did not appear in [FKPT08]). Since the proofs of those results are rather complicated, we shall give a simple, self-contained proof of the next proposition.

**Proposition 6.4** Let \( M_{12} \) be an s-t tgd mapping, let \( M_{21} = (S_2, S_1, \Sigma_{21}) \) be the canonical candidate inverse of \( M_{12} \), and let \( M_{21} = (S_2, S_1, \Sigma_{21}) \) be an arbitrary normal inverse of \( M_{12} \). Then \( \Sigma_{21} \) logically implies \( \Sigma_{21}' \).

**Proof** Let \( \sigma \) be an arbitrary member of \( \Sigma_{21} \); we must show that \( \Sigma_{21} \) logically implies \( \sigma \). Assume not; we shall derive a contradiction.

Since \( \Sigma_{21} \) does not logically imply \( \sigma \), there is \( (K, J) \) such that \( (K, J) \models \Sigma_{21} \) but \( (K, J) \not\models \sigma \). Let \( J' \) consist of the ground facts of \( J \); that is, \( J' \) consists of all facts \( P(a_1, \ldots, a_k) \) of \( J \) where \( a_1, \ldots, a_k \) are constants. Since \( (K, J') \models \Sigma_{21} \), it follows by normality of \( M_{12} \) that \( (K, J') \models \Sigma_{21} \) (intuitively, the non-ground facts of \( J \) have no effect in determining satisfaction of \( \Sigma_{21} \)). Similarly, since \( (K, J) \not\models \sigma \), it follows that \( (K, J') \not\models \sigma \).

Using the notation of Definition 6.1, let us write \( \sigma \) as \( \nu_A \land \chi_A \land \eta_A \rightarrow \tilde{A} \). Also, let \( V_A \) be as in Definition 6.1. Since \( (K, J') \not\models \sigma \), we have (by renaming constants if necessary) that there is a homomorphism \( h : V_A \rightarrow K \) (that preserves constants and respects the inequalities \( \eta_A \)), where \( \tilde{I}_A \not\subseteq J' \). Since \( V_A \) is the result of chasing \( I_A \) with \( \Sigma_{12} \), we have \( (I_A, V_A) \models \Sigma_{12} \). Therefore, since \( V_A \rightarrow K \), it follows from Lemma 3.1 that \( (I_A, K) \models \Sigma_{12} \). Since also \( (K, J') \models \Sigma_{21} \), it follows that \( (I_A, J') \models \Sigma_{12} \models \Sigma_{21} \). Since \( M_{21} \) is an inverse of \( M_{12} \), and since \( I_A \) and \( J' \) are ground instances, it follows that \( \tilde{I}_A \not\subseteq J' \). This is our desired contradiction. \( \square \)

### 7 Unique Inverses

Mathematicians are accustomed to inverses being unique. For example, an invertible function has a unique inverse, and an invertible finite matrix has a unique inverse. The reason that inverses are typically unique is that they are typically two-sided inverses. Thus, assume that \( Y \) and \( Y' \) are both two-sided inverses of \( X \) for some associative operation \( \ast \), and that \( I \) is the identity under this operation. Then \( X \ast Y = I \). Apply \( Y' \) to both sides, and we get

\[ Y' \ast (X \ast Y) = Y' \ast I. \tag{3} \]

By associativity, the left-hand side of (3) is \( (Y' \ast X) \ast Y = I \ast Y = Y \), and the right-hand side of (3) is \( Y' \). Therefore, \( Y = Y' \), and so the inverse of \( X \) is unique.

Sometimes inverses are not unique. This can happen in particular for infinite matrices, where multiplication is not necessarily associative. For a discussion of such phenomena on infinite matrices, we refer the interested reader to the first author’s first paper [Fag68]. In particular, Lemma 2 of that paper gives a sufficient condition for an infinite matrix to have a unique two-sided inverse. It also gives an example where that unique two-sided inverse is a unique right inverse, but where there are multiple left inverses.

In this section, we consider the notion of uniqueness of inverses of s-t tgd mappings (including questions such as whether or not a given s-t tgd mapping has a unique normal inverse). Our interest in this question was inspired by the following two examples of s-t tgd mappings, one with a unique normal inverse, and another with multiple normal inverses.

**Example 7.1** Let \( M_{12} = (S_1, S_2, \Sigma_{12}) \), where \( S_1 \) consists of the unary relation symbol \( R \), where \( S_2 \) consists of the unary relation symbol \( S \), and where \( \Sigma_{12} \) consists of the tgd \( R(x) \rightarrow S(x) \). Let \( \Sigma_{21} \) consist of the normal constraint \( S(x) \land \text{const}(x) \rightarrow \tilde{R}(x) \), and let \( M_{21} = (S_2, S_1, \Sigma_{21}) \). It is easy to see that \( M_{21} \) is a normal inverse of \( M_{12} \). A theorem we shall give later (Theorem 7.7) implies that \( M_{21} \) is in fact the unique normal inverse of \( M_{12} \). \( \square \)

19
Example 7.2  Let $S_1$ consist of the unary relation symbol $R$, and let $S_2$ consist of the binary relation symbol $S$. Let $\Sigma_{12}$ consist of the tgd $R(x) \rightarrow S(x, x)$. Let $\Sigma_{21}$ consist of the normal constraint $S(x, x) \land \text{const}(x) \rightarrow R(x)$, and let $\Sigma_{21}'$ consist of the normal constraint $S(x, y) \land \text{const}(x) \rightarrow R(x)$. Let $M_{12} = (S_1, S_2, \Sigma_{12})$, let $M_{21} = (S_2, \hat{S}_1, \Sigma_{21})$, and let $M_{21}' = (S_2, \hat{S}_1, \Sigma_{21}')$. It is straightforward to verify that $M_{21}$ and $M_{21}'$ are inequivalent normal inverses of $M_{12}$. □

Because of these two examples (but where the focus was on unique inverses specified by tgds), Fagin [Fag07] says, “It might be interesting to examine the question of when there is a unique inverse mapping specified in a given language.” We note that a referee of [Fag07] commented that a reason why $M_{12}$ in Example 7.2 does not have a unique inverse is that $M_{12}$ is “not onto”. In this section, we give a definition of what it means for a full s-t tgd mapping to be onto, and show that, this condition is a sufficient condition for an invertible full s-t tgd mapping to have a unique normal inverse (however, as we shall see, it is not a necessary condition).

In this section, we also show that no schema mapping has a unique inverse. Therefore, to attain uniqueness, we must restrict to a class of possible inverses, such as normal inverses. We give a necessary and sufficient condition for an s-t tgd mapping to have a unique normal inverse. We use this theorem to show that even a nonfull s-t tgd mapping can have a unique normal inverse. Since, as we noted, being onto is a sufficient but not necessary condition for an invertible full s-t tgd mapping to have a unique normal inverse, we find a larger class $C$ of schema mappings (those specified by disjunctive tgds with inequalities) such that being onto is a necessary and sufficient condition for an invertible full s-t tgd mapping $M$ to have a unique inverse in the class $C$. We also show the surprising result that such mappings $M$ are very special: they are very close to being copy mappings.

Say that two schema mappings $(S_1, S_2, \Sigma_{12})$ and $(S_1, S_2, \Sigma_{12}')$ are equivalent if $\Sigma_{12}$ and $\Sigma_{12}'$ are logically equivalent. It is useful to consider a weaker notion than equivalence. Recall that if $M_{12} = (S_1, S_2, \Sigma_{12})$ and $M_{21} = (S_2, \hat{S}_1, \Sigma_{21})$ are schema mappings, then $M_{21}$ is an inverse of $M_{12}$ if and only if for every pair $I, J$ of ground instances, we have that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ if and only if $I \subseteq J$. Therefore, for pairs $(J_1, J_2)$ where $J_2$ is not a ground instance, the pair $(J_1, J_2)$ satisfying or not satisfying $\Sigma_{21}$ plays no role whatever in determining whether or not $M_{21}$ is an inverse of $M_{12}$. Based on this intuition, let us say that $\Sigma_{21}$ and $\Sigma_{21}'$ are weakly equivalent if whenever $J_1$ is arbitrary and $J_2$ is a ground instance, then $(J_1, J_2) \models \Sigma_{21}$ if and only if $(J_1, J_2) \models \Sigma_{21}'$. □ We may also say that $(S_2, \hat{S}_1, \Sigma_{21})$ and $(S_2, \hat{S}_1, \Sigma_{21}')$ are then weakly equivalent. If $(S_2, \hat{S}_1, \Sigma_{21})$ and $(S_2, \hat{S}_1, \Sigma_{21}')$ are both normal mappings, then they are weakly equivalent if and only if they are equivalent. This follows easily from the observation that if $J'_2$ is the result of removing every non-ground fact from $J_2$, then $(J_1, J_2') \models \Sigma_{21}$ if and only if $(J_1, J'_2) \models \Sigma_{21}$.

We capture the intuition about the irrelevance of pairs $(J_1, J_2)$ where $J_2$ is not a ground instance in the following simple proposition.

Proposition 7.3  Let $M_{12}$ be a schema mapping, and let $M_{21}$ and $M_{21}'$ be weakly equivalent schema mappings. Then $M_{21}$ is an inverse of $M_{12}$ if and only if $M_{21}'$ is an inverse of $M_{12}$.

Proof  By symmetry, we need only show that if $M_{21}$ is an inverse of $M_{12}$, then $M_{21}'$ is an inverse of $M_{12}$. Let $I, J$ be ground instances. Since $M_{21}$ is an inverse of $M_{12}$, we know that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ if and only if $I \subseteq J$. To show that $M_{21}'$ is an inverse of $M_{12}$, we need only show that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ if and only if $(I, J) \models \Sigma_{12} \circ \Sigma_{21}'.

Now $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ if and only if there is $J'$ such that $(I, J') \models \Sigma_{12}$ and $(J', J) \models \Sigma_{21}$. Since $M_{21}$ and $M_{21}'$ are weakly equivalent, and $J$ is ground, we have that $(J', J) \models \Sigma_{21}$ if and only if $(J', J) \models \Sigma_{21}'$. Hence, $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ if and only if there is $J'$ such that $(I, J') \models \Sigma_{12}$ and $(J', J) \models \Sigma_{21}$, which happens if and only if $(I, J) \models \Sigma_{12} \circ \Sigma_{21}'$. Therefore, $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ if and only if $(I, J) \models \Sigma_{12} \circ \Sigma_{21}'. □

We now make use of Proposition 7.3 to show that no schema mapping has a unique inverse. In the following theorem (and later), when we speak of “uniqueness”, we mean uniqueness up to equivalence.

---

Footnote: This notion arises also in the full version of [FKPT08].
Theorem 7.4 No schema mapping has a unique inverse.

Proof Let \( \mathcal{M}_{12} = (S_1, S_2, \Sigma_{12}) \) be an invertible schema mapping. Assume that \( \mathcal{M}_{21} = (S_2, S_1, \Sigma_{21}) \) is an inverse of \( \mathcal{M}_{12} \). Define \( \Sigma'_{21} \) by having \((J_1, J_2) \models \Sigma'_{21}\) if and only if \((J_1, J_2) \models \Sigma_{21}\) and \(J_2\) is ground.\(^5\)

Define \( \Sigma''_{21} \) by having \((J_1, J_2) \models \Sigma''_{21}\) if and only if either \((J_1, J_2) \models \Sigma_{21}\) or \(J_2\) contains a null value. Let \( \mathcal{M}'_{21} = (S_2, S_1, \Sigma'_{21}) \) and let \( \mathcal{M}''_{21} = (S_2, S_1, \Sigma''_{21}) \). By construction, we see that \( \Sigma_{21}, \Sigma'_{21}\), and \( \Sigma''_{21}\) are all weakly equivalent. It follows from Proposition 7.3 that since \( \mathcal{M}_{21}\) is an inverse of \( \mathcal{M}_{12}\), so are \( \mathcal{M}'_{21}\) and \( \mathcal{M}''_{21}\). We also see from the construction that \( \Sigma_{21}\) and \( \Sigma''_{21}\) are not logically equivalent. So \( \mathcal{M}'_{21}\) and \( \mathcal{M}''_{21}\) are inverses of \( \mathcal{M}_{12}\) that are not equivalent. \( \square \)

Because of Theorem 7.4, if we wish to study uniqueness of inverses, we must restrict our attention to particular classes (such as normal inverses). We have seen that normal inverses are an important class (in particular, every invertible \( s\!-\!t \) tdg mapping has a normal inverse). In Example 7.1, we gave an \( s\!-\!t \) tdg mapping with a unique normal inverse (although we have not proven uniqueness yet), and in Example 7.2, we gave an example with multiple normal inverses.

The next theorem gives a necessary and sufficient condition, based on our notions of “essential” and “demanding”, for an invertible \( s\!-\!t \) tdg mapping to have a unique normal inverse.

Theorem 7.5 An invertible \( s\!-\!t \) tdg mapping has a unique normal inverse if and only if for every source atom \( A\), if \( \delta\) is an essential conjunction for \( A\), and \( \delta'\) is a demanding conjunction for \( A\), both with formulas \( \text{const}(x)\) for exactly the variables \( x\) that appear in \( A\), then \( I_\delta \rightarrow I_{\delta'} \).

Proof Assume first that \( \mathcal{M}_{12}\) is an invertible \( s\!-\!t \) tdg mapping with a unique normal inverse. Let \( A\) be a source atom. Assume that \( \delta\) is essential for \( A\), and \( \delta'\) is demanding for \( A\), and both have formulas \( \text{const}(x)\) for exactly the variables \( x\) that appear in \( A\). Assume that we do not have \( I_\delta \rightarrow I_{\delta'}\); we shall derive a contradiction. Assume without loss of generality that \( \delta'\) has no inequalities as conjuncts (if necessary, remove them). Let \( e\) be as in Definition 5.14 with \( e(A) = \delta\). It follows from Theorem 5.15 that \( \mathcal{M}''_{21}\) is an inverse of \( \mathcal{M}_{12}\). Let \( \sigma'\) be \( \delta' \land \eta_A \rightarrow \hat{A}\), where \( \eta_A\) is a conjunction of the inequalities \( x \neq y\) for distinct variables \( x, y\) of \( A\). Let \( \Sigma_{21} = \Sigma''_{21} \cup \{\sigma'\}\). Let \( \mathcal{M}_{21} = (S_2, S_1, \Sigma_{21})\).

We now show that \( \mathcal{M}_{21}\) is also an inverse of \( \mathcal{M}_{12}\). Let \( \delta''\) be \( \delta' \land \eta_A\). Let \( f\) be a weak renaming consistent with the inequalities of \( \delta''\). It follows easily from Theorem 5.13 that we need only show that \( \delta''\) is demanding for \( A'\). Let \( I\) be a ground instance such that \( I_{\delta''} \rightarrow \text{chase}_{12}(I)\); we must show that \( I_{A'} \subseteq I\). Since \( \text{const}(x)\) appears in \( \delta''\) only for variables \( x\) in \( A\), and since \( f\) is consistent with the inequalities in \( \eta_A\), we can assume without loss of generality (by renaming variables if needed) that \( f(x) = x\) for each variable \( x\) of \( A\). So \( A'\) is simply \( A\). Since \( \text{const}(x)\) appears in \( \delta''\) only for variables \( x\) where \( f(x) = x\), it follows easily that \( f\) is a homomorphism from \( I_{\delta''}\) to \( I_{\delta''}f\). Furthermore, \( I_{\delta''} = I_{\delta''f}\), since inequalities are ignored in computing \( I_{\delta''}\). Therefore, \( I_{\delta''} \rightarrow I_{\delta''}f\). Hence, since \( I_{\delta''}f \rightarrow \text{chase}_{12}(I)\), we have \( I_{\delta''} \rightarrow \text{chase}_{12}(I)\). Since \( \delta'\) is demanding for \( A\), this implies that \( I_A \subseteq I\). But \( A = A'\). Hence, \( I_{A'} \subseteq I\), as desired.

We now show that \( \mathcal{M}'_{21}\) and \( \mathcal{M}_{21}\) are not equivalent, which gives our desired contradiction. Let \( J = I_{\delta'}\). Let \( I\) be the result of chasing \( J\) with \( \Sigma_{21}\). Clearly \( (J, I) \models \Sigma_{21}\). We now show that \( (J, I) \not\models \sigma'\), and so \( (J, I) \not\models \Sigma_{21}\). Note that because of the structure of \( \Sigma_{21}\), it follows that \( I\) is a ground instance.

Let \( \hat{A}(\bar{c})\) be the result of chasing \( J\) with \( \sigma'\). We need only show that \( \hat{A}(\bar{c})\) does not appear in \( I\). Let \( \sigma\) be an arbitrary member of \( \Sigma_{21}\). By construction of \( \Sigma_{21}\), we know that \( \sigma\) is of the form \( e(A) \land \eta_{A'} \rightarrow \bar{A'}\), where \( A'\) is a prime source atom, and where \( \eta_{A'}\) is a conjunction of the inequalities \( x \neq y\) for distinct variables \( x, y\) of \( A'\). We must show that the result of chasing \( J\) with \( \sigma\) does not produce \( \hat{A}(\bar{c})\). There are two cases.

Case 1: \( A'\) involves a different relation symbol than \( A\) does. So certainly the result of chasing \( J\) with \( \sigma\) does not produce \( \hat{A}(\bar{c})\).

\(^{5}\)Note that we define \( \Sigma'_{21}\) not by writing formulas, but simply by saying which pairs \((J_1, J_2)\) satisfy \( \Sigma'_{21}\).
Case 2: $A'$ involves the same relation symbol as $A$. There are two subcases.

Subcase 2a: $A'$ equals $A$. Since there is no homomorphism from $I_{\delta}$ to $I_{\delta'}$, that is, from $I_{\delta}$ to $J$, and since $e(A) = \delta$, it follows that $\sigma$ does not fire on $J$.

Subcase 2b: $A'$ is different from $A$. Then the equality pattern of the variables in $A'$ is different from the equality pattern of the variables in $A$. Hence, the result of chasing $J$ with $\sigma$ again does not produce $\widehat{A}(\overline{c})$.

We now prove the converse. Assume that for every source atom $A$, if $\delta$ is an essential conjunction for $A$, and $\delta'$ is a demanding conjunction for $A$, both with formulas $\text{const}(x)$ for exactly the variables $x$ that appear in $A$, then $I_{\delta} \rightarrow I_{\delta'}$. Let $e$ be as in Definition 5.14. So $\mathcal{M}_{21} = (\widehat{S}_2, \widehat{S}_1, \Sigma_{21})$ is an inverse of $\mathcal{M}_{12}$, by Theorem 5.15.

Let $\mathcal{M}_{21} = (S_2, \widehat{S}_1, \Sigma_{21})$ be an arbitrary normal inverse of $\mathcal{M}_{12}$. We need only show that $\Sigma_{21}^e$ and $\Sigma_{21}$ are logically equivalent.

