The Structure of Inverses in Schema Mappings

RONALD FAGIN AND ALAN NASH

IBM Almaden Research Center

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 nonequivalent 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 $\mathcal{M}$ specified by *full* s-t tgds (those with no existential quantifiers), if $\mathcal{M}$ 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.2.4 [**Database Management**]: Systems—*Relational databases*; H.2.5 [**Database Management**]: Heterogeneous Databases—*Data translation*

General Terms: Algorithms, Theory

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

ACM Reference Format:

Fagin, R., and Nash, A. 2010. The structure of inverses in schema mappings. J. ACM 57, 6, Article 31, (October 2010), 57 pages.
DOI = 10.1145/1857914.1857915 http://doi.acm.org/10.1145/1857914.1857915

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

Authors' address: IBM Almaden Research Center, Dept. K53/B2, 650 Harry Road, San Jose, CA 95120; R. Fagin (corresponding author), e-mail: fagin@almaden.ibm.com.

DOI 10.1145/1857914.1857915 http://doi.acm.org/10.1145/1857914.1857915

$\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$, where $\mathbf{S}$ (the *source schema*) and $\mathbf{T}$ (the *target schema*) are sequences of distinct relation symbols with no relation symbols in common, and Σ 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 $\mathbf{S}$), and J is a target instance (an instance of the schema $\mathbf{T}$), then we say that J is a *solution* for I if the pair (I, J) together satisfies Σ. We sometimes identify the schema mapping $\mathcal{M}$ with the set of pairs (I, J) that satisfy Σ, and write $(I, J) \in \mathcal{M}$. The most widely studied case arises when Σ 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 [Lenzerini 2002]. 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 [Kolaitis 2005; Lenzerini 2002]), several different operators on schema mappings have been singled out as deserving study in their own right [Bernstein 2003]. 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 [Bernstein and Melnik 2007].

The composition operator has been investigated in depth [Fagin et al. 2005b; Madhavan and Halevy 2003; Melnik 2004; Nash et al. 2007]; 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 [2007] 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 $\mathcal{M}'$ (i.e., 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 et al. [2008] 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 et al. [2009] 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 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 et al. 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 et al. [2009] 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., nonground source instances) to consider. Fagin et al. 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 et al. [2009], 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" (i.e., Fagin's notion of inverse, Fagin et al.'s notions of quasi-inverse and of maximum extended recovery, and Arenas et al.'s 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 [2007] 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 et al. [2008] 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 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 nonground source instances. This is why we consider only ground source instances, in the tradition of the earlier articles [Arenas et al. 2009; Fagin 2007; Fagin et al. 2008]. Normal inverses are especially nice, in that if I is a source instance, $\mathcal{M}$ is an s-t tgd mapping specified by Σ, and $\mathcal{M}'$ is a normal inverse of $\mathcal{M}$ that is specified by Σ', then the result of chasing I with Σ and then chasing the result by Σ' 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 tgds have no existential quantifiers), there is an especially interesting story. Let us say that a full s-t tgd mapping $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ is *onto* if every target instance is the result of chasing some source instance with Σ. 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 Fagin et al. [2008], we define *disjunctive tgds 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 Fagin et al. [2008]). We show that a full s-t tgd mapping $\mathcal{M}$ has a unique inverse specified by disjunctive tgds with inequalities if and only if $\mathcal{M}$ is invertible and onto. Furthermore, we show that $\mathcal{M}$

satisfies these conditions if and only if $\mathcal{M}$ 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 $\mathcal{M}'$ is an inverse of $\mathcal{M}$, then $\mathcal{M}$ is an inverse of $\mathcal{M}'$. This is because M' is a right-inverse of $\mathcal{M}$ (that is, $\mathcal{M} \circ \mathcal{M}'$ is the identity) if and only if M is a left-inverse of $\mathcal{M}'$, 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 [2007], is only a right-inverse. In particular, it does *not* follow that if $\mathcal{M}'$ is an inverse of a schema mapping $\mathcal{M}$, then $\mathcal{M}$ is an inverse of $\mathcal{M}'$. In fact, it does not even follow that if $\mathcal{M}'$ is an inverse of a schema mapping, then $\mathcal{M}'$ 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 $\mathcal{M}$ is a full s-t tgd mapping with an invertible normal inverse if and only if $\mathcal{M}$ 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: $\mathcal{M}$ has an invertible normal inverse if and only if $\mathcal{M}$ has a unique inverse specified by disjunctive tgds with inequalities if and only if $\mathcal{M}$ is invertible and onto if and only if $\mathcal{M}$ is equivalent to a p-copy mapping. We also show that these nice properties are not equivalent if we remove the restriction that $\mathcal{M}$ 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 $\mathcal{M}$ such that the size of the smallest normal inverse of $\mathcal{M}$ is exponential in the size of $\mathcal{M}$. 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 $\mathcal{M}$ is a full s-t tgd mapping, with k source relation symbols and with exactly $m \geq 1$ inequivalent normal inverses, then $\mathcal{M}$ 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 [Fagin et al. 2008] that for invertible s-t tgd mappings $\mathcal{M}$, the canonical candidate inverse of $\mathcal{M}$ is indeed an inverse of $\mathcal{M}$. We now discuss the second result where we give a greatly simplified proof. Fagin [2007] introduced the

[1] We do not define the property of being *onto* when $\mathcal{M}$ 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.

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 tgds 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.

2. *Preliminaries*

In this section, we give a number of definitions. Most of these definitions are from Fagin et al. [2005a], which we refer the reader to for an introduction to data exchange.

Schemas and Schema Mappings. A *schema* $\mathbf{R}$ is a finite sequence $(R_1, \ldots, R_k)$ of relation symbols, each of a fixed arity. An *instance* I over $\mathbf{R}$ (which we may call an $\mathbf{R}$-*instance*, or simply an *instance*, when $\mathbf{R}$ 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 $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ consisting of a source schema $\mathbf{S}$, a target schema $\mathbf{T}$, and a set Σ of constraints (defined shortly). We say that $\mathcal{M}$ is *specified by* Σ. If Σ is a finite set of s-t tgds (defined shortly), then we may refer to $\mathcal{M}$ as an *s-t tgd mapping*. When $\mathbf{S}$ and $\mathbf{T}$ are clear from context, we will sometimes say Σ when we should say $(\mathbf{S}, \mathbf{T}, \Sigma)$, and talk about a set of constraints, when we should talk about a schema mapping.

Instances and Formulas. From now on, we assume that $\mathbf{S}$ and $\mathbf{T}$ are two fixed schemas. We call $\mathbf{S}$ the *source schema* and $\mathbf{T}$ the *target schema*. We refer to $\mathbf{S}$-instances as *source instances*, and $\mathbf{T}$-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 (i.e., 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 $\mathbf{S}$-instances as *ground instances* to emphasize the fact that all individual values in such instances are constants. Intuitively, schema mappings of the form $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ model the situation in which we perform data exchange from $\mathbf{S}$ to $\mathbf{T}$: 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 $\mathbf{S}$, 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 $\mathbf{S}$). We may refer to it as a P-*atom*. In the context of a schema mapping $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$, we may refer to a P-atom where P is in $\mathbf{S}$ as a *source atom*, and a P-atom where P is in $\mathbf{T}$ as a *target atom*. If P is an m-ary relation symbol in $\mathbf{S}$, 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 $\mathbf{S}$). 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 const (that is not part of any schema). We refer to the formula const(x) for a variable x as a const *formula*; the intended interpretation of const is that const(x) should hold precisely if x is a constant.