We first show that $\Sigma_{21}$ logically implies $\Sigma_{21}^e$. Let $\sigma$ be an arbitrary member of $\Sigma_{21}^e$. Then $\sigma$ is of the form $e(A) \land \eta_A \rightarrow \widehat{A}$. Let $\delta$ be $e(A)$. By part (2) of Theorem 5.13, we know that there is an essential conjunction $\delta'$ for $A$ such that $\delta' \rightarrow \widehat{A}$ is a weak renaming of a constraint in $\Sigma_{21}$. Since $\delta$ and $\delta'$ are both essential for $A$, it follows by assumption that $I_{\delta}$ and $I_{\delta'}$ are homomorphically equivalent. It is not hard to see that this implies that $\Sigma_{21}$ logically implies $\sigma$. Since $\sigma$ is an arbitrary member of $\Sigma_{21}^e$, it follows that $\Sigma_{21}$ logically implies $\Sigma_{21}^e$, as desired.

We now show that $\Sigma_{21}^e$ logically implies $\Sigma_{21}$. Let $\sigma$ be an arbitrary member of $\Sigma_{21}$. By part (1) of Theorem 5.13, we know that $\sigma$ is of the form $\delta'' \land \eta_{A'} \rightarrow \widehat{A}'$, where $A$ is a source atom and where $(\delta'')'$ is demanding for $A'$, where $\delta'$ is a source atom and where $(\delta')'$ is demanding for $A'$, and $\eta_{A'}$ is the conjunction of all inequalities $x \neq y$ for distinct variables $x, y$ of $A'$. By further renaming variables if needed, we can assume that $A'$ is a prime atom. Now there is an essential conjunction $\delta$ for $A'$ such that $\delta \land \eta_{A'} \rightarrow \widehat{A}'$ is a normal constraint in $\Sigma_{21}^e$. Let us denote this constraint by $\gamma$. Since $\gamma$ is essential for $A$, and $\delta''$ is demanding for $A$, it follows by assumption that $I_{\delta} \rightarrow I_{\delta''}$. It follows easily that $\gamma$ logically implies $\tau_f$. So $\Sigma_{21}^e$ logically implies $\tau_f$, as desired. \qed

From Theorem 7.5, we obtain a sufficient condition, that we shall utilize shortly, for a unique normal inverse.

**Proposition 7.6** Let $\mathcal{M}_{12} = (S_1, S_2, \Sigma_{12})$ be an invertible s-t tgd mapping. Assume that whenever $A$ is a source atom and $\delta'$ is a demanding conjunction for $A$ with formulas $\text{const}(x)$ precisely for the variables $x$ of $A$, then chase$_{12}(I_A) \rightarrow I_{\delta'}$. Then $\mathcal{M}_{12}$ has a unique normal inverse.

**Proof** We shall make use of Theorem 7.5. Let $A$ be arbitrary source atom. Assume that $\delta$ is an essential conjunction for $A$, and $\delta'$ is a demanding conjunction for $A$, both with formulas $\text{const}(x)$ for exactly the variables $x$ that appear in $A$; we must show that $I_{\delta} \rightarrow I_{\delta'}$. Since $\delta$ is relevant for $A$, we have $I_{\delta} \rightarrow \text{chase}_{12}(I_A)$. By assumption, we have chase$_{12}(I_A) \rightarrow I_{\delta'}$. So $I_{\delta} \rightarrow I_{\delta'}$, as desired. \qed

Let us say that a full s-t tgd mapping is onto if every target instance is the result of chasing some source instance. That is, the full s-t tgd mapping $\mathcal{M}_{12} = (S_1, S_2, \Sigma_{12})$ is onto if for every target instance $J$ there is a source instance $I$ such that chase$_{12}(I) = J$. Note that the mapping $\mathcal{M}_{12}$ of Example 7.1 is onto, whereas the mapping $\mathcal{M}_{12}$ of Example 7.2 is not onto.

**Theorem 7.7** A full s-t tgd mapping that is invertible and onto has a unique normal inverse.

We defer the proof until later, when we have developed more tools.
As an example of the use of Theorem 7.7, let us consider the mapping $M_{12}$ of Example 7.1. This mapping is invertible and onto, and so has a unique normal inverse by Theorem 7.7.

Does the converse hold? That is, is every full s-t tgd mapping with a unique normal inverse necessarily onto? The next example shows that this is false.

**Example 7.8** Let $S_1$ consist of four unary relation symbols $P_i$, for $1 \leq i \leq 4$, and let $S_2$ consist of the four unary relation symbols $Q_i$, for $1 \leq i \leq 4$ and the unary relation symbol $R$. Let $\Sigma_{12}$ consist of the full s-t tgds $P_i(x) \rightarrow Q_i(x)$, for $1 \leq i \leq 4$, along with the full s-t tgds $P_1(x) \wedge P_2(x) \rightarrow R(x)$ and $P_3(x) \wedge P_4(x) \rightarrow R(x)$. Let $M_{12} = (S_1, S_2, \Sigma_{12})$. The mapping $M_{12}$ is not onto, since the target instance whose set of facts is $\{Q_1(0), Q_2(0)\}$ is not a solution for any source instance $I$ (such an instance $I$ must contain the facts $P_1(0), P_2(0)$, and so every solution for $I$ must also contain the fact $R(0)$). Let $M_{21} = (S_2, \bar{S}_1, \Sigma_{21})$, where $\Sigma_{21} = \{Q_i(x) \wedge \text{const}(x) \rightarrow \bar{P}_i(x) : 1 \leq i \leq 4\}$. Although $M_{12}$ is not onto, we now show (by using Proposition 7.6) that $M_{12}$ has a unique normal inverse, namely $M_{21}$.

Let $A$ be a source atom. By symmetry of the roles of the source atoms, we can assume without loss of generality that $A$ is the source atom $P_1(x)$. Let $\delta'$ be a demanding conjunction for $A$ with $\text{const}$ formula $\text{const}(x)$ (and no other $\text{const}$ formula). We now show that $\delta'$ must contain $Q_1(x)$. Assume not; we shall derive a contradiction.

Now $I_A = \{P_1(c_2)\}$, where $c_2$ is the constant associated with the variable $x$ as in Section 2. Let $d$ be a constant different from $c_2$. Let $I$ consist of the facts $P_i(d)$ for $1 \leq i \leq 4$, along with the facts $P_i(c_2)$ for $2 \leq i \leq 4$. To chase $\Sigma_{12}(I)$ consists of the facts $Q_i(d)$ for $1 \leq i \leq 4$, along with the facts $\bar{P}_i(c_2)$ and $R(d)$. Since the only $\text{const}$ formula in $\delta'$ is $\text{const}(x)$, it follows that $I_{\delta'}$ contains only one constant, namely the constant $c_2$, and possibly also null values. Let $h$ be a function where $h(c_2) = c_2$ and $h(n) = d$ for every null $n$. Since $\delta'$ does not contain $Q_1(x)$, it follows that $I_{\delta'}$ does not contain $Q_1(c_2)$. So $I_{\delta'}$ contains some subset of $\{Q_2(c_2), Q_3(c_2), Q_4(c_2)\}$, possibly along with some facts $Q_i(n)$ for some nulls $n$ and for $1 \leq i \leq 4$, possibly along with $R(c_2)$, and possibly some facts $R(n)$ for some nulls $n$. Hence, $h$ is a homomorphism that maps $I_{\delta'}$ to chase $\Sigma_{12}(I)$. Since $I_{\delta'} \rightarrow \text{chase}_{12}(I)$ but $I_A \not\subset I$, this contradicts the assumption that $\delta'$ is demanding for $A$. This contradiction shows that $\delta'$ must contain $Q_1(x)$. Hence, chase $\Sigma_{12}(I_A) \subset I_{\delta'}$. So by Proposition 7.6, it follows that $M_{12}$ has a unique normal inverse. \qed

Can a nonfull s-t tgd mapping have a unique normal inverse? The next example shows that the answer is “yes”.

**Example 7.9** Let $S_1$ consist of two unary relation symbols $P$ and $Q$, and let $S_2$ consist of the binary relation symbol $R$. Let $\Sigma_{12}$ consist of the s-t tgds $P(x) \rightarrow \exists yR(x, y)$, $Q(y) \rightarrow \exists xR(x, y)$, and $P(x) \wedge Q(y) \rightarrow R(x, y)$. Note that the first tgd is not a logical consequence of the third tgd, since if the $P$ relation in the source is nonempty and the $Q$ relation in the source is empty, then the first tgd fires but the third tgd does not fire. Similarly, the second tgd is not a logical consequence of the third tgd. Let $M_{12} = (S_1, S_2, \Sigma_{12})$, and let $M_{21}$ be the canonical candidate inverse of $M_{12}$. Thus, $M_{21} = (S_2, \bar{S}_1, \Sigma_{21})$, where $\Sigma_{21}$ consists of the normal constraints $R(x, y) \wedge \text{const}(x) \rightarrow \bar{P}(x)$ and $R(x, y) \wedge \text{const}(y) \rightarrow \bar{Q}(y)$.

It is straightforward to verify that $M_{21}$ is an inverse of $M_{12}$. We now show that $M_{21}$ is the unique normal inverse of $M_{12}$. Let $M_{21} = (S_2, \bar{S}_1, \Sigma_{21})$ be another normal inverse of $M_{12}$. By Proposition 6.4, we know that $\Sigma_{21}$ logically implies $\Sigma_{21}'$. So we need only show that $\Sigma_{21}'$ logically implies $\Sigma_{21}$. Let $\tau$ be an arbitrary member of $\Sigma_{21}$: we must show that $\Sigma_{21}'$ logically implies $\tau$. By symmetry in the roles of $P$ and $Q$, we can assume without loss of generality that the conclusion $Q$ of $\tau$ is of the form $P(x)$, where $x$ is a variable. By normality of $\tau$, we know that $\text{const}(x)$ appears in $\tau$. We now show that there is a variable $y$ such that the relational atom $R(x, y)$ appears in $\tau$. Assume not; we shall derive a contradiction.

We now define a ground instance $I$. For each relational atom $R(v, w)$ in $\delta$, let $P(c_v)$ and $Q(c_w)$ be facts of $I$, where $c_v$ and $c_w$ are constants, as before. Because of the tgd $P(x) \wedge Q(y) \rightarrow R(x, y)$ in $\Sigma_{12}$, we see that the fact $R(c_v, c_w)$ is in chase $\Sigma_{12}(I)$ for each relational atom $R(v, w)$ in $\delta$. Thus $\delta$ is onto, since it follows by construction of $I$ that $I_A \not\subset I$. Because $I_{\delta} \rightarrow \text{chase}_{12}(I)$ and $I_A \not\subset I$, we know that $\delta$ is not demanding for $A$. But $\delta \rightarrow A$ is
in \( \Sigma_{21} \), so there is a violation of condition (1) of Theorem 5.13 (where the weak renaming \( f \) is the identity map). This is our desired contradiction. Therefore, there is a variable \( y \) such that the relational atom \( R(x, y) \) appears in \( \delta \).

Since \( \delta \) contains \( R(x, y) \) and \( \text{const}(x) \), it follows easily that \( \tau \) is a logical consequence of the tgd \( R(x, y) \land \text{const}(x) \rightarrow \check{P}(x) \), which is in \( \Sigma_{21} \). So \( \Sigma_{21} \) logically implies \( \tau \), as desired. This completes the proof that \( \mathcal{M}_{21} \) is the unique normal inverse of \( \mathcal{M}_{12} \).

Let \( \Sigma'_{12} \) consist of the first two s-t tgds in \( \Sigma_{12} \); that is, \( \Sigma'_{12} \) consists of the s-t tgds \( P(x) \rightarrow \exists y R(x, y) \) and \( Q(y) \rightarrow \exists x R(x, y) \). Let \( \mathcal{M}'_{12} = (S_1, S_2, \Sigma'_{12}) \). The curious reader may wonder whether \( \mathcal{M}'_{12} \), like \( \mathcal{M}_{12} \), has a unique normal inverse. The answer is “no”, as we now show. The canonical candidate inverse of \( \mathcal{M}'_{12} \) is the same as the canonical candidate inverse of \( \mathcal{M}_{12} \), and it is straightforward to see that this canonical candidate inverse is indeed an inverse of \( \mathcal{M}'_{12} \). We now give another, inequivalent inverse. Let \( \Sigma_{21} \) consist of the normal constraints \( R(x, y) \land \text{const}(x) \rightarrow \check{P}(x) \) and \( R(x, y) \land \text{const}(y) \rightarrow \check{Q}(y) \), as before, along with a third normal constraint \( R(x,x) \land R(x,y) \land \text{const}(y) \rightarrow \check{P}(y) \). Let \( \mathcal{M}'_{21} = (S_2, S_1, \Sigma_{21}) \). Since the conclusion of the third constraint is \( \check{P}(y) \) rather than \( \check{Q}(y) \), this third constraint is not a logical consequence of the first two. However, since \( R(x, x) \) does not arise in a chase, it is easy to see that \( \text{chase}_{21}'(\text{chase}_{12}(I)) = I \) for each ground instance \( I \), and so \( \mathcal{M}'_{21} \) is another normal inverse of \( \mathcal{M}_{12} \). \( \square \)

As we see from Example 7.8, being invertible and onto is not a necessary and sufficient condition for a full s-t tgd mapping to have a unique normal inverse. Is there a language with a richer set of constructs such that being invertible and onto is a necessary and sufficient condition for a full s-t tgd mapping to have a unique inverse in this language? We now give such a language.

**Definition 7.10** A disjunctive tgd with inequalities is a constraint of the form \( \alpha \land \eta \rightarrow \beta \), where \( \alpha \) is a conjunction of source atoms, \( \beta \) is a disjunction of formulas of the form \( \exists \beta' \) with \( \beta' \) a conjunction of target atoms, and \( \eta \) is a conjunction (possibly empty) of inequalities of the form \( x \neq y \) for distinct variables \( x, y \) of \( \beta \) that are not existentially quantified. Further, there is the safety condition that every variable in \( \beta \) that is not existentially quantified must appear in \( \alpha \). Again, we have suppressed writing the leading universal quantifiers. \( \square \)

Disjunctive tgds with inequalities were defined in [FKPT08], where they were shown to be rich enough to specify quasi-inverses of quasi-invertible full s-t tgd mappings.\(^6\) It was also shown there that inequalities in the premise, and both disjunctions and existential quantifiers in the conclusion, are needed in general to specify quasi-inverses of quasi-invertible full s-t tgd mappings. Note that \( \text{const} \) formulas are not part of the syntax. Every invertible full s-t tgd mapping has an inverse specified in this language, even without the disjunctions, namely the canonical candidate inverse with the \( \text{const} \) formulas dropped. The reason it is all right to drop the \( \text{const} \) formulas is because of a simple result in [FKPT08], which we now state, which says that \( \text{const} \) formulas play no role for inverses of full s-t tgd mappings.

**Proposition 7.11** [FKPT08] Let \( \mathcal{M}_{12} = (S_1, S_2, \Sigma_{12}) \) be a full s-t tgd mapping. Let \( \mathcal{M}_{21} = (S_2, S_1, \Sigma_{21}) \), where \( \Sigma_{21} \) is a set of disjunctive s-t tgds with constants and inequalities. Let \( \mathcal{M}'_{21} = (S_2, S_1, \Sigma'_{21}) \), where \( \Sigma'_{21} \) is obtained from \( \Sigma_{21} \) by removing every \( \text{const} \) formula. Let \( I \) and \( J \) be ground instances. Then \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \) if and only if \( (I, J) \models \Sigma_{12} \circ \Sigma'_{21} \). In particular, \( \mathcal{M}_{21} \) is an inverse of \( \mathcal{M}_{12} \) if and only if \( \mathcal{M}'_{21} \) is an inverse of \( \mathcal{M}_{12} \).

We now give a variant of Proposition 4.6 that holds for inverses specified by disjunctive tgds with inequalities and that we shall find useful later.

\(^6\)In the definition of disjunctive tgds with inequalities that is given in [FKPT08], inequalities \( x \neq y \) are allowed in the premise for any pair \( x, y \) of variables in the premise, and not just for variables \( x, y \) that appear in the conclusion and are not existentially quantified. However, the disjunctive tgds with inequalities used in their construction of quasi-inverses all satisfy our restriction on variables that may appear in inequalities. This restriction is not needed for our results here, but we make this restriction by analogy with our restriction on inequalities that we have for normal mappings, that the inequalities must involve only variables in the conclusion.
**Proposition 7.12** Assume that \( \mathcal{M}_{12} = (\mathbf{S}_1, \mathbf{S}_2, \Sigma_{12}) \) is a full s-t tgd mapping, \( \mathcal{M}_{21} = (\mathbf{S}_2, \mathbf{S}_1, \Sigma_{21}) \) is an inverse of \( \mathcal{M}_{12} \), and \( \Sigma_{21} \) is a set of disjunctive tgds with inequalities. Let \( I \) be a ground instance, and let \( U = \text{chase}_{12}(I) \). Then \( (U, \hat{I}) \models \Sigma_{21} \), and \( U \) witnesses \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \) when \( \hat{I} \subseteq J \).

**Proof** Since \( \mathcal{M}_{21} \) is an inverse of \( \mathcal{M}_{12} \), we know that \( (I, \hat{I}) \models \Sigma_{12} \circ \Sigma_{21} \) and therefore there exists some \( K \) such that \( (I, K) \models \Sigma_{12} \) and \( (K, \hat{I}) \models \Sigma_{21} \). Since \( U \) is a universal solution for \( I \) with respect to \( \mathcal{M}_{12} \), there is a homomorphism \( h : U \rightarrow K \) that is the identity on \( I \). Pick a constraint \( \phi \in \Sigma_{21} \); by assumption, it must be of the form

\[
\alpha(\bar{x}, \bar{y}) \land \eta(\bar{x}) \rightarrow \psi(\bar{x}),
\]

where \( \eta(\bar{x}) \) is a conjunction of existentially-quantified conjunctions, and where the variables in \( \psi(\bar{x}) \) that are not existentially quantified are exactly the members of \( \bar{x} \). Assume that \( U \) satisfies the premise of \( \phi \) on \( \bar{a}, \bar{b} \). That is, \( U \models \alpha(\bar{a}, \bar{b}) \land \eta(\bar{a}) \). Since \( h \) is a homomorphism from \( U \) to \( K \), we have that \( K \models \alpha(h(\bar{a}), h(\bar{b})) \). Now every member of \( U \) is a constant, since \( U = \text{chase}_{12}(I) \) and \( \Sigma_{12} \) is full. Therefore \( h(\bar{a}) = \bar{a} \) and \( h(\bar{b}) = \bar{b} \), and so \( K \models \alpha(\bar{a}, \bar{b}) \land \eta(\bar{a}) \). Since \( (K, \hat{I}) \models \Sigma_{21} \), we must have \( \hat{I} \models \psi(\bar{a}) \). This shows that \( (U, \hat{I}) \models \phi \). Since \( \phi \) is an arbitrary member of \( \Sigma_{21} \), it follows that \( (U, \hat{I}) \models \Sigma_{21} \), as desired. Since \( \hat{I} \subseteq J \), it follows easily that \( (U, J) \models \Sigma_{21} \). Since \( (I, U) \models \Sigma_{12} \) and \( (U, J) \models \Sigma_{21} \), we have that \( U \) witnesses \( (I, J) \models \Sigma_{12} \circ \Sigma_{21} \), as desired. \( \square \)

Recall that the *copy mapping*, which is used to define the inverse, is the schema mapping \( \text{Id} = (\mathbf{S}, \mathbf{S}, \Sigma_{\text{id}}) \), where \( \Sigma_{\text{id}} \) consists of the s-t tgds \( R(\bar{x}) \rightarrow \bar{R}(\bar{x}) \) as \( R \) ranges over the relation symbols in \( \mathbf{S} \). We now define a *p-copy mapping* (where the \( p \) stands for “pseudo” or “permutation”) that is a generalization of the copy mapping.

**Definition 7.13** The schema mapping \( (\mathbf{S}, \mathbf{T}, \Sigma) \) is a *p-copy mapping* if:

1. Every member of \( \Sigma \) is of the form

\[
P(x_1, \ldots, x_k) \rightarrow Q(x_{f(1)}, \ldots, x_{f(k)}),
\]

where \( P \) is a source relation symbol, \( Q \) is a target relation symbol, \( x_1, \ldots, x_k \) are distinct variables, and \( f \) is a permutation of \( \{1, \ldots, k\} \).

2. Every source relation symbol appears in exactly one premise of \( \Sigma \).

3. Every target relation symbol appears in exactly one conclusion of \( \Sigma \). \( \square \)

For example, assume that \( \mathbf{S}_1 \) consists of the binary relation symbol \( P_1 \) and the ternary relation symbol \( P_2 \), and \( \mathbf{S}_2 \) consists of the binary relation symbol \( Q_1 \) and the ternary relation symbol \( Q_2 \). Assume that \( \Sigma_{12} \) consists of the s-t tgds \( P_1(x, y) \rightarrow Q_1(y, x) \) and \( P_2(x, y, z) \rightarrow Q_2(y, x, z) \). Then \( (\mathbf{S}_1, \mathbf{S}_2, \Sigma_{12}) \) is a p-copy mapping.

The next theorem says that disjunctive tgds with inequalities form a rich enough language that a full s-t tgd mapping has a unique inverse in this language if and only if it is invertible and onto. It also says that the only full s-t tgd mappings with these properties are p-copy mappings.

**Theorem 7.14** Let \( \mathcal{M}_{12} = (\mathbf{S}_1, \mathbf{S}_2, \Sigma_{12}) \) be a full s-t tgd mapping. The following are equivalent.

1. \( \mathcal{M}_{12} \) has a unique inverse specified by disjunctive tgds with inequalities.
2. \( \mathcal{M}_{12} \) is invertible and onto.
3. \( \mathcal{M}_{12} \) is equivalent to a p-copy mapping.
Proof We begin by showing that (3) $\Rightarrow$ (2). From (3), we know that there is a schema mapping $M_{12}'$ that is equivalent to $M_{12}$ and that is a p-copy mapping. Clearly, $M_{12}$ is invertible and onto. It follows easily that $M_{12}$ is invertible and onto, as desired.

We now show that (2) $\Rightarrow$ (1). Assume that (2) holds. Since $M_{12}$ is invertible, Theorem 6.3 tells us that the canonical candidate inverse is indeed an inverse of $M_{12}$, so $M_{12}$ has a normal inverse. By Proposition 7.11, the const formulas are irrelevant, and so $M_{12}$ has an inverse specified by tgd s with inequalities, and hence by disjunctive tgd s with inequalities. Now assume that $M_{21} = (S_2, S_1, \Sigma_21)$ and $M_{21} = (S_2, S_1, \Sigma_21)$ are both inverses of $M_{12}$, where $\Sigma_{21}$ and $\Sigma_{21}'$ are disjunctive tgd s with inequalities. We must show that $\Sigma_{21}$ and $\Sigma_{21}'$ are logically equivalent. We now show that $\Sigma_{21}$ logically implies $\Sigma_{21}'$. By symmetry, we have that $\Sigma_{21}'$ logically implies $\Sigma_{21}$. Assume that $(J, K) \models \Sigma_{21}$. We must show that $(J, K) \models \Sigma_{21}'$. By replacing each null in $(J, K)$ by a new constant, we obtain $(J', K')$ where every entry of every tuple is a constant, such that $(J', K')$ is isomorphic to $(J, K)$ (but where the isomorphism may map constants into either constants or nulls, and may map nulls into either constants or nulls). Since $\Sigma_{21}$ has no const formulas, it follows easily that $(J', K') \models \Sigma_{21}$. Since $M_{12}$ is onto, there is a ground instance $I$ such that $J' = \text{chase}_{12}(I)$. So $(I, J') \models \Sigma_{12}$. Since also $(J', K') \models \Sigma_{21}$, we have that $(I, K') \models \Sigma_{12} \circ \Sigma_{21}$. Therefore, since $M_{21}$ is an inverse of $M_{12}$, we have that $I \subseteq K'$. Hence, since $M_{21}'$ is an inverse of $M_{12}$, we have that $(I, K') \models \Sigma_{12} \circ \Sigma_{21}'$. So by Proposition 7.12, we have that $(J', K') \models \Sigma_{21}'$. Since $\Sigma_{21}$ has no const formulas, we have as before $(J, K) \models \Sigma_{21}'$. This was to be shown.

We conclude by showing that (1) $\Rightarrow$ (3). Assume that (1) holds. We must show that $M_{12}$ is equivalent to a p-copy mapping. By Theorem 5.19, we know that every source atom has an essential target atom. Let $e$ be a function that maps every prime source atom $A$ onto a target atom $e(A)$ that is essential for $A$. By Theorem 5.15, we know that $(S_2, S_1, \Sigma_{21})$ is an inverse of $A$. Let $\Sigma_{21}'$ be the result of removing every const formula from every member of $\Sigma_{21}$, and let $M_{21} = (S_2, S_1, \Sigma_{21}')$. Note that every member of $\Sigma_{21}'$ is of the form $B \land \eta_A \rightarrow \hat{A}$, where $B$ is an essential atom for $A$, and where $\eta_A$ consists of all inequalities of the form $x \neq y$ where $x$ and $y$ are distinct variables of $A$. Note that by Proposition 5.20, the variables in $\eta_A$ and $A$ are the same, so $\eta_A$ has inequalities among all distinct variables of $B$ also. By Proposition 7.11 we know that $M_{21}'$ is an inverse of $M_{12}$. Since (1) holds, $M_{21}'$ is the unique inverse that is specified by disjunctive tgd s with inequalities. We now prove the following claim:

Claim 1: $\text{chase}_{12}(I_A)$ is a singleton for each source atom $A$.

Assume that $A$ is a source atom where $\text{chase}_{12}(I_A)$ is not a singleton; we shall derive a contradiction. Denote $e(A)$ by $B$. Thus, $B$ is essential for $A$. Since $B$ is relevant for $A$, we have that $I_B \subseteq \text{chase}_{12}(I_A)$, so $\text{chase}_{12}(I_A)$ is nonempty. Assume that $\text{chase}_{12}(I_A)$ has more than one member; we shall derive a contradiction. Now $\Sigma_{21}'$ contains the formula $B \land \eta_A \rightarrow \hat{A}$. Let $\nu_A$ be as in Definition 6.1. In particular, $\nu_A$ contains $B$ and at least one more distinct relational atom $B'$. Form $\Sigma_{21}$ from $\Sigma_{21}'$ by replacing $B \land \eta_A \rightarrow \hat{A}$ by $\nu_A \land \eta_A \rightarrow \hat{A}$, and let $M_{21} = (S_2, S_1, \Sigma_{21})$. Since $B$ is demanding for $A$, it is easy to see that $\nu_A$ is demanding for $A$. It follows from Theorem 5.13 (and the fact that every weak renaming consistent with $\nu_A$ is a strict renaming, since $A$ and $\nu_A$ have the same variables) that $M_{21}$ is an inverse of $M_{12}$. We now show that $\Sigma_{21}$ is not logically equivalent to $\Sigma_{21}'$.

Assume that $B$ is an atom $Q(x_1, \ldots, x_4)$, and that $F$ is a fact $Q(a_1, \ldots, a_4)$. Let us say that $F$ is an exact match for $B$ if $a_i = a_j$ if and only if $x_i$ and $x_j$ are the same variable, for all $i, j$. Similarly, we define what it means for one atom to be an exact match for another atom (with the same relational symbol). Let $J$ consist of a single fact $F$ that is an exact match for $B$. We now show that $(J, \emptyset) \not\models \Sigma_{21}$, but $(J, \emptyset) \models \Sigma_{21}$. So $\Sigma_{21}$ is not logically equivalent to $\Sigma_{21}'$, as desired. The fact that $(J, \emptyset) \not\models \Sigma_{21}$ follows from the fact that $\Sigma_{21}'$ contains the formula $B \land \eta_A \rightarrow \hat{A}$, and $J$ contains a fact that is an exact match for $B$. It remains to show that $(J, \emptyset) \models \Sigma_{21}$. Let $\sigma$ be a member of $\Sigma_{21}$ except for $\nu_A \land \eta_A \rightarrow \hat{A}$. Since $J$ consists of a single fact that is an exact match for $B$, it follows that $\sigma$ does not fire on $J$, because otherwise $F$ would be an exact match for the atom $B'$ in the premise of $\sigma$, and so $B'$ and $B$ would be an exact match, which is not possible since they are essential for atoms that are not an exact match for each other. So $(J, \emptyset) \models \sigma$. We now show that $(J, \emptyset) \models \nu_A \land \eta_A \rightarrow \hat{A}$. To show this, we must show that $\nu_A \land \eta_A \rightarrow \hat{A}$ does not fire on $J$. If it were to fire on $J$, then there would be a mapping $h$ on the variables in $\nu_A$ that maps each atom in $\nu_A$ onto $F$ and (because of $\eta_A$) is one-to-one on the variables in $\nu_A$. Recall that $B'$ is a relational atom in $\nu_A$ other than $B$. Since $B$ and $B'$ map onto the same fact $F$, it follows that
\(B'\) is, like \(B\), a \(Q\)-atom. Assume that \(B'\) is \(Q(x_{i_1}, \ldots, x_{i_t})\), where each \(x_{i_r}\) is in \(\{x_1, \ldots, x_t\}\). Assume that \(F\) is the fact \(Q(a_1, \ldots, a_t)\). Now \(h(x_{i_r}) = a_r = h(x_r)\) for each \(r\), since both \(B\) and \(B'\) map onto \(F\). Since \(h\) is one-to-one on variables, it follows that \(x_{i_r}\) and \(x_r\) are the same variable for each \(r\). So \(B'\) and \(B\) are the same atom, a contradiction. This contradiction shows that there is no \(\alpha\) such that \(\alpha \rightarrow \neg \alpha\). Let \(\alpha\) be a function with domain the variables in the premise \(\alpha\) and conclusion \(e(A)\) are the same by Proposition 5.20. Since \(e(A)\) is essential for \(A\), it follows that \(e(A)\) is essential for \(A\), so \(I_{e(A)} \subseteq \text{chase}_{12}(I_A)\). Therefore, from Claim 1, we see that \(I_{e(A)} = \text{chase}_{12}(I_A)\). It is clear that \(\Sigma_{12}\) logically implies \(\Sigma'_{12}\). Later, we shall show that \(\Sigma_{12}\) is logically equivalent to \(\Sigma'_{12}\). First, we prove another claim.

**Claim 2:** Assume that \(A\) and \(B\) are atoms, and that \(B\) is essential for \(A\) with respect to \(\Sigma_{12}\). Then \(\Sigma'_{12}\) logically implies the \(s-t\) tgd \(A \rightarrow B\).

Assume that Claim 2 were false; we shall derive a contradiction. Assume that \(A\) is a \(P\)-atom. Define \(\nu'_A\) like \(\nu_A\), except that the chase is with \(\Sigma'_{12}\) instead of \(\Sigma_{12}\). Let \(B'\) be \(\nu'_A\). Note that \(B'\) is a singleton atom, because it arises only by firing the \(s-t\) tgd in \(\Sigma_{12}\) whose premise is the \(P\)-atom with all variables distinct. Since \(\Sigma_{12}\) does not logically imply the \(s-t\) tgd \(A \rightarrow B\), we know that \(B\) is different from \(B'\). Since \(B'\) is derived as the result of a chase with \(\Sigma_{12}\), and since \(\Sigma_{12}\) logically implies \(\Sigma'_{12}\), it follows that \(B'\) is in \(\nu_A\). So \(\nu_A\) contains at least the two distinct atoms \(B\) and \(B'\). This contradicts Claim 1, which is our desired contradiction.

**Claim 3:** \(\Sigma_{12}\) is logically equivalent to \(\Sigma'_{12}\).

We already noted that \(\Sigma_{12}\) logically implies \(\Sigma'_{12}\), so Claim 3 is proven if we show that \(\Sigma'_{12}\) logically implies \(\Sigma_{12}\). Assume not; we shall derive a contradiction. We can assume without loss of generality that every member of \(\Sigma_{12}\) has a singleton conclusion. Let \(\alpha \rightarrow B\) be a member of \(\Sigma_{12}\) that is not a logical consequence of \(\Sigma'_{12}\). We now show that there is no atom \(A\) such that \(B\) is essential for \(A\) with respect to \(\Sigma_{12}\). Assume that there were. By Claim 2, it follows that \(\Sigma_{12}\) logically implies \(A \rightarrow B\). Since \(\alpha \rightarrow B\) is a member of \(\Sigma_{12}\), we have \(I_B \subseteq \text{chase}_{12}(I_{\alpha})\). Therefore, since \(B\) is demanding for \(A\) with respect to \(\Sigma_{12}\), it follows that \(I_A \subseteq I_{\alpha}\), that is, the atom \(A\) appears in \(\alpha\). Hence, since \(\Sigma'_{12}\) logically implies \(A \rightarrow B\), we have that \(\Sigma'_{12}\) logically implies \(\alpha \rightarrow B\), a contradiction. This contradiction shows that there is no atom \(A\) such that \(B\) is essential for \(A\) with respect to \(\Sigma_{12}\).

Assume that \(B\) is the atom \(Q(x_{i_1}, \ldots, x_{i_t})\) where \(x_{i_1}, \ldots, x_{i_t}\) are variables (not necessarily distinct). Let \(\tau\) be an arbitrary member of \(\Sigma_{12}\) of the form \(\delta \rightarrow Q(z_{i_1}, \ldots, z_{i_t})\), where \(z_{i_1}, \ldots, z_{i_t}\) are variables, not necessarily distinct, and where \(x_{i_1}\) and \(x_{i_2}\) are the same variable whenever \(z_{i_1}\) and \(z_{i_2}\) are the same variable. Let \(h_\tau\) be a function with domain the variables in \(\tau\) such that \(h_\tau(z_i) = x_i\) for each \(i\) (this is well-defined, since \(x_{i_1}\) and \(x_{i_2}\) are the same variable whenever \(z_{i_1}\) and \(z_{i_2}\) are the same variable), and where \(h_\tau\) maps each variable in the premise \(\delta\) of \(\tau\) that is not in the conclusion \(Q(z_{i_1}, \ldots, z_{i_t})\) of \(\tau\) onto a new variable. Let \(\tau'\) be the image of \(\tau\) under \(h_\tau\). Thus, \(\tau'\) is a weak renaming of \(\tau\). Note that \(\tau'\) is not necessary a strict renaming of \(\tau\), since two distinct variables \(z_i\) and \(z_j\) will map onto the same variable if \(x_i\) and \(x_j\) are the same variable. It is straightforward to verify that if \(I\) is a source instance and the chase of \(I\) with \(\tau\) produces a fact \(F\) that is an exact match for \(B\), then the chase of \(I\) with \(\tau'\) also produces \(F\).

By construction, the conclusion of \(\tau'\) is \(B\). Assume that the premise of \(\tau'\) is \(\delta'\). Let \(\psi_\tau\) be the formula \(\exists y \hat{\delta}'\), where \(y\) consists of the variables in \(\tau'\) that are not in \(B\). Let \(Z\) consist of all such formulas \(\psi_\tau\). In particular, \(Z\) contains \(\exists y \hat{\delta}\), where \(y\) consists of all variables in \(\alpha\) that are not in \(B\). Now \(Z\) is finite, since its size is at most the number of members of \(\Sigma_{12}\). Let \(\gamma\) be the formula \(B \land \eta \rightarrow \zeta\), where \(\eta\) is the conjunction of all inequalities of the form \(x_i \neq x_j\) where \(x_i\) and \(x_j\) are distinct variables in \(B\) (and hence distinct variables in \(\alpha\)), and where \(\zeta\) is the disjunction of the members of \(Z\). Let \(\Sigma_{21} = \Sigma_{21}' \cup \{\gamma\}\), and let \(\mathcal{M}_{21} = (S_2, \hat{S}_1, \Sigma_{21})\).

We now show that \(\mathcal{M}_{21}\) is an inverse of \(\mathcal{M}_{12}\), and that \(\Sigma_{21}\) is not logically equivalent to \(\Sigma'_{21}\). To show that
M_{21} is an inverse of M_{12}, we must show that for all ground instances I and J:

\[(I, J) |= \Sigma_{12} \circ \Sigma_{21} \text{ if and only if } \widehat{I} \subseteq J.\]  
(4)

Since M_{21}^E is an inverse of M_{12}, we know that for all ground instances I and J:

\[(I, J) |= \Sigma_{12} \circ \Sigma_{21}^E \text{ if and only if } \widehat{I} \subseteq J.\]  
(5)

Now \Sigma_{21} logically implies \Sigma_{21}^E, since \Sigma_{21} is a superset of \Sigma_{21}^E. It follows easily that \Sigma_{12} \circ \Sigma_{21} logically implies \Sigma_{12} \circ \Sigma_{21}^E. So if \( (I, J) |= \Sigma_{12} \circ \Sigma_{21} \), then \( (I, J) |= \Sigma_{12} \circ \Sigma_{21}^E \), which, by (5), implies that \( \widehat{I} \subseteq J \). Assume now that \( \widehat{I} \subseteq J \); we must show that \( (I, J) |= \Sigma_{12} \circ \Sigma_{21} \). Let \( J^* = \text{chase}_{12}(I) \). Since \( (I, J^*) |= \Sigma_{12} \), we need only show that \( (J^*, J) |= \Sigma_{21} \). Now \( (J^*, I) |= \Sigma_{21}^E \), by Proposition 7.12, and so \( (J^*, J) |= \Sigma_{21}^E \) since \( \widehat{I} \subseteq J \). Therefore, since \( \Sigma_{21} = \Sigma_{21}^E \cup \{\gamma\} \), we need only show that \( (J^*, J) |= \gamma \). Assume that \gamma fires on \( J^* \). Then there is a one-to-one mapping \( h \) (one-to-one because of \eta) from the variables of \( B \) to constants, that maps \( B \) onto a fact \( F \) of \( J^* \). So \( F \) is an exact match for \( B \). Since \( F \) is in \( J^* \), there is a member \( \tau \) of \( \Sigma_{12} \) that generates \( F \) in the chase of \( I \) with \( \Sigma_{12} \). Let \( \tau^* \) and \( \delta^* \) be as before. Since \( F \) is an exact match for \( B \), it follows from an earlier comment that the chase of \( I \) with \( \tau^* \) generates \( F \). It follows fairly easily that \( \exists y \delta^* \) is satisfied in \( J \) under \( h \), so \( \exists y \delta^* \) is satisfied in \( \widehat{I} \) under \( h \). Since \( \widehat{I} \subseteq J \), it follows that \( \exists y \delta^* \) is satisfied in \( J \) under \( h \). But \( \exists y \delta^* \) is a disjunct in the conclusion of \gamma. Therefore, \( (J^*, J) |= \gamma \), as desired. This concludes the proof that \( M_{21} \) is an inverse of \( M_{12} \).