If δ is a conjunction of relational atoms (but no const formulas), then we define I_δ to be an instance obtained from δ as follows. For each variable v, assign a distinct fixed constant c_v, and let the facts of I_δ consist of the facts $P(c_{v_1}, \ldots, c_{v_k})$ where $P(v_1, \ldots, v_k)$ is an atom in δ. For example, if δ is $P(x, y) \wedge Q(y)$, then I_δ is the instance $\{P(c_x, c_y), Q(c_y)\}$. If δ is a conjunction of relational atoms and at least one const formula, then we define I_δ as follows. For each variable v such that const(v) is in δ, assign a distinct fixed constant c_v, and for each remaining variable v assign a distinct fixed null n_v. Define I_δ be the facts that result by taking each relational atom in δ and doing a replacement using the assignment we just described. For example, if δ is $P(x, y) \wedge Q(y) \wedge \text{const}(x)$, then I_δ is the instance $\{P(c_x, n_y), Q(n_y)\}$. It is sometimes convenient to allow δ to contain also inequalities $x \neq y$, where x and y are distinct variables among the variables appearing in relational atoms of δ. In that case, we simply ignore the inequalities in defining I_δ. Note that if δ and δ' contain the same relational atoms, and δ has no const formula, while δ' contains the formulas const(x) for every variable x in the relational atoms of δ', then I_δ and $I_{\delta'}$ are the same, and contain only constants (no nulls). Throughout this article, we reserve the symbol δ (possibly with subscripts or primes) to be a conjunction of relational atoms, const formulas 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 *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, $P(x_1, x_2, x_1, x_3, x_2)$ 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.

Constraints. All sets of constraints we consider are finite, unless otherwise specified. We consider constraints of several forms. A *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 α is a conjunction of source atoms and β is a conjunction of target atoms. The variables in $\bar{x}$ are exactly the variables that appear in both α and β. Note in particular that this gives us the safety condition that every variable in β that is not existentially quantified appears also in α. We will generally suppress writing the $\forall\bar{x}\forall\bar{y}$ part. If $\bar{z}$ is empty, we say that φ is *full*. We use the standard notion of *satisfaction* of constraints, denoted $\models$. For example, if I is a source instance, J is a target instance, and Σ 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 Σ.

Homomorphisms. Let J, J' be two instances. A function h that maps values to values is a *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 *homomorphically equivalent* if there are homomorphisms from J to J' and from J' to J. We then write $J \leftrightarrow J'$.

Solutions and Universal Solutions. Let $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ be a schema mapping. Then J is a *solution* for I (under $\mathcal{M}$) if $(I, J) \models \Sigma$. We write $\mathrm{Sol}(\mathcal{M}, I)$ to denote the solutions for I under $\mathcal{M}$. We say that a solution U for the source instance I is a *universal solution* [Fagin et al. 2005a], if $U \rightarrow J$ for every solution J for I.

Composition and Inverse. We recall the concept of the *composition* of two schema mappings, introduced in Fagin et al. [2005b] and Melnik [2004], and the concept of an *inverse* of a schema mapping, introduced in [Fagin 2007].

Let $\mathcal{M}_{12} = (\mathbf{S}_1, \mathbf{S}_2, \Sigma_{12})$ and $\mathcal{M}_{23} = (\mathbf{S}_2, \mathbf{S}_3, \Sigma_{23})$ be schema mappings. The *composition* $\mathcal{M}_{12} \circ \mathcal{M}_{23}$ is a schema mapping $(\mathbf{S}_1, \mathbf{S}_3, \Sigma_{13})$ such that for every $\mathbf{S}_1$-instance I and every $\mathbf{S}_3$-instance J, we have that $(I, J) \models \Sigma_{13}$ if and only if there is an $\mathbf{S}_2$-instance K such that $(I, K) \models \Sigma_{12}$ and $(K, J) \models \Sigma_{23}$. As noted in Fagin et al. [2005b], the composition is unique. When the schemas are understood from the context, we will often write $\Sigma_{12} \circ \Sigma_{23}$ for the composition $\mathcal{M}_{12} \circ \mathcal{M}_{23}$.