We now show that \( \Sigma_{21} \) is not logically equivalent to \( \Sigma_{21}^E \). It is clear that \( (I_B, \emptyset) \not|= \gamma \), and so \( (I_B, \emptyset) \not|= \Sigma_{21} \). We now show that \( (I_B, \emptyset) |= \Sigma_{21}^E \). Since, as we showed, there is no atom \( A \) such that \( B \) is essential for \( A \) with respect to \( \Sigma_{12} \), no member of \( \Sigma_{21} \) has a strict renaming of \( B \) and satisfies \( \eta_A \). It follows that \( (I_B, \emptyset) |= \Sigma_{21}^E \), as desired.

We have shown that \( \Sigma_{12} \) has two distinct, inequivalent inverses given by disejunctive tgd with inequalities, namely \( M_{21}^E \) and \( M_{21} \). This contradiction shows that Claim 3 holds.

We now state and prove our final claim.

**Claim 4:** For every target relation symbol \( Q \), there is exactly one member of \( \Sigma_{12}^t \) whose conclusion is a \( Q \)-atom. No two variables appearing in this \( Q \)-atom are the same.

To prove this claim, we begin by showing that there must be at least one member of \( \Sigma_{12}' \) whose conclusion is a \( Q \)-atom, where \( Q \) is the only \( Q \) variable appearing in this \( Q \)-atom. Assume not; we shall derive a contradiction. Let \( \gamma \) be the formula \( Q(x_1, \ldots, x_t) \land \eta \rightarrow \beta \), where the variables \( x_1, \ldots, x_t \) are distinct, where \( \eta \) is a conjunction of the inequalities \( x_i \neq x_j \) whenever \( i \neq j \), and where \( \beta \) is an arbitrary disjunction of source atoms whose variables altogether are exactly \( x_1, \ldots, x_t \).

Let \( \Sigma_{21} = \Sigma_{21}^E \cup \{\gamma\} \), and let \( M_{21} = (S_2, S_1, \Sigma_{21}) \).

We now show that \( M_{21} \) is an inverse of \( M_{12} \), and that \( \Sigma_{21} \) is not logically equivalent to \( \Sigma_{21}^E \). Let \( J^* = \text{chase}_{12}(I) \). Let \( K = \text{chase}_{12}^t(I) \), the result of chasing \( I \) with \( \Sigma_{21}^t \). Since (by Claim 3) \( \Sigma_{12} \) is logically equivalent to \( \Sigma_{21}^t \), and since both \( \Sigma_{12} \) and \( \Sigma_{21}^t \) are full, it follows that \( K = J^* \) (both \( K \) and \( J^* \) are the unique, null-free core of the universal solutions for \( I \)). As in the proof of Claim 3, to show that \( M_{21} \) is an inverse of \( M_{12} \), we need only show that \( (J^*, J) |= \gamma \). Since by assumption there is no member of \( \Sigma_{12} \) whose conclusion is a \( Q \)-atom with no two variables appearing in this \( Q \)-atom the same, and since, as we showed, \( J^* = \text{chase}_{12}^t(I) \) it follows easily that \( J^* \) does not contain a \( Q \)-fact \( Q(c_1, \ldots, c_t) \), where \( c_1, \ldots, c_t \) are distinct constants. So \( \gamma \) does not fire on \( J^* \), and so \( (J^*, J) |= \gamma \), as desired.

We now show that \( \Sigma_{21} \) is not logically equivalent to \( \Sigma_{21}^E \). Let \( J \) consist of the singleton fact \( Q(c_1, \ldots, c_t) \), where \( c_1, \ldots, c_t \) are distinct constants. Then \( (J, \emptyset) \not|= \gamma \), and so \( (J, \emptyset) \not|= \Sigma_{21} \). We now show that \( (J, \emptyset) |= \Sigma_{21}^E \). Each member of \( \Sigma_{21}^E \) is of the form \( e(A) \land \eta_A \rightarrow A \). Since, by assumption, there is no member of \( \Sigma_{12}^t \) whose conclusion is a \( Q \)-atom with no two variables appearing in this \( Q \)-atom the same, and since each atom in the premise of a member of \( \Sigma_{21}^E \) is a conclusion of a member of \( \Sigma_{12}^t \), it follows that no atom in the premise of a member of \( \Sigma_{21}^E \) is a \( Q \)-atom with no two variables appearing in this \( Q \)-atom the same. Therefore, no member of \( \Sigma_{21}^E \) fires on \( J \). So \( (J, \emptyset) |= \Sigma_{21}^E \), as desired.

\footnote{A disjunction is required if no source atom has arity at least \( t \).}
We have shown that $M_{12}$ has two distinct, inequivalent inverses given by disjunctive tgds with inequalities, namely $M^I_{21}$ and $M_{21}$. This contradiction shows that there must be at least one member $\sigma$ of $\Sigma_{12}'$ whose conclusion is a $Q$-atom with no two variables appearing in this $Q$-atom the same.

We now show that there can be no other member $\sigma'$ of $\Sigma_{12}'$ whose conclusion is a $Q$-atom. Assume not; we shall derive a contradiction. Assume that the premise of $\sigma$ is a $P$-atom and the premise of $\sigma'$ is a $P'$-atom. Since $\sigma$ and $\sigma'$ are different, we know that $P$ and $P'$ are different, by construction of $\Sigma_{12}'$. Since no two variables appearing in the conclusion of $\sigma$ are the same, there is a mapping $h$ that maps the variables in $\sigma$ to the variables in $\sigma'$ that maps the conclusion of $\sigma$ onto the conclusion of $\sigma'$. Let $A$ be the $P$-atom that is the result of applying $h$ to the premise of $\sigma$. So the chase of $I_A$ with $\Sigma_{12}'$ is $I_{P'}$, where $B'$ is the conclusion of $\sigma'$. This contradicts the fact that conclusion of $\sigma'$ is essential for the premise of $\sigma'$. This contradiction shows that the only member of $\Sigma_{12}$ whose conclusion is a $Q$-atom is $\sigma$, where every variable in the conclusion is distinct. This completes the proof of Claim 4.

As we noted earlier, the variables in the source and target of each member of $\Sigma_{12}'$ are the same by Proposition 5.20, since the target is essential for the source with respect to $\Sigma_{12}$. By construction, for every source relation symbol $P$, there is exactly one member of $\Sigma_{12}'$ whose premise is a $P$-atom, and every variable is distinct in this $P$-atom. By Claim 4, for every target relation symbol $Q$, there is exactly one member of $\Sigma_{12}'$ whose conclusion is a $Q$-atom, and no two variables appearing in this $Q$-atom are the same. It follows that $M_{12}$ is a p-copy mapping. Also, by Claim 3, $\Sigma_{12}$ is logically equivalent to $\Sigma_{12}'$. So (3) holds, as desired. This completes the proof that (1) $\Rightarrow$ (3). □

Note that we cannot replace (2) in the statement of the theorem by simply “$M_{12}$ is onto”, because of the schema mapping with source relation symbols $P$ and $R$ and the single target relation symbol $Q$, that is specified by the tgds $P(x) \rightarrow Q(x)$, $R(x) \rightarrow Q(x)$. This schema mapping is clearly onto but not invertible.

We can now give the proof of Theorem 7.7.

**Proof of Theorem 7.7:**

Assume that $M_{12} = (S_1, S_2, \Sigma_{12})$ is a full s-t tgd mapping that is invertible and onto. By Theorem 7.14, we know that $M_{12}$ is equivalent to a p-copy mapping. We can assume without loss of generality that $M_{12}$ itself is a p-copy mapping. Since $M_{12}$ is invertible, we know that the canonical candidate inverse is a normal inverse of $M_{12}$. We now use Proposition 7.6 to show that $M_{12}$ has a unique normal inverse. To apply Proposition 7.6, assume that $A$ is a source atom and $\delta'$ is a demanding conjunction for $A$ with formulas $\text{const}(x)$ precisely for the variables $x$ of $A$; we must show that $\text{chase}_{12}(I_A) \subseteq I_{\delta'}$.

If $\Sigma_{12}$ includes the tgd $P(x_1, \ldots, x_k) \rightarrow Q(x_{f(1)}, \ldots, x_{f(k)})$, and if $y_1, \ldots, y_k$ are variables, not necessarily distinct, then let us refer to the atoms $P(y_1, \ldots, y_k)$ and $Q(y_{f(1)}, \ldots, y_{f(k)})$ as *buddies*. Let $\chi_A$ be the conjunction of the formulas $\text{const}(x)$ for the variables $x$ of $A$. Let $\gamma$ be the conjunction of the buddies of the relational atoms in $\delta'$, and let $\gamma' = \gamma \land \chi_A$. Since $\delta'$ is demanding for $A$, and since $I_{\delta'} = \text{chase}_{12}(I_{\gamma'})$, it follows that $I_A \subseteq I_{\gamma'}$. So $\text{chase}_{12}(I_A) \subseteq \text{chase}_{12}(I_{\gamma'})$. Clearly $\text{chase}_{12}(I_{\gamma'}) = I_{\delta'}$. Hence, $\text{chase}_{12}(I_A) \subseteq I_{\delta'}$, as desired. □

Let us reconsider the schema mapping $M_{12}$ from Example 7.8. It has a unique normal inverse, but since $M_{12}$ is not equivalent to a p-copy mapping, it follows from Theorem 7.14 that $M_{12}$ does not have a unique inverse specified by disjunctive tgds with inequalities. In addition to $M_{21} = (S_2, S_1, \Sigma_{21})$ from Example 7.8, another inverse is specified by $\Sigma_{21}$ along with the disjunctive tgd $R(x) \rightarrow (P_1(x) \lor P_3(x))$.

Define a *near p-copy mapping* to be a full s-t tgd mapping $M = (S, T, \Sigma)$ where (i) for each member $\sigma$ of $\Sigma$, the premise and conclusion of $\sigma$ are each singletons, with the same variables in the premise as in the conclusion, and with the variables in the conclusion all distinct, and where (ii) every member of $T$ appears in the conclusion of exactly one member of $\Sigma$, and every member of $S$ appears in the premise of at most one member of $\Sigma$. Thus, a near p-copy mapping may differ from being a p-copy mapping for two reasons. First, the variables in the premise are not necessarily distinct. Second, some member of $S$ may fail to appear in $\Sigma$. By the *constants-added version* of an s-t tgd mapping, we mean the mapping that results by adding to the premise of every tgd the formulas $\text{const}(x)$ for every variable $x$ that appears in the conclusion. Returning again to Example 7.8, we see that the unique normal inverse $M_{21} = (S_2, S_1, \Sigma_{21})$ is the constants-added version of a near p-copy mapping (it is only
“near”, because the relation symbol \( R \) does not appear in \( \Sigma_{21} \). This is not a coincidence. As a consequence of a later result (Theorem 10.2) that relates the number of normal inverses to the number of constraints in an inverse, we obtain the following result, whose proof we shall give in Section 10.

**Theorem 7.15** If a full s-t tgd mapping has a unique normal inverse \( \mathcal{M}_{21} \), then \( \mathcal{M}_{21} \) is equivalent to the constants-added version of a near p-copy mapping.

It is straightforward to verify that the schema mapping \( \mathcal{M}_{21} \) in Example 7.9 is not equivalent to the constants-added version of a near p-copy mapping. Since also \( \mathcal{M}_{21} \) is the unique normal inverse of a schema mapping, it follows that Theorem 7.15 fails in the nonfull case.

In addition to the fact that Theorem 7.14 allows disjunctions in the inverse and Theorem 7.15 does not, another difference between the two theorems is that Theorem 7.14 characterizes the mapping \( \mathcal{M}_{12} \), whereas Theorem 7.15 characterizes the inverse mapping \( \mathcal{M}_{21} \).

We close this section with an explanation of why const formulas are not allowed in the language for inverses used in Theorem 7.14. Would the theorem still be true if we were to enrich the language for inverses still further to be disjunctive tgds with constants and inequalities? It turns out that uniqueness is then hopeless. For example, consider the schema mapping \( \mathcal{M}_{12} \) of Example 7.1. Let \( \sigma_1 \) be the constraint \( S(x) \land \text{const}(x) \rightarrow R(x) \), and let \( \sigma_2 \) be the constraint \( S(x) \rightarrow \exists y R(y) \). In addition to the inverse \( \mathcal{M}_{21} \) given in Example 7.1, which is specified by \( \sigma_1 \), another inverse is specified by \( \{\sigma_1, \sigma_2\} \). The constraint \( \sigma_2 \) is not logically implied by the constraint \( \sigma_1 \), because of the const formula in the premise of \( \sigma_1 \) but not \( \sigma_2 \). More generally, if there were a full, invertible s-t tgd mapping \( \mathcal{M}'_{12} \) with a unique inverse specified by disjunctive tgds with constants and inequalities, then all the more so it would have a unique inverse specified by disjunctive tgds with inequalities (there is at least one such inverse, namely the canonical candidate inverse). So from the implication (1) \( \Rightarrow \) (3) of Theorem 7.14, it would follow that \( \mathcal{M}'_{12} \) is equivalent to a p-copy mapping. But then the obvious generalization of the construction we just gave for a second inverse of \( \mathcal{M}_{12} \) of Example 7.1 would show that \( \mathcal{M}'_{12} \) has inequivalent inverses specified by disjunctive tgds with constants and inequalities, a contradiction.

### 8 Inverse of the Inverse

In this section, we consider the question as to when a normal inverse of a schema mapping is itself invertible. Surprisingly, it turns out to be rare that a normal inverse of an s-t tgd mapping is invertible. We focus on the full case, and then show by example that our results do not hold in the nonfull case.

We now give an example of a schema mapping with an invertible normal inverse.

**Example 8.1** Let \( \mathcal{M}_{12} \) and \( \mathcal{M}_{21} \) be as in Example 7.1. Then \( \mathcal{M}_{21} \) is a normal inverse of \( \mathcal{M}_{12} \), and \( \mathcal{M}_{12} \) is an inverse of \( \mathcal{M}_{21} \). So \( \mathcal{M}_{21} \) is an invertible normal inverse of \( \mathcal{M}_{12} \). \( \square \)

More generally, it is straightforward to see that every p-copy mapping has an invertible normal inverse. Thus, as in Example 8.1, let \( \mathcal{M}_{12} \) be a p-copy mapping, and let \( \mathcal{M}_{21} \) be obtained from \( \mathcal{M}_{12} \) by “reversing the arrows” and adding const formulas to the premises. Then, as in Example 8.1, \( \mathcal{M}_{21} \) is a normal inverse of \( \mathcal{M}_{12} \), and \( \mathcal{M}_{12} \) is an inverse of \( \mathcal{M}_{21} \). The next theorem tells us that p-copy mappings are the only full s-t tgd mappings with an invertible normal inverse. Before we state and prove this theorem, we need a simple lemma from [Fag07].

**Lemma 8.2** [Fag07] Let \( \mathcal{M}_{12} \) be an invertible s-t tgd mapping. If \( I_1 \neq I_2 \), then \( \text{chase}_{12}(I_1) \neq \text{chase}_{12}(I_2) \).

We now give the main result of this section.

**Theorem 8.3** Let \( \mathcal{M}_{12} \) be a full s-t tgd mapping. Then \( \mathcal{M}_{12} \) has an invertible normal inverse if and only if \( \mathcal{M}_{12} \) is equivalent to a p-copy mapping.
Lemma 8.2 (where $M$ chase $I$ also for normal mappings, by the same proof). So property and its homomorphic version are shown to be equivalent to invertibility for s-t tgd mappings, this holds invertible, it satisfies (the homomorphic version of) the subset property, as given in Section 3 (although the subset constraint $\delta$ of weakly renaming $B$ includes $A$. Thus, let $M_{21} = (S_2, S_1, \Sigma_{21})$ be an invertible normal inverse of $M_{12}$. We shall show that $M_{12}$ is equivalent to a p-copy mapping. Whenever we speak of relevant, demanding, or essential atoms in this proof, we mean with respect to $\Sigma_{12}$. We shall reserve $A$ and $A'$ for source atoms (with relation symbols in $S_1$), and $B$ and $B'$ for target atoms (with relation symbols in $S_2$).

Claim 1: If $B$ is a relevant atom for a source atom $A$, then $\text{chase}_{21}(I_B) = \widehat{I_A}$.

We now prove Claim 1. Assume that $B$ is a relevant atom for $A$. Now $\text{chase}_{21}(I_B)$ is nonempty, since otherwise $\text{chase}_{21}(I_B) = \text{chase}_{21}(\emptyset)$, which is impossible by Lemma 8.2 (where $M_{21}$ plays the role of $M_{12}$). Since $M_{21}$ is full, we know that that $\text{chase}_{21}(I_B)$ has no nulls, and so every fact in $\text{chase}_{21}(I_B)$ is of the form $\widehat{I}_{A'}$ for some atom $A'$. The claim is proven if we show that whenever $\widehat{I}_{A'} \subseteq \text{chase}_{21}(I_B)$, then $A'$ is the same atom as $A$. So assume that $\widehat{I}_{A'} \subseteq \text{chase}_{21}(I_B)$. By Lemma 5.17, where the role of $A$ is played by $A'$, we know that $B$ is demanding for $A'$. Since $B$ is also relevant for $A$, it follows from Lemma 5.16 that $A'$ is the same atom as $A$, as desired.

Claim 2: Each source atom $A$ has exactly one relevant atom, and $B$ is essential for $A$.

We now prove Claim 2. Since $M_{12}$ is invertible, it follows from Theorem 5.19 that $A$ has some essential atom $B$. So $B$ is relevant for $A$. Assume that $A$ has another relevant atom $B'$; we shall derive a contradiction. By Claim 1, we have that $\text{chase}_{21}(I_B)$ and $\text{chase}_{21}(I_{B'})$ both equal $\widehat{I_A}$, and so are equal. But this is impossible by Lemma 8.2 (where $M_{21}$ plays the role of $M_{12}$).

Let us denote the unique relevant atom for $A$ by $B_A$. For the next claim, recall that if $\varphi$ is a formula, and $f$ is a weak renaming, then $\varphi^f$ is the result of replacing every variable $x$ in $\varphi$ by $f(x)$.

Claim 3: Let $f$ be a weak renaming. Then $(B_A)^f = B_{A'}$.

We now prove Claim 3. Assume that $A' = A$. Since $B_A$ is relevant for $A$, it is clear that the result $(B_A)^f$ of weakly renaming $B_A$ using $f$ is relevant for $A'$. That is, $(B_A)^f$ is relevant for $A'$. By definition, the unique relevant atom for $A'$ is $B_A$. Therefore, $(B_A)^f = B_A = B_{A'}$, as desired.

Claim 4: Let $B$ be a target atom. Then $B$ is relevant for some source atom.

We now prove Claim 4. We prove it first when every variable in $B$ is distinct. Since $M_{21}$ is invertible, we know that $\text{chase}_{21}(I_B) \neq \emptyset$, since otherwise $\text{chase}_{21}(I_B) = \text{chase}_{21}(\emptyset)$, which is impossible by Lemma 8.2 (where $M_{21}$ plays the role of $M_{12}$). So there is some member $\delta \rightarrow \widehat{A}$ of $\Sigma_{21}$ that fires on $I_B$. Hence, $\text{chase}_{21}(I_\delta)$ includes $\widehat{I_A}$. By Claim 1, we have that $\text{chase}_{21}(I_B) = \widehat{I_A}$. So $\text{chase}_{21}(I_{B_A}) \subseteq \text{chase}_{21}(I_\delta)$. Since $M_{21}$ is invertible, it satisfies (the homomorphic version of) the subset property, as given in Section 3 (although the subset property and its homomorphic version are shown to be equivalent to invertibility for s-t tgd mappings, this holds also for normal mappings, by the same proof). So $I_{B_A} \subseteq I_\delta$. Therefore, $\delta$ has $B_A$ as a conjunct. Since the constraint $\delta \rightarrow \widehat{A}$ of $\Sigma_{21}$ fires on $I_B$, there is a homomorphism from $B_A$ to $B$. Since every variable in $B$ is distinct, it follows that $B_A$ and $B$ are the same up to a renaming of variables. Therefore, since $B_A$ is relevant for $A$, we know that $B$ is relevant for some atom obtained by renaming the variables of $A$. This completes the proof of Claim 4 when all of the variables in $B$ are distinct.

Let $B'$ be a target atom where the variables need not be distinct. Let $B$ be an atom where all of the variables are distinct and where $B'$ is obtained from $B$ by a weak renaming $f$, that is, $B' = B^f$. Since all of the variable in $B$ are distinct, it follows from what we have shown that $B$ is relevant for some source atom $A$, and so $B$ is simply $B_A$. Therefore, $B' = (B_A)^f$. Hence, by Claim 3, we know that $B'$ is $B_{A'}$. So $B'$ is relevant for $A'$. 

Claim 5: Let $A$ be a source atom with all variables distinct. Then every variable in $B_A$ is distinct.

We now prove Claim 5. Since $B_A$ is essential for $A$, it follows from Proposition 5.11 that $B_A$ has exactly the same variables as $A$. We now show that every variable in $B_A$ is distinct. Assume not; we shall derive a contradiction. Let $B'$ be an atom with the same relation symbol as $B_A$ but with every variable distinct. So $B'$ has strictly more variables than $A$. Since also every variable in $A$ is distinct, and every variable in $B'$ is distinct, it
follows that the arity of $B'$ is strictly bigger than the arity of $A$. By Claim 4, $B'$ is relevant for some source atom $A'$. Since by Claim 2 we know that $A'$ has a unique relevant atom, and this atom is essential for $A'$, it follows that $B'$ is essential for $A'$. So by Proposition 5.11, we know that $B'$ and $A'$ have the same variables. Therefore, since every variable in $B'$ is distinct, the arity of $A'$ is at least the arity of $B'$, which as we noted is strictly bigger than the arity of $A$. So the arity of $A'$ is strictly bigger than the arity of $A$. Since $B_A$ is obtained from $B'$ by a weak renaming, and $B'$ is relevant for $A'$, it follows that $B_A$ is relevant for an atom $A''$ obtained from $A'$ by a weak renaming. Since $B_A$ is demanding for $A$, it follows from Lemma 5.16 that $A''$ and $A$ are the same atom. But this is impossible, since $A''$ has the same arity as $A'$, and the arity of $A'$ is strictly bigger than the arity of $A$. This is our desired contradiction. This completes the proof of Claim 5.

Let $\Sigma'_{12}$ consist of all of the constraints $A \rightarrow B_A$, where $A$ is a prime atom with all variables distinct. Let $\mathcal{M}'_{12} = (S_1, S_2, \Sigma'_{12})$.

Claim 6: $\Sigma_{12}$ and $\Sigma'_{12}$ are logically equivalent.

We now prove Claim 6. Clearly $\Sigma_{12}$ logically implies $\Sigma'_{12}$. We now show that $\Sigma'_{12}$ logically implies $\Sigma_{12}$. We first show that each of the constraints $A' \rightarrow B_{A'}$ is a logical consequence of $\Sigma'_{12}$. Let $A$ be an atom with the same relation symbol as $A'$ and with all variables distinct. So there is a renaming $f$ where $A'$ is $A^f$. By Claim 3, we know that $(B_A)^f = B_{A'}$. So $A' \rightarrow B_{A'}$ is $(A \rightarrow B_A)^f$. Therefore $A' \rightarrow B_{A'}$ is a logical consequence of $A \rightarrow B_A$, and so of $\Sigma_{12}$, as desired.

Assume now that $\varphi \rightarrow B$ is a member of $\Sigma_{12}$. By Claim 4, we know that $B$ is $B_A$ for some source atom $A$. Since chase$_{12}(I_A)$ is $I_{B_A}$, and chase$_{12}(I_A)$ includes $I_{B_A}$, it follows that chase$_{12}(I_A) \subseteq$ chase$_{12}(I_A)$, and therefore chase$_{12}(I_A) \rightarrow$ chase$_{12}(I_A)$. So by the homomorphic version of the subset property, $I_A \subseteq I_A$. Therefore, $A$ is in $\varphi$. So $\varphi \rightarrow B$ is a logical consequence of $A \rightarrow B$, that is, of $A \rightarrow B_A$, and so is a logical consequence of $\Sigma'_{12}$. This completes the proof of Claim 6.

We conclude by showing that $\mathcal{M}'_{12}$ is a p-copy mapping. Let $A \rightarrow B_A$ be a member of $\Sigma'_{12}$. By construction, every variable in $A$ is distinct. As noted earlier, $B_A$ has exactly the same variables as $A$, and by Claim 5, every variable in $B_A$ is distinct. By construction, every source relation symbol appears in exactly one premise of $\Sigma'_{12}$. To complete the proof that $\mathcal{M}'_{12}$ is a p-copy mapping, all that is left to show is that every target relation symbol appears in exactly one conclusion of $\Sigma'_{12}$.

Let $Q$ be an arbitrary target relation symbol, and let $B'$ be a $Q$-atom with every variable distinct. By Claim 4, we have that $B'$ is relevant for some source atom $A'$, and so $B'$ equals $B_{A'}$. Assume that $A'$ is a P-atom. Let $A$ be the prime $P$-atom with all variables distinct. Let $f$ be a weak renaming where $A'$ is $A^f$. By Claim 3, we know that $B_{A'}$, that is, $B'$, is $(B_A)^f$. Hence, since $B'$ is a $Q$-atom, so is $B_A$. So $Q$ appears in some conclusion of $\Sigma'_{12}$.

We now show that $Q$ cannot be in more than one conclusion in $\Sigma'_{12}$. Say $Q$ were in the conclusion of the member of $\Sigma'_{12}$ whose premise has relation symbol $P$ and also in the conclusion of the member of $\Sigma'_{12}$ whose premise has relation symbol $P'$. Let $F$ be the fact $P(0, \ldots, 0)$, where every variable is set to 0. Similarly, let $F'$ be the fact $P'(0, \ldots, 0)$, where every variable is set to 0. Then the result of chasing $F$ with $\Sigma'_{12}$ is $Q(0, \ldots, 0)$, and identically the result of chasing $F'$ with $\Sigma'_{12}$ is $Q(0, \ldots, 0)$. But this is impossible by Lemma 8.2, since $\Sigma_{12}$ and $\Sigma'_{12}$ are logically equivalent by Claim 6. This concludes the proof that $\mathcal{M}'_{12}$ is a p-copy mapping.

In Theorem 7.14, we characterized full s-t tgd mappings that have a unique inverse specified by disjunctive tgds with inequalities, by showing that these are exactly those schema mappings equivalent to a p-copy mapping. In Theorem 8.3, we characterized full s-t tgd mappings with an invertible normal inverse by showing that these, too, are exactly those schema mappings equivalent to a p-copy mapping. We thereby obtain the unexpected result that a full s-t tgd mapping has a unique inverse specified by disjunctive tgds with inequalities if and only if it has an invertible normal inverse. The next theorem states this equivalence (along with the other equivalences that we obtain from Theorems 7.14 and 8.3).

Theorem 8.4 Let $\mathcal{M}_{12}$ be a full s-t tgd mapping. The following are equivalent.

1. $\mathcal{M}_{12}$ has a unique inverse specified by disjunctive tgds with inequalities.
2. $\mathcal{M}_{12}$ has an invertible normal inverse.
3. $M_{12}$ is equivalent to a p-copy mapping.

4. $M_{12}$ is invertible and onto.

What about the nonfull case? The technical condition (4) of Theorem 8.4 no longer makes sense then, because we have not even defined what it means for a nonfull mapping to be onto. We now show by example that this equivalence of (1) and (2) in Theorem 8.4 fails when we drop the assumption that $M_{12}$ be full.

**Example 8.5** Let $S_1$ consist of the unary relation symbols $P_1$ and $P_2$, and let $S_2$ consist of the unary relation symbols $Q_1$ and $Q_2$. Let $\Sigma_{12}$ consist of the s-t tgds $P_1(x) \rightarrow Q_1(x)$ and $P_2(x) \rightarrow \exists y(Q_2(x) \land Q_1(y))$. Let $\Sigma_{12}$ consist of the s-t tgds $P_1(x) \rightarrow Q_1(x)$ and $P_2(x) \rightarrow Q_2(x)$. Let $\Sigma_{21}$ consist of the normal constraints $Q_1(x) \land \text{const}(x) \rightarrow \overline{P_1}(x)$ and $Q_2(x) \land \text{const}(x) \rightarrow \overline{P_2}(x)$. Let $\Sigma'_{21}$ consist of the normal constraints $Q_1(x) \land \text{const}(x) \rightarrow \overline{P_1}(x)$ and $Q_2(x) \land Q_1(y) \land \text{const}(x) \rightarrow \overline{P_2}(x)$. Let $M_{12} = (S_1, S_2, \Sigma_{12})$, let $M'_{12} = (S_1, S_2, \Sigma'_{12})$, let $M_{21} = (S_2, S_1, \Sigma_{21})$, and let $M'_{21} = (S_2, S_1, \Sigma'_{21})$. It is straightforward to verify that $M_{21}$ and $M'_{21}$ are inequivalent normal inverses of $M_{12}$, and $M'_{12}$ is an inverse of $M_{21}$. So condition (2) of Theorem 8.4 holds, since $M_{12}$ is a normal inverse of $M_{12}$, and $M'_{12}$ is an inverse of $M_{21}$. However, condition (1) of Theorem 8.4 fails, since $M_{12}$ has two inequivalent normal inverses, namely $M_{21}$ and $M'_{21}$. Furthermore, it is not hard to see that condition (3) of Theorem 8.4 fails also. □

### 9 The Size of an Inverse, and Complexity of Computing an Inverse

In this section, we consider the question of whether there is a polynomial-size inverse in some language. The exact definition of size is not very important. For simplicity, we take the size of a relational atom to be its arity, the size of an inequality or a const formula to be 1, the size of a formula to be the sum of the sizes of the relational atoms, inequalities, and const formulas in it, and the size of a schema mapping $(S, T, \Sigma)$ to be the sum of the sizes of the members of $\Sigma$. We show that there is a family of invertible, full s-t tgd mappings $M$ where the minimal number of constraints in a normal inverse of $M$ is exponential in the size of $M$ (and so, since the size of a mapping is at least equal to the number of constraints in it, the normal inverse of $M$ with the smallest size has size exponential in $|M|$). We also show, however, that if we expand the language to allow the premise to contain Boolean combinations of equalities rather than simply conjunctions of inequalities, then every invertible, full s-t tgd mapping has a polynomial-size inverse, that can be computed in polynomial time. Note that we cannot tell if the output of this polynomial-time algorithm is actually an inverse, since (by Corollary 3.4), the complexity of deciding invertibility, even in the full case, is coNP-complete. Instead, what we know is that if the schema mapping $M$ is invertible, then the output of this polynomial-time algorithm is an inverse of $M$. It is an interesting open problem as to whether similar results can be obtained for s-t tgd mappings that are not full.

**Theorem 9.1** There is a family of full s-t tgd mappings, each of which is invertible, but where the minimal number of constraints in a normal inverse is exponential in the size of the schema mapping.

**Proof** The family is parameterized by the positive integer $k$. Let $S_1 = \{P_0, \ldots, P_k\}$, and let $S_2 = \{P'_0, \ldots, P'_k, Q_0, \ldots, Q_{k-1}\}$. Assume that all of the relation symbols in $S_1$ and $S_2$ are $4k$-ary.

Let $x_1, \ldots, x_{4k}$ be distinct variables. Let $S_1$ consist of the s-t tgds $P_1(x_1, x_2, \ldots, x_{4k}) \rightarrow P'_1(x_1, x_2, \ldots, x_{4k})$, for $0 \leq i \leq k$. Define $\bar{x}^0$, for $0 \leq i \leq k - 1$, by letting $x_{4i+2}^0 = x_{4i+1}$, $x_{4i+4}^0 = x_{4i+3}$, and $x_j^0 = x_j$ if $j \notin \{4i+2, 4i+4\}$. For example, $\bar{x}^0$ is

$$(x_1, x_1, x_3, x_3, x_5, x_6, \ldots, x_{4k-1}, x_{4k}),$$

and $\bar{x}^1$ is

$$(x_1, x_2, x_3, x_4, x_5, x_7, x_7, x_9, x_{10}, \ldots, x_{4k-1}, x_{4k}).$$
Let $S_2$ consist of the s-t tgds $P_{i+1}(\bar{x}^i) \rightarrow P'_0(\bar{x}^i)$, for $0 \leq i \leq k - 1$. Let $S_3$ consist of the s-t tgds $P_0(\bar{x}^i) \rightarrow Q_i(\bar{x}^i)$, for $0 \leq i \leq k - 1$. Let $\Sigma_{12} = S_1 \cup S_2 \cup S_3$, and let $M_{12} = (S_1, S_2, \Sigma_{12})$.

We begin by showing that $M_{12}$ is invertible. Let $T_1$ consist of the s-t tgds $P'_0(x_1, x_2, \ldots, x_{4k}) \rightarrow \hat{P}_j(x_1, x_2, \ldots, x_{4k})$, for $1 \leq j \leq k$ (note that we do not include the case $j = 0$). Let $T_2$ consist of the formula

$$P'_0(x_1, x_2, \ldots, x_{4k}) \land ((x_1 \neq x_2) \lor (x_3 \neq x_4)) \land ((x_5 \neq x_6) \lor (x_7 \neq x_8)) \land \ldots \land ((x_{4k-3} \neq x_{4k-2}) \lor (x_{4k-1} \neq x_{4k})) \rightarrow \hat{P}_0(x_1, x_2, \ldots, x_{4k}).$$

Let $T_3$ consist of the s-t tgds $Q_i(\bar{x}^i) \rightarrow \hat{P}_0(\bar{x}^i)$, for $0 \leq i \leq k-1$. Let $\Sigma_{21} = T_1 \cup T_2 \cup T_3$, and let $M_{21} = (S_2, \widehat{S_1}, \Sigma_{21})$.

Note that $\text{chase}_{21}$ is well-defined, even in the presence of the formula in $T_2$.

We now show that $M_{21}$ is an inverse of $M_{12}$. It is sufficient to show that $\widehat{I} = \text{chase}_{21}(\text{chase}_{12}(I))$ for each ground instance $I$ (this is because the analogue of Theorem 4.10 holds, by the same proof). We first show that $\widehat{I} \subseteq \text{chase}_{21}(\text{chase}_{12}(I))$. If $\hat{P}_1(a)$ is a fact of $\widehat{I}$ (and so $P_1(a)$ is a fact of $I$), and if $1 \leq i \leq k$, then we see from the tgds in $S_1$ and $T_1$ that $\hat{P}_i(a)$ is in $\text{chase}_{21}(\text{chase}_{12}(I))$. So assume that $\hat{P}_0(a)$ is a fact of $\widehat{I}$ (and so $P_0(a)$ is a fact of $I$). There are two cases.

**Case 1:** There is $i$ with $0 \leq i \leq k-1$ such that $a_{4i+2} = a_{4i+1}$ and $a_{4i+4} = a_{4i+3}$. Then the s-t tgd $P_0(\bar{x}^i) \rightarrow Q_i(\bar{x}^i)$ in $S_3$ and the s-t tgd $Q_i(\bar{x}^i) \rightarrow \hat{P}_0(\bar{x}^i)$ in $T_3$ guarantee that $\hat{P}_0(a)$ is a fact of $\text{chase}_{21}(\text{chase}_{12}(I))$.

**Case 2:** There is no $i$ with $0 \leq i \leq k-1$ such that $a_{4i+2} = a_{4i+1}$ and $a_{4i+4} = a_{4i+3}$. Then the s-t tgd $P_0(x_1, x_2, \ldots, x_{4k}) \rightarrow P'_0(x_1, x_2, \ldots, x_{4k})$ in $S_1$ and the formula in $T_2$ guarantee that $\hat{P}_0(a)$ is a fact of $\text{chase}_{21}(\text{chase}_{12}(I))$.

We now show the reverse inclusion, that $\text{chase}_{21}(\text{chase}_{12}(I)) \subseteq \widehat{I}$. Because the premise of each member of $\Sigma_{12}$ and of $\Sigma_{21}$ contains a single relational atom, we see that each fact of $\text{chase}_{21}(\text{chase}_{12}(I))$ is obtained by chasing a single fact $P_1(a_1, a_2, \ldots, a_{4k})$ of $I$ with a single member $\sigma_1$ of $\Sigma_{12}$, and then chasing the single tuple that results from this chase by a single member $\sigma_2$ of $\Sigma_{21}$. We must show that the result of this second chase is either the empty set, or is the fact $\hat{P}_1(a_1, a_2, \ldots, a_{4k})$.

We now consider cases.

**Case 1:** $\sigma_1$ is the s-t tgd $P_0(x_1, x_2, \ldots, x_{4k}) \rightarrow P'_0(x_1, x_2, \ldots, x_{4k})$ of $S_1$. Assume that $\sigma_1$ was applied to the fact $P_0(a_1, a_2, \ldots, a_{4k})$ of $I$ to obtain the fact $P'_0(a_1, a_2, \ldots, a_{4k})$. The only member of $\Sigma_{21}$ whose premise contains $P'_0$ is the formula in $T_2$, and so we may assume that $\sigma_2$ is this formula. Since the conclusion of $\sigma_2$ is $\hat{P}_0(x_1, x_2, \ldots, x_{4k})$, it follows that chasing $P_0(a_1, a_2, \ldots, a_{4k})$ with $\sigma_2$ gives either the empty set or the fact $\hat{P}_0(a_1, a_2, \ldots, a_{4k})$, as desired.

**Case 2:** $\sigma_1$ is the s-t tgd $P_i(x_1, x_2, \ldots, x_{4k}) \rightarrow P'_i(x_1, x_2, \ldots, x_{4k})$ of $S_1$, for some $i$ with $1 \leq i \leq k$. Assume $\sigma_1$ was applied to the fact $P_i(a_1, a_2, \ldots, a_{4k})$ of $I$ to obtain the fact $P'_i(a_1, a_2, \ldots, a_{4k})$. The only member of $\Sigma_{21}$ whose premise contains $P'_i$ is the tgd $P'_i(x_1, x_2, \ldots, x_{4k}) \rightarrow \hat{P}_i(x_1, x_2, \ldots, x_{4k})$, and so we may assume that $\sigma_2$ is this tgd. Clearly, chasing $P'_i(a_1, a_2, \ldots, a_{4k})$ with $\sigma_2$ gives the fact $\hat{P}_i(a_1, a_2, \ldots, a_{4k})$, as desired.