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

The *copy mapping* is the schema mapping $\mathrm{Id} = (\mathbf{S}, \widehat{\mathbf{S}}, \Sigma_{\mathrm{Id}})$, where Σ_{Id} consists of the s-t tgds $R(x_1, \ldots, x_k) \rightarrow \widehat{R}(x_1, \ldots, x_k)$, where R is k-ary, and where R ranges over the relation symbols in $\mathbf{S}$. Thus, if I_1 is an $\mathbf{S}$-instance and I_2 is an $\widehat{\mathbf{S}}$-instance, then $(I_1, I_2) \models \Sigma_{\mathrm{Id}}$ if and only if $\widehat{I_1} \subseteq I_2$.

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

Chasing. If $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is an s-t tgd mapping, then *chasing*, or applying the *chase process* [Maier et al. 1979] to the $\mathbf{S_1}$-instance I with Σ_{12} produces an $\mathbf{S_2}$-instance U such that U is a universal solution for I under $\mathcal{M}_{12}$ Fagin et al. [2005a]. We may write $U = \mathrm{chase}_{12}(I)$, and say that U is *the* result of the chase. For definiteness, we use the version of the chase as defined in Fagin et al. [2005a], although it does not really matter, since whatever version of the chase we use, the results are all homomorphically equivalent. Similarly, we may write $\mathrm{chase}_{21}(I)$ for the result of chasing an $\mathbf{S_2}$-instance I with Σ_{21}. We shall also extend this notation to cases where Σ_{12} or Σ_{21} are not simply sets of s-t tgds, but where we also allow const formulas and inequalities in the premises.

3. *Deciding Invertibility*

In Fagin [2007], 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 Fagin [2007] 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 Fagin et al. [2008]) that if $\mathcal{M}_{12}$ is an s-t tgd mapping, and I and I' are source instances where $I \subseteq I'$, then $\mathrm{Sol}(\mathcal{M}_{12}, I') \subseteq \mathrm{Sol}(\mathcal{M}_{12}, I)$. If the opposite implication also necessarily holds, then $\mathcal{M}_{12}$ is said to have the *subset property* [Fagin et al. 2008]. Thus, an s-t tgd mapping $\mathcal{M}_{12}$ has the *subset property* if whenever I and I' are source instances where $\mathrm{Sol}(\mathcal{M}_{12}, I') \subseteq \mathrm{Sol}(\mathcal{M}_{12}, I)$, then $I \subseteq I'$. It was shown in Fagin et al. [2008] 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 $\mathrm{chase}_{12}(I) \to \mathrm{chase}_{12}(I')$, then $I \subseteq I'$.

To prove our next result, we shall make use of the following simple lemma by Fagin et al. [2005a].

LEMMA 3.1 [FAGIN ET AL. 2005A]. *If $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is an s-t tgd mapping, then the solutions for a ground instance I are exactly the target instances J such that* $\mathrm{chase}_{12}(I) \to J$.

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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, the following are equivalent:*

(1) *$\mathcal{M}_{12}$ is invertible.*
(2) *Whenever I and I' are ground instances where $\mathrm{Sol}(\mathcal{M}_{12}, I') \subseteq \mathrm{Sol}(\mathcal{M}_{12}, I)$, then $I \subseteq I'$ (the subset property).*
(3) *Whenever I and I' are ground instances where $\mathrm{chase}_{12}(I) \to \mathrm{chase}_{12}(I')$, then $I \subseteq I'$ (the homomorphic version of the subset property).*
(4) *For every relational atom A and ground instance I,*

$$\mathrm{chase}_{12}(I_A) \to \mathrm{chase}_{12}(I) \textit{ implies } I_A \subseteq I.$$

(5) *For every relational atom A and ground instance I with at most n facts,*

$$\mathrm{chase}_{12}(I_A) \to \mathrm{chase}_{12}(I) \textit{ implies } I_A \subseteq I$$

where n is the number of facts in $\mathrm{chase}_{12}(I_A)$.