**Case 3:** $\sigma_1$ is the s-t tgd $P_{i+1}(\bar{x}^i) \rightarrow P'_0(\bar{x}^i)$ of $S_2$, for some $i$ with $0 \leq i \leq k-1$. Assume that $\sigma_1$ was applied to the fact $P_{i+1}(a_1, a_2, \ldots, a_{4k})$ of $I$ to obtain the fact $P'_0(a_1, a_2, \ldots, a_{4k})$. So necessarily $a_{4i+2} = a_{4i+1}$ and $a_{4i+4} = a_{4i+3}$. The only member of $\Sigma_{21}$ whose
premise contains $P_0'$ is the formula in $T_2$, and so we may assume that $\sigma_2$ is this formula. But this formula is not fired by $P_0'(a_1, a_2, \ldots, a_{4k})$, since $a_{4i+2} = a_{4i+1}$ and $a_{4i+4} = a_{4i+3}$. So this case is not possible.

Case 4: $\sigma_1$ is the s-t tgd

$$P_0(\bar{x}) \rightarrow Q_1(\bar{x})$$

of $S_3$, for some $i$ with $0 \leq i \leq k - 1$. Assume that $\sigma_1$ was applied to the fact $P_0(a_1, a_2, \ldots, a_{4k})$ of $I$ to obtain $Q_1(\bar{a}, a_1, a_2, \ldots, a_{4k})$. The only member of $\Sigma_{21}$ whose premise contains $Q_1$ is the s-t tgd $Q_1(\bar{x}) \rightarrow P_0(\bar{x})$ of $T_3$, so we may assume that $\sigma_2$ is this s-t tgd. Clearly, the result of chasing $Q_1(a_1, a_2, \ldots, a_{4k})$ with $\sigma_2$ is either the empty set or the fact $P_0(a_1, a_2, \ldots, a_{4k})$, as desired.

This concludes the proof that $\text{chase}_{21}(\text{chase}_{12}(I)) \subseteq \hat{I}$, which was the final step in the proof that $M_{21}$ is an inverse of $M_{12}$.

We now show that the size of the smallest normal inverse of $M_{12}$ is exponential in the size of $M_{12}$. Assume that $M_{21}' = (S_2, S_1, \Sigma_{21}')$ is a normal inverse of $M_{12}$. It follows from Theorem 4.10 that for every ground instance $I$:

$$\hat{I} = \text{chase}_{21}(\text{chase}_{12}(I)).$$

Let us refer to $4i + 1$ and $4i + 3$ as buddies, for $0 \leq i \leq k - 1$. Let $\bar{a} = (a_1, \ldots, a_{4k})$ be a $4k$-tuple of constants. Let us call $\bar{a}$ special if:

1. for each pair $i_1, i_2$ of buddies, exactly one of the equalities $a_{i_1} = a_{i_1 + 1}$ or $a_{i_2} = a_{i_2 + 1}$ holds; and
2. these are the only equalities among members of $\bar{a}$ (that is, if $a_i = a_j$ for distinct values $i, j$, then there is an odd $t$ such that $\{i, j\} = \{t, t + 1\}$).

Let the equality profile of $\bar{a}$ be the $2k$-tuple $(\rho_1, \rho_2, \rho_5, \ldots, \rho_{4k-1})$ where $\rho_i = 0$ if $a_i = a_{i+1}$, and $\rho_i = 1$ if $a_i \neq a_{i+1}$. Let us say that an equality profile is special if it is the equality profile of a special tuple $\bar{a}$. It is easy to see that an equality profile is special precisely if exactly one of $\rho_1$ or $\rho_3$ is 0, exactly one of $\rho_5$ or $\rho_7$ is 0, etc.

For simplicity in what follows, when we say that an inequality $x_i \neq x_j$ appears in a formula, we mean that either the inequality $x_i \neq x_j$ or the inequality $x_j \neq x_i$ actually appears. Let $\sigma$ be a member of $\Sigma_{21}'$ whose conclusion is of the form $P_0(\bar{x})$, where $\bar{x} = (x_{m_1}, x_{m_2}, \ldots, x_{m_{4k}})$. For each odd number $i$ with $1 \leq i \leq 4k - 1$, let us say that $i$ is of type 1 with respect to $\sigma$ if $x_{m_i}$ and $x_{m_{i+1}}$ are distinct variables and the inequality $x_{m_i} \neq x_{m_{i+1}}$ appears in the premise of $\sigma$. If $i$ is not of type 1 with respect to $\sigma$, then let us say that $i$ is of type 0 with respect to $\sigma$. Thus, $i$ is of type 0 precisely if either (a) $x_{m_i}$ and $x_{m_{i+1}}$ are the same variable, or else (b) they are different variables and the inequality $x_{m_i} \neq x_{m_{i+1}}$ does not appear in the premise of $\sigma$.

Let $\rho = (\rho_1, \rho_3, \rho_5, \ldots, \rho_{4k-1})$ be a special equality profile, and let $\bar{a}$ be a special $4k$-tuple of constants with equality profile $\rho$. Let $I$ be a ground instance whose only fact is $P_0'(\bar{a})$. Now $\text{chase}_{12}(I)$ consists of the single fact $P_0'(\bar{a})$ (the s-t tgd $S_3$ cannot be applied in the chase since $\bar{a}$ is special). It follows from (6) that there must be a member $\sigma_\rho$ of $\Sigma_{21}'$ such that the chase of $P_0'(\bar{a})$ with $\sigma_\rho$ produces $P_0'(\bar{a})$. It is clear that $\sigma_\rho$ must have the following properties:

1. The conclusion of $\sigma_\rho$ is of the form $P_0'(\bar{x})$, where $\bar{x} = (x_{m_1}, x_{m_2}, \ldots, x_{m_{4k}})$;
2. variables $x_{m_r}$ and $x_{m_s}$ can be the same variable only if $a_{m_r} = a_{m_s}$;
3. $i$ is of type 0 with respect to $\sigma_\rho$ for each $i$ where $\rho_i = 0$; and
4. the only relation symbol that appears in the premise of $\sigma_\rho$ is $P_0'$.

We now show that for each odd $i$ with $1 \leq i \leq 4k - 1$, we have that $i$ is of type $\rho_i$ with respect to $\sigma_\rho$. We already have that if $\rho_i = 0$, then $i$ is of type 0 with respect to $\sigma_\rho$ (this follows from the third condition above for $\sigma_\rho$). So we need only show that if $\rho_i = 1$, then $i$ is of type 1 with respect to $\sigma_\rho$. Let $i_1 = i$, and let $i_2$ be the
buddy of $i$. Note for later use that $i_1$ is odd (because $i$ is odd, and $i_1 = i$). Since $a$ is special, and since $\rho_{i_1} = 1$, it follows that $\rho_{i_1} = 0$. Therefore, as noted before, $i_2$ is of type 0 with respect to $\sigma_\rho$. Assume that $i_1$ is also of type 0 with respect to $\sigma_\rho$; we shall derive a contradiction.

Let $h$ be a mapping from the variables in $\sigma_\rho$ to constants where $h(x_{m_i}) = a_{m_i}$ for each $i$ with $1 \leq i \leq 4k$. This function is well-defined by the second condition about $\sigma_\rho$.

We now show that $h$ respects each of the inequalities of $\sigma_\rho$, that is, that if $y \neq y'$ is an inequality that appears in the premise of $\sigma_\rho$, then $h(y) \neq h(y')$. Assume that $y \neq y'$ is an inequality that appears in the premise of $\sigma_\rho$, but $h(y) = h(y')$; we shall derive a contradiction. Since $h(y) = h(y')$, and since $\bar{a}$ is a special 4k-tuple of constants with equality profile $\rho$, we know that there is some odd $j$ such that $\{y, y'\} = \{x_{m_j}, x_{m_{j+1}}\}$ and $\rho_j = 0$. By property (3) above (with $j$ playing the role of $i$), we know that $j$ is of type 0 with respect to $\sigma_\rho$. Hence, the inequality $x_{m_j} \neq x_{m_{j+1}}$, that is, the inequality $y \neq y'$, does not appear in the premise of $\sigma_\rho$. This is our desired contradiction.

Let $J_1$ be the target instance that consists of all of the facts $P_0^\rho(h(y_1), \ldots, h(y_{4k}))$, where the atom $P_0^\rho(y_1, \ldots, y_{4k})$ appears in the premise of $\sigma_\rho$. Obtain $J_2$ from $J_1$ by replacing each occurrence of $a_{i+1}$ by $a_i$. Define $\bar{a}' = (a'_1, \ldots, a'_{4k})$ by letting $a'_{i+1} = a_i$, and letting $a'_j = a_j$ if $j \neq i + 1$. Since $a_{i+1} = a_i$ (because $\rho_{i+1} = 0$), we have $a'_{i+1} = a_{i+1} = a_i = a'_i$. Thus, $a'_{i+1} = a'_i$.

Define $h'$ by letting $h'(y) = h(y)$ if $y$ is not $x_{i+1}$, and letting $h'(x_{i+1}) = h(x_i)$, that is, $h'(x_{i+1}) = a_i$. So $J_2$ consists of all of the facts $P_0'(h'(y_1), \ldots, h'(y_{4k}))$, where the atom $P_0'(y_1, \ldots, y_{4k})$ appears in the premise of $\sigma_\rho$.

We now show that $h'$ respects each of the inequalities of $\sigma_\rho$. There are three cases.

Case 1: $\{y, y'\}$ does not contain $x_{i+1}$. Then $h'(y) = h(y)$ and $h'(y') = h(y)$. Now $h(y) \neq h(y')$, since $h$ respects the inequalities of $\sigma_\rho$. Therefore, $h'(y) \neq h'(y')$, as desired.

Case 2: $\{y, y'\}$ contains $x_{i+1}$ but not $x_i$. In particular, either $y$ or $y'$ is $x_{i+1}$, assume without loss of generality that $y$ is $x_{i+1}$. Then $h'(y) = h(x_i)$ and $h'(y') = h(y')$. Since $y$ is not $x_i$ or $x_{i+1}$, and $i$ is odd, we know by the fact that $\bar{a}$ is a special 4k-tuple that $h(y') \neq h(x_i)$. So $h'(y') = h(y') \neq h(x_i) = h'(y)$. Therefore, $h'(y) \neq h'(y')$, as desired.

Case 3: $\{y, y'\} = \{x_i, x_{i+1}\}$. This case is not possible, since $i_1$ is of type 0 with respect to $\sigma_\rho$, and so the inequality $x_{m_{i_1}} \neq x_{m_{i_1+1}}$ does not appear in the premise of $\sigma_\rho$.

Since $J_2$ consists of all of the facts $P_0'(h'(y_1), \ldots, h'(y_{4k}))$, where the atom $P_0'(y_1, \ldots, y_{4k})$ appears in the premise of $\sigma_\rho$, and since $h'$ respects each of the inequalities of $\sigma_\rho$, it follows that the chase of $J_2$ with $\sigma_\rho$ contains $P_0'(\bar{a}')$. Form the source instance $I_2$ from the target instance $J_2$ by replacing each fact $P_0'(\bar{a}')$ by $P_0'(\bar{a})$. Let $\tau_1$ be the s-tgd $P_0(x_1, x_2, \ldots, x_{4k}) \rightarrow P_0'(x_1, x_2, \ldots, x_{4k})$. Clearly, the chase of $I_2$ with $\tau_1$ is $J_2$.

Since $i_1$ and $i_2$ are buddies, there is $s$ with $0 \leq s \leq k - 1$ such that $\{i_1, i_2\} = \{4s + 1, 4s + 3\}$. Let $I''_s$ be the set difference $I_2 \setminus \{P_0'(\bar{a}')\}$, and let $J''_s$ be the set difference $J_2 \setminus \{P_0'(\bar{a}')\}$. Then the chase of $I''_s$ with $\tau_1$ is $J''_s$. Let $I''_s$ consist of the fact $P_{s+1}(\bar{a}')$. Let $\tau_2$ be the s-tgd $P_{s+1}(\bar{x}^s) \rightarrow P_0'(\bar{x}^s)$. Since $a'_{i+1} = a_{i+1}$ (by construction) and $a'_j = a_j$ (as noted earlier), the chase of $I''_s$ with $\tau_2$ contains the fact $P_0'(\bar{a}')$. Let $I_3 = I''_s \cup I''_s$. So the chase of $I_3$ with $\{\tau_1, \tau_2\}$ contains $J''_s \cup \{P_0'(\bar{a}')\}$, which contains $J_2$. Since $\tau_1$ and $\tau_2$ are members of $\Sigma_{12}$, it follows that the chase of $I_3$ with $\Sigma_{12}$ contains $J_2$. Since the chase of $I_3$ with $\Sigma_{12}$ contains $J_2$, and the chase of $J_2$ with $\sigma_\rho$ contains $P_0'(\bar{a}')$, it follows that the chase of $I_3$ with $\Sigma_{12}$ contains $P_0'(\bar{a}')$, which is not in $I_3$. Therefore, $I_3$ is not a variable instance of $\Sigma_{12}$, which contradicts (6) when $I$ is $I_3$. This is our desired contradiction.

We just showed that if $\rho = (\rho_1, \rho_2, \rho_3, \ldots, \rho_{4k-1})$ is a special equality profile, then there is a member $\sigma_\rho$ of $\Sigma_{21}'$ such that for each odd $i$ with $1 \leq i \leq 4k - 1$, we have that $i$ is of type $\rho_i$ with respect to $\sigma_\rho$. Therefore, $\sigma_\rho$ and $\sigma_\rho'$ are different when $\rho \neq \rho'$. Hence, $\Sigma_{21}'$ has at least as many members as there are special equality profiles. Clearly, there are $2^k$ distinct special equality profiles. So $\Sigma_{21}'$ has at least $2^k$ members. Since the size of the schema mapping $\Sigma_{12}$ is linear in $k$, it follows that the number of constraints in $\mathcal{M}_{21}'$ is exponential in the size of $\Sigma_{12}$. Since $\mathcal{M}_{21}$ is an arbitrary normal inverse of $\mathcal{M}_{12}$, this proves the theorem.

**Definition 9.2**: A constraint is *Boolean normal* if it is of the form $\alpha \land \chi_A \land \theta \rightarrow A$, where $\alpha$ is a conjunction of source atoms, $A$ is a target atom, $\chi_A$ is the conjunction of the formulas $\text{const}(x)$ for every variable $x$ of $A$, and $\theta$
is a Boolean combination (possibly empty) of equalities \( x = y \) for variables \( x, y \) of \( A \). Further, there is the safety condition that every variable in \( A \) must appear in \( \alpha \). Again, we have suppressed writing the leading universal quantifiers. A schema mapping is said to be Boolean normal if all of its constraints are Boolean normal. 

Thus, we obtain the definition of “Boolean normal” from the definition of “normal” by allowing Boolean combinations of equalities in the premise, rather than simply conjunctions of inequalities. Of course, every normal schema mapping is a Boolean normal schema mapping. Furthermore, it is easy to see that every Boolean normal schema mapping is equivalent to a normal schema mapping. That is, allowing Boolean combinations of equalities in the premise, rather than simply conjunctions of inequalities, does not increase the expressive power. However, allowing Boolean combinations of equalities in the premise does potentially allow a more compact representation. In particular, we can see that this happens in the inverse mappings in the proof of Theorem 9.1. There (except for the fact that we did not bother to include the const formulas) we produced examples of inverses, specified by Boolean normal constraints, that are are of polynomial size, and hence exponentially more compact than any inverse specified by normal constraints. The next theorem says that this is a general phenomenon: in the full case, we can always find, in polynomial time, a polynomial-size inverse (if an inverse exists). Before we prove this next theorem, we need some more machinery.

Let \( \sigma \) be an s-t tgd whose premise consists only of \( P \)-atoms for some single relational symbol \( P \). Define an equivalence relation \( E_\sigma \) on the variables that appear in \( \sigma \) as follows. Assume that \( P \) is \( t \)-ary. For each \( i \) with \( 1 \leq i \leq t \), let \( Y_i \) be the set of all variables that appear in the \( i \)th position of some atom in the premise of \( \sigma \). Let \( E_\sigma \) be the most refined equivalence relation (largest number of equivalence classes) such that each \( Y_i \) is a subset of an equivalence class of \( E_\sigma \). It is easy to see that each equivalence class of \( E_\sigma \) is a union of \( Y_i \)'s. For each equivalence class, select a unique representative, and let \( [x] \) denote the representative of the equivalence class containing \( x \). Form \( \sigma' \) from \( \sigma \) by replacing each variable \( x \) by \( [x] \). It is easy to see that each of the atoms in the premise of \( \sigma' \) is the same atom (call it \( A \)). Intuitively, we have done a minimum unification of the atoms in the premise of \( \sigma \), so that they all “unify” to \( A \). Form \( \sigma^\dagger \) from \( \sigma' \) by replacing the premise of \( \sigma' \) by \( A \). Now \( \sigma^\dagger \) is a “special case” of \( \sigma \) (thus, \( \sigma^\dagger \) is obtained from \( \sigma \) by identifying some variables and then replacing the conjunction \( A \land \cdots \land A \) by simply \( A \)), and so \( \sigma^\dagger \) is a logical consequence of \( \sigma \). As an example, let \( \sigma \) be \( P(x, y, z) \land P(y, w, z) \rightarrow Q(z, w) \). Then \( Y_1 = \{x, y\} \), \( Y_2 = \{y, w\} \), and \( Y_3 = \{z\} \). There are two equivalence classes of \( E_\sigma \), namely \( Y_1 \cup Y_2 = \{x, y, w\} \) and \( \{z\} \). If we take the representative of the equivalence class \( \{x, y, w\} \) to be \( x \), then \( \sigma' \) is \( P(x, x, z) \land P(x, x, z) \rightarrow Q(z, x) \), and \( \sigma^\dagger \) is \( P(x, x, z) \rightarrow Q(z, x) \). Note that if \( \sigma \) has a singleton premise, then \( \sigma^\dagger \) is simply \( \sigma \) itself.

The next lemma shows why we are interested in these formulas \( \sigma^\dagger \). We state this lemma in the full case (which is the only case where we shall apply it), although it holds also for the nonfull case, provided we carefully define what we mean when we say that two different chases (which might introduce different nulls) are equal.

**Lemma 9.3** Let \( \sigma \) be a full s-t tgd whose premise consists only of \( P \)-atoms for some fixed relational symbol \( P \), and let \( I \) be an instance consisting of a single \( P \)-fact. Then the chase of \( I \) with \( \sigma \) equals the chase of \( I \) with \( \sigma^\dagger \).

**Proof** Since \( \sigma \) logically implies \( \sigma^\dagger \), it follows easily that the chase of \( I \) with \( \sigma^\dagger \) is contained in the chase of \( I \) with \( \sigma \). We now show the opposite inclusion. Assume that \( \sigma \) fires on \( I \). Then there is a homomorphism \( h \) from the the premise of \( \sigma \) to \( I \). So if the \( Y_i \)'s are as in the definition of \( \sigma^\dagger \), then \( h(x) = h(y) \) whenever \( x \) and \( y \) are both in \( Y_i \). It follows easily that \( h(x) = h(y) \) whenever \( x \) and \( y \) are in the same equivalence class of \( E_\sigma \). Define \( h' \) on the variables \( [x] \) in the premise of \( \sigma^\dagger \) by letting \( h'([x]) = h(x) \). Since \( h(x) = h(y) \) whenever \( x \) and \( y \) are in the same equivalence class of \( E_\sigma \), it follows that \( h' \) is well-defined. It is straightforward to verify that \( h' \) is a homomorphism from the premise of \( \sigma^\dagger \) to \( I \), and so \( \sigma^\dagger \) fires on \( I \). Further, it is not hard to see that the homomorphic image of the conclusion of \( \sigma \) under \( h \) equals the homomorphic image of the conclusion of \( \sigma^\dagger \) under \( h' \). It follows that the chase of \( I \) with \( \sigma \) is contained in the chase of \( I \) with \( \sigma^\dagger \). This was to be shown. 

**Theorem 9.4** There is a polynomial-time algorithm such that if the input is a schema mapping \( M_{12} \) specified by a finite set of full s-t tgds, then the output is a polynomial-size Boolean normal schema mapping that is an inverse
of $M_{12}$ if $M_{12}$ has an inverse.\footnote{Note that by part (1) of Corollary 3.4, we cannot hope for a polynomial-time algorithm for deciding invertibility. Therefore, the output of the algorithm is left unspecified if $M_{12}$ has no inverse.}

**Proof** Let $M_{12} = (S_1, S_2, \Sigma_{12})$, where $\Sigma_{12}$ is a finite set of full s-t tgds. For each member $\varphi(x) \rightarrow (A_1 \land \ldots \land A_r)$ of $\Sigma_{12}$, where each $A_i$ is an atom, let $\Sigma_{12}$ contain the s-t tgds $\varphi(x) \rightarrow A_1, \ldots, \varphi(x) \rightarrow A_r$. Thus, $\Sigma_{12}$ is a finite set of full s-t tgds, each with a singleton conclusion, that is logically equivalent to $\Sigma_{12}$.

We now give a procedure to augment $\Sigma_{12}$ to a set $\Sigma''_{12}$. Let $U$ be the set of all of the s-t tgds $\sigma^i$, as defined earlier, and let $\Sigma''_{12} = \Sigma_{12} \cup U$. Since $\Sigma''_{12}$ consists of $\Sigma_{12}$ along with some logical consequences of $\Sigma_{12}$, it follows that $\Sigma''_{12}$ is logically equivalent to $\Sigma_{12}$. Since also $\Sigma''_{12}$ is logically equivalent to $\Sigma_{12}$, it follows that $\Sigma''_{12}$ is logically equivalent to $\Sigma_{12}$. By renaming variables if needed, we can assume that no two distinct members of $\Sigma''_{12}$ have a variable in common. Furthermore, we find it convenient to assume that for each member $\sigma$ of $\Sigma''_{12}$, there is another member $\sigma^o$ of $\Sigma''_{12}$ that is obtained from $\sigma$ by renaming the variables in a one-to-one manner and with a disjoint set of variables from $\sigma$ (we add $\sigma^o$ to $\Sigma''_{12}$ if needed). It is easy to see that there is a polynomial-time procedure for generating $\Sigma''_{12}$ from $\Sigma_{12}$.

Let us say that a member $\sigma$ of $\Sigma''_{12}$ is special if the premise contains a single atom, and if every variable in the premise appears in the conclusion (and hence the same variables appear in the premise and the conclusion). Let $\sigma$ be a special member of $\Sigma''_{12}$. Assume that $\sigma$ is $P(x_{z_1}, \ldots, x_{z_k}) \rightarrow Q(x_{i_1}, \ldots, x_{i_j})$. So the conclusion of $\sigma$ is a Q-atom. Let $\tau$ be an arbitrary member of $\Sigma''_{12}$, other than $\sigma$, such that the conclusion of $\tau$ is a Q-atom. Assume that the conclusion of $\tau$ is $Q(x_{j_1}, \ldots, x_{j_r})$. Recall that $\sigma$ and $\tau$ have no variables in common. Let $E^\tau$ be the most refined equivalence relation (largest number of equivalence classes) on the variables in $\sigma$ and $\tau$ such that $x_{i_1}$ and $x_{j_1}$ are in the same equivalence class, for $1 \leq i \leq k$. Let $\theta_1^\tau$ be a conjunction of equalities among the variables in $\sigma$, where the equality $x_{i_1} = x_{j_1}$ is an atom in $\theta_1^\tau$ precisely if $x_{i_1}$ and $x_{j_1}$ are in the same equivalence class of $E^\tau$. Intuitively, $\theta_1^\tau$ tells us how to equate variables to obtain a minimum unification of the conclusions of $\sigma$ and $\tau$. For each equivalence class $E$ of $E^\tau$, select a unique representative. If this equivalence class $E$ contains a variable in $\sigma$, then choose the representative of $E$ to be a variable in $\sigma$. (The only times that the equivalence class $E$ does not contain a variable in $\sigma$ is when $E$ consists of a variable in the premise of $\tau$ but not in the conclusion of $\tau$. This is possible, since we are not assuming that $\tau$ is special.) Let $[x]^\tau$ denote the representative of the equivalence class of $E^\tau$ containing $x$. Let us refer to the variables $[x_1]_1^\tau, \ldots, [x_k]_1^\tau$ as distinguished. Let us say that a $P$-atom $P(x_{u_1}, \ldots, x_{u_r})$ in the premise of $\sigma$ is distinguished if $[x_{u_1}]_1^\tau$ is distinguished for $1 \leq \ell \leq t$. If $A$ is the distinguished $P$-atom $P(x_{u_1}, \ldots, x_{u_r})$, define $\gamma_A$ to be the conjunction of the equalities $[x_{u_1}]_1^\tau = [x_{z_1}]_1^\tau$ for $1 \leq \ell \leq t$. Intuitively, $\gamma_A$ tells us how to equate variables to obtain a minimum unification of $A$ and the premise of $\sigma$. Define $\theta_2^\tau$ to be the disjunction of the formulas $\gamma_A$ for each distinguished $P$-atom $A$ of $\tau$. If this disjunction is empty (because $\tau$ has no distinguished $P$-atom), then $\theta_2^\tau$ is the empty disjunction, which is logically equivalent to False. Intuitively, $\theta_2^\tau$ tells us what possible collections of equalities among variables are needed to provide a minimum unification of some atom in the premise of $\tau$ with the premise of $\sigma$. Let $\theta_\tau$ be the formula $\theta_1^\tau \land \theta_2^\tau$. Intuitively, $\theta_\tau$ says that if the conclusions of $\tau$ and $\sigma$ can unify to the same atom, then some atom in the premise of $\tau$ can unify to the same atom as the premise of $\sigma$. Note that if $\tau$ has no distinguished $P$-atom, then $\theta_\tau$ is logically equivalent to $\theta_1^\tau$. Let $\theta$ be the conjunction of the formulas $\theta_\tau$, over all members $\tau$ of $\Sigma_{12}$ other than $\sigma$, where the conclusion of $\tau$ is a Q-atom. We now define $\sigma^*$ to be $Q(x_{i_1}, \ldots, x_{i_j}) \land \theta \rightarrow P(x_{z_1}, \ldots, x_{z_k})$. Note that (the hatted version of) the premise of $\sigma^*$ is the conclusion of $\sigma^o$, and the conclusion of $\sigma$ is the relational atom in the premise of $\sigma^o$. Let $\Sigma_{21}$ consist of all of the formulas $\sigma^o$, where $\sigma$ is a special member of $\Sigma_{12}$. Obtain $\Sigma'_{21}$ from $\Sigma_{21}$ by adding to the premise of every member $\tau$ of $\Sigma_{21}$ the conjuncts $\text{const}(x)$ where $x$ is a variable that appears in $\tau$. Let $M_{21}' = (S_2, \Sigma_1, \Sigma'_{21})$. The output of the algorithm is $M_{21}'$. It is clear that our algorithm runs in polynomial time (and so, of course, the output $M_{21}'$ is of polynomial size). Clearly $M_{21}'$ is a Boolean normal schema mapping. We shall show that $M_{21}'$ is an inverse of $M_{12}$ if $M_{12}$ has an inverse.

Assume that $\theta$ is a Boolean combination of equalities, and $f$ is a weak renaming of variables. Let us say that $\theta$ holds under $f$ if the Boolean expression that results by replacing each equality $x = y$ by $\text{True}$ when $f(x)$ and $f(y)$ are the same variable, and replacing each equality $x = y$ by $\text{False}$ when $f(x)$ and $f(y)$ are different variables, evaluates to $\text{True}$. Similarly, if $g$ is a function that maps variables to constants, then say that $\theta$ holds under $g$ if
the Boolean expression that results by replacing each equality $x = y$ by $\text{True}$ when $g(x)$ and $g(y)$ are the same constant, and replacing each equality $x = y$ by $\text{False}$ when $g(x)$ and $g(y)$ are different constants. Let us say that $f$ and $g$ agree on equalities if for each $x$, we have that $f(x) = g(x)$ if and only if $g(x) = g(y)$. Clearly, if $f$ and $g$ agree on equalities, then $\theta$ holds under $f$ if and only if $\theta$ holds under $g$. As before, if $\varphi$ is a formula, let $\varphi^f$ be the result of replacing every variable $x$ in $\varphi$ by $f(x)$. If $A$ is an atom, let $A^\varphi$ be the fact that arises by replacing every variable $x$ in $A$ by $g(x)$.

**Claim:** For every constraint $\sigma^*$ in $\Sigma_{21}$, which must be of the form $\beta \land \theta \rightarrow \alpha$, where (1) $\alpha$ is a source atom, (2) $\beta$ is a target atom with the same variables as $\alpha$, and (3) $\theta$ is a Boolean combination of equalities among the variables, and for every weak renaming $f$, we have that $\theta$ holds under $f$ if and only if $\beta^f$ is an essential atom for $\alpha^f$ (with respect to $\Sigma_{12}$).

Note that $\sigma^*$ is derived from $\sigma$ in $\Sigma'_{12}$, where $\sigma$ is $\alpha \rightarrow \beta$. Assume that $\alpha$ is a $P$-atom and $\beta$ is a $Q$-atom. We now prove the Claim. Assume first that $\beta^f$ is essential for $\alpha^f$; we wish to show that $\theta$ holds under $f$. To show this, we must show that if $\tau$ is a member of $\Sigma''_{12}$ other than $\sigma$, and the conclusion of $\tau$ is a $Q$-atom, then $\theta$ holds under $f$ (this is because $\theta$ is the conjunction of such formulas $\theta_\tau$). So assume that $\theta_1^\tau$ holds under $f$; we must show that $\theta_2^\tau$ holds under $f$. Now the conclusion of $\sigma^f$ is $\beta^f$. Since $\theta_1^\tau$ holds under $f$, it follows that $\sigma^f$ and $\tau^f$ have the same conclusion. So the conclusion of $\tau^f$ is $\beta^f$. Let $g$ be a function that maps variables into constants and that agrees with $f$ on equalities. So the conclusion of of $\sigma^g$ is $\beta^g$. Let $I$ be an instance whose facts are the facts $A^\varphi$ for each atom $A$ in the premise of $\tau$. So the chase of $I$ with $\tau$ includes $\beta^g$. Since $\beta^f$ is essential for $\alpha^f$, it follows that $\alpha^g$ is a fact in $I$. So $\alpha^g$ is $A^\varphi$ for some atom $A$ in the premise of $\tau$. It follows that $g_A$, as defined earlier, holds under $g$, and so $\theta_2^\tau$ holds under $g$. Since $f$ and $g$ agree on equalities, this implies that $\theta_2^\tau$ holds under $f$, as desired.

Assume now that $\theta$ holds under $f$; we must show that $\beta^f$ is an essential atom for $\alpha^f$. Certainly $\beta^f$ is relevant for $\alpha^f$, since $\sigma$ is in $\Sigma_{12}$. So we must show that $\beta^f$ is demanding for $\alpha^f$. Let $g$ be a function that maps variables into constants and that agrees with $f$ on equalities. So $\theta$ holds under $g$. Let $I$ be an instance where $\text{chase}_{12}(I)$ contains $\beta^g$; we need only show that $\alpha^g$ is a fact in $I$. It is easy to see that the result of chasing with $\Sigma_{12}$ and $\Sigma''_{12}$ are the same. So the result of chasing $I$ with $\Sigma''_{12}$ contains $\beta^g$. Hence, there is a constraint $\tau$ in $\Sigma''_{12}$ that fires on $I$ and produces $\beta^g$. If $\tau = \sigma$, then it is not hard to see that this implies that $\alpha^g$ is in $I$, as desired. If $\tau$ is not $\sigma$, it is straightforward to verify that $\theta_1^\tau$ holds under $g$. Since also $\theta$ holds under $g$, this implies that $\theta_2^\tau$ holds under $g$. So there is some distinguished atom $A$ in the premise of $\tau$ such that $g_A$ holds under $g$. Hence, $A^\varphi$ and $\alpha^g$ are the same fact. Let us denote the conclusion of $\tau$ by $C$. Since $\theta_1^\tau$ holds under $g$, we have that $\beta^g = C^g$. So $C^g$ arises by chasing $\tau$ with $\tau$. Since also $\tau$ has $A$ in its premise, we have that $A^\varphi$ is in $I$. But we showed that $A^\varphi$ and $\alpha^g$ are the same fact. So $\alpha^g$ is in $I$, as desired. This concludes the proof of the Claim.

Assume that $M_{12}$ has an inverse. We now use the Claim to prove that $M_{21}'$ is an inverse of $M_{12}$.

Assume that $\sigma^*$ is a member of $\Sigma_{21}$, and $\sigma^*$ is $\beta \land \theta \rightarrow \alpha$. Let $k$ be the number of variables that appear in $\sigma^*$. Define the set $T_{\sigma^*}$ as follows. For each weak renaming $f$ of the variables in $\sigma^*$ such that the range of $f$ is in $\{x_1, \ldots, x_k\}$ and such that $\theta$ holds for $f$, let $T_{\sigma^*}$ contain the constraints $\beta^f \land \eta_y \rightarrow \alpha^f$, where $\eta_y$ is the conjunction of the inequalities $f(x) \neq f(y)$ where $x$ and $y$ are variables of $\sigma^*$ and where $f(x)$ and $f(y)$ are different variables. (The assumption that range of $f$ is in $\{x_1, \ldots, x_k\}$ is only to assure that $T_{\sigma^*}$ be finite.) It is straightforward to see that $\sigma^*$ is logically equivalent to $T_{\sigma^*}$, and similarly the constants-added version of $\sigma^*$ is logically equivalent to the constants-added version of $T_{\sigma^*}$ (recall that the constants-added version of an s-t tgd mapping is the mapping that results by adding to the premise of every tgd the formulas $\text{const}(x)$ for every variable $x$ that appears in the conclusion). Let $\Sigma_{21}'$ be the union of the constants-added version of the sets $T_{\sigma^*}$ over all $\sigma^*$ in $\Sigma_{21}$, and let $M''_{21} = (S_2, S_1, \Sigma_{21}')$. By Proposition 7.11, we know that $M''_{21}$ is an inverse of $M_{12}$ if and only if $M''_{21}$ is an inverse of $M_{12}$. Hence we need only show that $M''_{21}$ is an inverse of $M_{12}$. We now use Theorem 5.13 to show that $M''_{21}$ is an inverse of $M_{12}$. Since each $T_{\sigma^*}$ was obtained by considering weak renamings $f$ such that $\theta$ holds for $f$, it follows easily from the Claim that for every member $\varphi$ of $\Sigma_{21}'$, the premise of $\varphi$ is essential for the conclusion of $\varphi$. Hence, the first condition of Theorem 5.13 holds (when $\Sigma_{21}'$ plays the role of $\Sigma_{21}$). We now show that the second condition also holds. Let $A$ be a source atom. Since $M_{12}$ is invertible, we know by Proposition 6.2 that $\omega_A$ is essential for $A$, and so contains an atom $B$ that is essential for $A$ (with respect to $\Sigma_{12}$).
It follows from Lemma 9.3 that there is a member $\sigma$ of $\Sigma_{12}$ with a singleton premise (and a singleton conclusion) such that the chase of $I_A$ is the same with $\sigma$ as it is with $\Sigma_{12}$. Write $\sigma$ as $\alpha \rightarrow \beta$. So there is a weak renaming $f$ such that $\alpha f$ is $A$ and $\beta f$ is $B$. We now show that every variable in $\alpha$ appears in $\beta$, and so $\sigma$ is special. Assume that some variable $x$ appears in $\alpha$ but not in $\beta$; we shall derive a contradiction. Let $f'$ be a weak renaming that is like $f$ except that $f'(x)$ is a new variable. So $\alpha f'$ is different from $A$, although $\beta f'$ is the same as $\beta f$, that is, $B$. So chase$_{12}(I, \mu')$ contains $I_B$, even though $I_A \not\subseteq I, \mu'$. This contradicts the fact that $B$ is essential for $A$. Hence, $\sigma$ is special, as desired.

So there is $\theta$ such that $\sigma\theta^* = \beta \theta^* \rightarrow \tilde{A}$, and $\sigma^*$ is in $\Sigma_{21}$. Since $\beta f'$ (namely, $B$) is essential for $\alpha f$ (namely, $A$), it follows from the Claim that $\theta$ holds under $f$. Therefore, if $\chi_A$ is the conjunction of the formulas $\text{const}(x)$ for every variable $x$ of $A$, and if $\eta_f$ is as above, then $B \wedge \chi_A \wedge \eta_f \rightarrow \tilde{A}$ is a weak renaming of a constraint in $\Sigma_{21}$. Hence, the second condition of Theorem 5.13 holds (when $\Sigma_{21}'$ plays the role of $\Sigma_{21}$), as desired. This completes the proof that $M_{21}'$ is an inverse of $M_{12}$. ☐

It is open as to whether such a polynomial-time algorithm exists in the nonfull case. It is even open in the nonfull case as to whether or not there always exists a Boolean normal inverse of polynomial size if an inverse exists.

### 10 Relating Number of Inverses to Number of Constraints in an Inverse

In this section, we show that for each full s-t tgd mapping $M_{12}$, there is a relationship between the number of normal inverses of $M_{12}$ and the minimal number of constraints in a Boolean inverse for $M_{12}$. We first show that we cannot bound the number of inverses in terms of the minimal number of constraints in a Boolean normal inverse, since there is a full s-t tgd mapping with infinitely many distinct normal inverses. We then show that we can bound the minimal number of constraints in a Boolean normal inverse in terms of the number of inverses (and the number of relation symbols).

We begin by giving an example of a full s-t tgd mapping with infinitely many distinct normal inverses.

**Example 10.1** Let $S_1$ consist of the unary relation symbol $P$, and let $S_2$ consist of the binary relation symbol $Q$. Let $\Sigma_{12}$ consist of the s-t tgd $P(x) \rightarrow Q(x, x)$. Let $\Sigma_{21}$ consist of the normal constraint $Q(x, y_1) \wedge Q(y_1, y_2) \wedge \ldots \wedge Q(y_{k-1}, y_k) \wedge Q(y_k, x) \wedge \text{const}(x) \rightarrow P(x)$.

Let $M_{12} = (S_1, S_2, \Sigma_{12})$, and let $M_{21}^k = (S_2, \widehat{S_1}, \widehat{\Sigma_{21}})$. It is straightforward to verify that for every ground instance $I$ and for each $k \geq 1$ we have chase$_{21}^k(\text{chase}_{12}(I)) = \widehat{I}$ (where chase$_{21}^k$ is the result of chasing $J$ with $\Sigma_{21}^k$). It therefore follows from Theorem 4.10 that $M_{21}^k$ is an inverse of $M_{12}$ for every $k$. It is also straightforward to verify that $\Sigma_{21}^k$ and $\Sigma_{21}^k$ are not logically equivalent if $k \neq k'$. So $M_{12}$ has infinitely many inequivalent normal inverses. ☐

The next theorem says that we can bound the minimal number of constraints in a Boolean normal inverse in terms of the number of inverses (and the number of relation symbols).

**Theorem 10.2** Let $M_{12}$ be a full s-t tgd mapping, with $k$ source relation symbols. Assume that $M_{12}$ has exactly $m \geq 1$ inequivalent normal inverses. Then $M_{12}$ has a Boolean normal inverse with at most $k + \log_2(m)$ constraints.

**Proof** Assume that $M_{12} = (S_1, S_2, \Sigma_{12})$. Let us say that the source atom $B$ is good if chase$_{12}(I_B)$ has exactly one member. Let us say that $B$ is bad if $B$ is not good. Let $b$ be the number of bad prime source atoms. We shall show that $2^b \leq m$ (where $m$ is the number of inequivalent normal inverses of $M$) and that $M$ has a Boolean normal inverse with $k + b$ constraints. Since $2^b \leq m$, we have that $b \leq \log_2(m)$, and so $k + b \leq k + \log_2(m)$. The theorem follows.
For each source relation symbol $P$, let $A_P$ be the $P$-atom $P(x_1, \ldots, x_r)$ where $x_1, \ldots, x_r$ are distinct. If $B$ is a $P$-atom $P(y_1, \ldots, y_r)$, let $\varphi_B$ be the formula that is the conjunction of (1) the equalities $x_i = y_j$ for each $i, j$ where $y_i$ and $y_j$ are the same variable, and (2) the inequalities of the form $x_i \neq x_j$ for each $i, j$ where $y_i$ and $y_j$ are different variables. Intuitively, $\varphi_B$ completely describes the equality pattern of the variables in $B$. Let $\theta_B$ be the disjunction of the formulas $\varphi_B$ where $B$ is a good $P$-atom. Let $\sigma_B$ be the formula $\omega_A \cap \theta_B \rightarrow \hat{A}_P$, where $\omega_A$ is defined as in Definition 6.1.

Let $B_1, \ldots, B_b$ be precisely the bad prime source atoms (they may involve various relation symbols). By Proposition 6.2, we know that $\omega_B$ is essential for $B_i$ with respect to $\Sigma_{12}$, for each $i$. Since $B_i$ is bad, it follows that $\omega_{B_i}$ has more than one atom. By the construction in Definition 6.1, we see that $\omega_{B_i}$ contains the formulas $\text{const}(x)$ for every variable $x$ of $B_i$. So by Proposition 5.18, we know that some atom $C_i$ in $\omega_{B_i}$ is essential for $B_i$ with respect to $\Sigma_{12}$, for each $i$. If $B_i$ is $P(y_1, \ldots, y_r)$, define $\eta_i$ to be the conjunction of the inequalities of the form $y_i \neq y_j$ where $y_i$ and $y_j$ are distinct variables. Let $\psi_i^0$ be the constraint $C_i \land \eta_i \rightarrow B_i$, and let $\psi_i^1$ be the constraint $\omega_{B_i} \land \eta_i \rightarrow B_i$.

Let $v = (v_1, \ldots, v_b)$ be an arbitrary $\{0, 1\}$-vector of length $b$. Define $\Sigma_{21}^Y$ to consist of the $k$ formulas $\sigma_B$ (one for each source relation symbol $P$) along with the $b$ constraints $\psi_i^0$ for $1 \leq i \leq b$. Let $M_{21}^Y = (S_2, S_1, \Sigma_{21}^Y)$. We now show that each $M_{21}^Y$ is an inverse of $M_{12}$, and that $M_{21}^Y$ and $M_{21}^Y'$ are not equivalent if $v \neq v'$. Since the number of vectors $v$ is $2^b$, this shows that $2^b \leq m$. Further, since each $M_{21}^Y$ is a Boolean normal inverse with $k + b$ constraints, this shows that $M$ has a Boolean normal inverse with $k + b$ constraints (in fact, it has at least $2^b$ Boolean normal inverse with $k + b$ constraints). This is sufficient to complete the proof.

Fix $v = (v_1, \ldots, v_b)$. We begin by showing that $M_{21}^Y$ is an inverse of $M_{12}$. We now define a function $e$ that maps each prime source atom $B$ to an essential conjunction $e(B)$ with respect to $\Sigma_{12}$. For the bad prime source atom $B_i$, we let $e(B_i) = C_i$ if $v_i = 0$, and $e(B_i) = \omega_{B_i}$ if $v_i = 1$. By construction, $e(B_i)$ is essential for $B_i$ if $v_i = 0$, and by Proposition 6.2, we know that $e(B_i)$ is essential for $B_i$ if $v_i = 1$. For each good prime source atom $A$, we let $e(B) = \omega_B$. Again by Proposition 6.2, we know that $e(B)$ is then essential for $B$. So by Theorem 5.15, $M_{21}^Y$ is an inverse of $M$. We now show that $M_{21}^Y$ is equivalent to $M_{21}^Y$, which completes the proof that $M_{21}^Y$ is an inverse of $M_{12}$.

For each prime source atom $B$ where $B$ is bad, $M_{21}^Y$ and $M_{21}^Y$ contain the same constraint with conclusion $B$. Let us now consider the good prime source atoms $B$. The formula $\sigma_B$ is logically equivalent to the set consisting of all of the formulas $\omega_A \cap \varphi_B \rightarrow \hat{A}_P$, where $B$ is a good prime $P$-atom. Assume that $B$ is a good prime source atom. Let $\sigma_1$ be the formula $\omega_A \cap \varphi_B \rightarrow \hat{A}_P$, and let $\sigma_2$ be the formula $\omega_B \land \eta_B \rightarrow \hat{B}$, where as before $\eta_B$ is the conjunction of all inequalities of the form $x \neq y$ where $x$ and $y$ are distinct variables in $B$. By construction, $\sigma_2$ is the unique member of $M_{21}^Y$ with conclusion $B$. So to complete the proof that $M_{21}^Y$ is equivalent to $M_{21}^Y$, we need only show that the formula $\sigma_1$ is logically equivalent to the formula $\sigma_2$.

Assume that $B$ is the good atom $P(y_1, \ldots, y_r)$, where $y_1, \ldots, y_r$ are variables, not necessarily distinct. Let $\psi_B$ be the formula obtained from $\sigma_1$ by replacing the variable $x_i$ by $y_i$, for $1 \leq i \leq r$. We now show that $\psi_B$ is logically equivalent to both $\sigma_1$ and $\varphi_B$, which implies that $\sigma_1$ and $\sigma_2$ are logically equivalent, as desired. In forming $\psi_B$, two variables $x_i$ and $x_j$ in $\sigma_1$ are replaced by the same variable precisely if $y_i$ and $y_j$ are the same variable, which holds precisely if the equality $x_i = x_j$ appears in $\varphi_B$. It follows easily that $\psi_B$ is logically equivalent to $\sigma_1$. We now show that $\psi_B$ is logically equivalent to $\sigma_2$.

It is easy to see that the conclusions of $\psi_B$ and $\sigma_2$ are the same, and that the result of replacing the variable $x_i$ by $y_i$, for $1 \leq i \leq r$, in $\varphi_B$ is equivalent to $\eta_B$. Let $\tau_B$ be the result of replacing the variable $x_i$ by $y_i$ in $\omega_A$, for $1 \leq i \leq r$. So we need only show that $\tau_B$ is equivalent to $\omega_B$. Now the conjunction(s) of $\tau_B$ must be in $\omega_B$, by properties of the chase with s-t-gds. Since $\omega_B$ is a singleton (because $B$ is good), it follows easily that $\tau_B$ is the same as $\omega_B$. This concludes the proof that $M_{21}^Y$ is equivalent to $M_{21}^Y$. We conclude the proof by showing that $M_{21}^Y$ and $M_{21}^Y$ are not equivalent if $v \neq v'$. Say $v \neq v'$, and that $v = (v_1, \ldots, v_b)$ and $v' = (v'_1, \ldots, v'_b)$. So there is $i$ with $1 \leq i \leq b$ such that $v_i \neq v'_i$. Assume without loss of generality that $v_i = 0$ and $v'_i = 1$. We now show that $(I_{C_i}, \emptyset)$ satisfies $\Sigma_{21}^Y$ but not $\Sigma_{21}^Y$. This of course shows that $M_{21}^Y$ and $M_{21}^Y$ are not equivalent. Clearly $\psi_i^0$ fires on $I_{C_i}$, and so $(I_{C_i}, \emptyset)$ does not satisfy $\psi_i^0$. Hence, $(I_{C_i}, \emptyset)$ does not satisfy $\Sigma_{21}^Y$, because $\Sigma_{21}^Y$ contains $\psi_i^0$. We now show that no member of $\Sigma_{21}^Y$ fires on $I_{C_i}$. Since $B_i$ is
bad, we know that $\omega_{B_i}$ has some other atom $A$ in addition to $C_i$ as a conjunct. Since $C_i$ is essential for $B_i$, it follows from Proposition 5.11 that $B_i$ and $C_i$ have the same variables. Since $\Sigma_{12}$ is full, every variable in $A$ is in $B_i$, and hence in $C_i$. Assume that $C_i$ is $Q(c_1, \ldots , c_m)$. Then $I_{C_i}$ consists of the fact $Q(c_1, \ldots , c_m)$. If $\psi^i_1$ were to fire on $I_{C_i}$, then there would be a homomorphism $h$ from the premise of $\psi^i_1$ to $I_{C_i}$. Since $C_i$ is part of the premise of $\psi^i_1$, we must have $h(y_i) = c_y$, for $1 \leq i \leq m$. Since $h$ must map $A$ onto $Q(c_1, \ldots , c_m)$, and since every variable in $A$ is among $y_1, \ldots , y_m$, it is easy to see that $A$ must be $C_i$, which is a contradiction. So $\psi^i_1$ does not fire on $I_{C_i}$. We now show that no other member of $\Sigma_{21}'$ fires on $I_{C_i}$. If some member $\psi^i_j'$ were to fire on $I_{C_i}$, where $j \neq i$, then because of the inequalities in $\psi^i_j'$, it would follow that some (and in fact, every) member of $\text{chase}_{12}(I_{B_j})$ is of the form $Q(c_1, \ldots , c_m)$, where $c_k = c_{k'}$ if and only if $y_k = y_{k'}$ (intuitively, each member of the prefix of $\psi^i_j'$ is a $Q$-atom with the same equality pattern of variables as $C_i$). So there would be a homomorphism $I_{C_i} \rightarrow \text{chase}_{12}(I_{B_j})$. Since $C_i$ is demanding for $B_j$, it follows that $I_{B_i} \subseteq I_{B_j}$. But this is impossible, since $i \neq j$. So no member $\psi^i_j'$ fires on $I_{C_i}$, where $j \neq i$. A similar argument shows that no $\sigma_P$ fires on $I_{C_i}$. So no member of $\Sigma_{21}'$ fires on $I_{C_i}$, and hence $(I_{C_i}, \emptyset)$ satisfies $\Sigma_{21}'$, as desired. □

It is open problem as to whether a version of Theorem 10.2 holds in the nonfull case.

Note in particular from Theorem 10.2 that if the $s$-$t$ tgd mapping $\mathcal{M}_{12}$ has a unique normal inverse (so that $m = 1$ in Theorem 10.2) then $\mathcal{M}_{12}$ has a Boolean normal inverse with at most $k$ constraints, where $k$ is the number of source relation symbols. This is the key to proving Theorem 7.15. We now give that proof.

**Proof of Theorem 7.15:** Assume that $\mathcal{M}_{12} = (\mathcal{S}_1, \mathcal{S}_2, \Sigma_{12})$ is a full $s$-$t$ tgd mapping with a unique normal inverse $\mathcal{M}_{21} = (\mathcal{S}_2, \mathcal{S}_1, \Sigma_{21})$, and $\mathcal{M}_{21}$ is not equivalent to the constants-added version of a near $p$-copy mapping. Assume without loss of generality that $\Sigma_{21}$ has the minimal number of constraints among the various sets of normal constraints logically equivalent to $\Sigma_{21}$, and each constraint $\sigma$ in $\Sigma_{21}$ has the minimal size among normal constraints logically equivalent to $\sigma$. Since $\mathcal{M}_{12}$ has a unique normal inverse, it follows from Theorem 10.2 (where $m = 1$) that $\mathcal{M}_{21}$ has at most $k$ constraints, where $k$ is the number of source relation symbols.

Let $P$ be an arbitrary source relation symbol, and let $A_P$ be a $P$-atom with all variables distinct. If $\tilde{P}$ is in the conclusion of no member of $\Sigma_{21}$, then $\text{chase}_{21}(\text{chase}_{12}(I_{A_P}))$ contains no $\tilde{P}$-fact, so $\Sigma_{21}$ is too weak. Therefore, $\tilde{P}$ is in the conclusion of some member of $\Sigma_{21}$. Since also $\Sigma_{21}$ has at most $k$ constraints, it follows that $\tilde{P}$ is in the conclusion of exactly one member of $\Sigma_{21}$. Let $\sigma_P$ be the member of $\Sigma_{21}$ whose conclusion is a $\tilde{P}$-atom. Every variable in the conclusion of $\sigma_P$ is distinct, or else $\text{chase}_{21}(\text{chase}_{12}(I_{A_P}))$ does not contain $I_{A_P}$, so $\Sigma_{21}$ is too weak. Let $B_P$ be a $P$-atom with all variables the same. Now $\sigma_P$ has no inequalities, or else $\text{chase}_{21}(\text{chase}_{12}(I_{B_P}))$ does not contain $I_{B_P}$, so $\Sigma_{21}$ is too weak.

The proof of Theorem 10.2 shows that $A_P$ is good, that is, that $\text{chase}_{12}(I_{A_P})$ is a singleton. This singleton is the only relevant atom (with respect to $\Sigma_{12}$) for $A_P$, so it follows fairly easily from part (2) of Theorem 5.13 (and the assumption that constraints in $\Sigma_{21}$ are of minimal size) that the premise of $\sigma_P$ contains (along with $\text{const}$ formulas) a single relational atom, that is essential for the conclusion of $\sigma_P$. Note that this is also true about each weak renaming of $\sigma_P$ (that is, the atom $B'$ in the premise of the weak renaming is essential for the conclusion $A'$ of the weak renaming). This is because $A'$ must have an essential atom, and the only candidate is $B'$.

Let $Q$ be an arbitrary relation symbol in $\mathcal{S}_2$. We now show that at most one member of $\Sigma_{21}$ can have $Q$ appear in its premise. Assume that $\sigma_P$ and $\sigma_{P'}$ both have $Q$ appear in its premise; we must show that $P$ and $P'$ are the same. Let $B_Q$ be a $Q$-atom with all variables the same. Then $B_Q$ is equivalent for both $B_P$ and $B_{P'}$ (this follows from our earlier comment about weak renamings of $\sigma_P$). Hence, by Lemma 5.16, it follows that $B_P$ and $B_{P'}$ are the same atom, so $P$ and $P'$ are the same, as desired. By Proposition 5.20, the variables in the premise and conclusion of $\sigma_P$ are the same.

It follows from what we have shown that $\mathcal{M}_{21}$ is equivalent to the constants-added version of a near $p$-copy mapping. This was to be shown. □
11 Invertibility in the LAV Case

Recall that a schema mapping has the unique-solutions property if no two distinct source instances have the same set of solutions. Fagin [Fag07] showed that the unique-solutions property is a necessary condition for a schema mapping to have an inverse. Fagin [Fag07] also showed that for LAV mappings (those specified by s-t tgds with a singleton premise), the unique-solutions property is not only a necessary condition but also a sufficient condition for invertibility. The proof of this latter result was quite complicated. In this section, we give a very simple proof.

Just as we defined a homomorphic version of the subset property in Section 3, there is a homomorphic version of the unique-solutions property, namely, that \( I = I' \) whenever \( \text{chase}_{12}(I) \iff \text{chase}_{12}(I') \). The reason that the unique-solutions property is equivalent to its homomorphic version is because of the fact, shown in [FKMP05], that two source instances \( I \) and \( I' \) have the same solutions if and only if they have homomorphic universal solutions. Note that it follows immediately from the two homomorphic versions that the subset property implies the unique-solutions property.

We now give our greatly simplified proof that the unique solutions property characterizes invertibility in the LAV case.

**Theorem 11.1** [Fag07] A LAV s-t tgd mapping is invertible if and only if it has the unique-solutions property.

**Proof** We just noted that the subset property implies the unique-solutions property. Since satisfying the subset property is equivalent to invertibility, the “only if” direction follows (even when the s-t tgd mapping is not LAV).

Assume now that \( M_{12} = (S_1, S_2, \Sigma_{12}) \) is a LAV mapping that satisfies the unique-solutions property. We now show that \( M_{12} \) satisfies the homomorphic version of the subset property, and so is invertible. Assume that \( I \) and \( I' \) are such that \( \text{chase}_{12}(I) \rightarrow \text{chase}_{12}(I') \). Then

\[
\text{chase}_{12}(I \cup I') = \text{chase}_{12}(I) \cup \text{chase}_{12}(I') \leftrightarrow \text{chase}_{12}(I'),
\]

where the equation follows from the fact that \( M_{12} \) is LAV, and the homomorphism \( \text{chase}_{12}(I) \cup \text{chase}_{12}(I') \rightarrow \text{chase}_{12}(I') \) follows from the fact that we can select the variables generated in the chase so that \( \text{chase}_{12}(I) \) and \( \text{chase}_{12}(I') \) have no nulls in common. Then by the homomorphic version of the unique-solutions property, \( I \cup I' = I' \) and therefore \( I \subseteq I' \). This shows that \( M_{12} \) satisfies the homomorphic version of the subset property, and so is invertible, as desired. \( \square \)

12 Concluding Remarks and Open Problems

In addition to resolving the key problem left open in [Fag07] as to the complexity of deciding if an s-t tgd mapping has an inverse, and also providing greatly simplified proofs of some known results, we have explored a number of interesting issues, about the structure of inverses, unique inverses, number of inverses, inverses of inverses, and sizes of inverses. We have shown that in the full case, these issues are, surprisingly, quite interrelated. We have also shown that in the nonfull case, these tight interconnections do not hold. We showed that in the full case, there is a polynomial-size Boolean inverse, and a polynomial-time algorithm for producing it. As we noted in Sections 9 and 10, there remain open problems about the size and about the number of constraints in inverses in the nonfull case. Perhaps the most interesting open problem is whether every invertible s-t tgd mapping (not necessarily full) has a polynomial-size Boolean inverse, and if so, whether there is a polynomial-time algorithm for producing it.

References