PROOF. The equivalence of (1) and (2) was shown in Fagin et al. [2008]. We now show that (2) and (3) are equivalent. Assume first that (2) holds. To prove that (3) holds, assume that $\mathrm{chase}_{12}(I) \to \mathrm{chase}_{12}(I')$; we must show that $I \subseteq I'$. By (2), it is sufficient to show that $\mathrm{Sol}(\mathcal{M}_{12}, I') \subseteq \mathrm{Sol}(\mathcal{M}_{12}, I)$. Assume that $J \in \mathrm{Sol}(\mathcal{M}_{12}, I')$; we must show that $J \in \mathrm{Sol}(\mathcal{M}_{12}, I)$. Since $J \in \mathrm{Sol}(\mathcal{M}_{12}, I')$, it follows from

Lemma 3.1 that $\text{chase}_{12}(I') \to J$. Since also $\text{chase}_{12}(I) \to \text{chase}_{12}(I')$, it follows by transitivity of homomorphism that $\text{chase}_{12}(I) \to J$. It then follows from Lemma 3.1 that $J \in \text{Sol}(\mathcal{M}_{12}, I)$, as desired.

Assume now that (3) holds. To prove that (2) holds, assume that $\text{Sol}(\mathcal{M}_{12}, I') \subseteq \text{Sol}(\mathcal{M}_{12}, I)$; we must show that $I \subseteq I'$. By (3), it is sufficient to show that $\text{chase}_{12}(I) \to \text{chase}_{12}(I')$, Now $\text{chase}_{12}(I') \in \text{Sol}(\mathcal{M}_{12}, I')$. So, since $\text{Sol}(\mathcal{M}_{12}, I') \subseteq \text{Sol}(\mathcal{M}_{12}, I)$, it follows that $\text{chase}_{12}(I') \in \text{Sol}(\mathcal{M}_{12}, I)$. So by Lemma 3.1, it follows that $\text{chase}_{12}(I) \to \text{chase}_{12}(I')$, 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 $\text{chase}_{12}(I) \to \text{chase}_{12}(I')$ and $I \not\subseteq I'$. Since $I \not\subseteq I'$, 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 $\text{chase}_{12}(I_A) \to \text{chase}_{12}(I)$. Since also $\text{chase}_{12}(I) \to \text{chase}_{12}(I')$, it follows by transitivity of homomorphism that $\text{chase}_{12}(I_A) \to \text{chase}_{12}(I')$, 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 $\text{chase}_{12}(I_A) \to \text{chase}_{12}(I)$. Then necessarily also $\text{chase}_{12}(I_A) \to \text{chase}_{12}(I')$ 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-t tgd mappings, as the next proof shows.

THEOREM 3.3. *The problem of deciding if an s-t tgd mapping is invertible is coNP-complete.*

PROOF. The proof of coNP-hardness is in Fagin [2007]. We now show the coNP upper bound. We make use of the equivalence of (1) and (5) in Proposition 3.2. To check that $M_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ 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 $h : \text{chase}_{12}(I_A) \to \text{chase}_{12}(I)$, 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-t tgds 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-t tgd mapping is invertible is coNP-complete.*
(2) *The problem of deciding if a LAV s-t tgd mapping is invertible is coNP-complete.*
(3) *The problem of deciding if a full and LAV s-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-t tgd mappings. If an s-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 $\mathcal{M}$ [Fagin et al. 2008]. Since we are interested in inverses $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ of an s-t tgd mapping $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, the normal mappings of interest to us have source schema $\mathbf{S_2}$ and target schema $\widehat{\mathbf{S_1}}$.

Definition 4.1. A constraint is *normal* if it is of the form $\alpha \wedge \chi_A \wedge \eta \rightarrow A$, where α is a conjunction of source atoms, A is a target atom, χ_A is the conjunction of the formulas const(x) for every variable x of A, and η 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 α. 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 const predicate on all variables in A, but just *allow* inequalities on variables in A. In particular, χ_A is fully determined by A (which is why we write it with the subscript A), whereas η 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

$$P(x_1, x_2, x_3, x_3) \wedge \text{const}(x_2) \wedge \text{const}(x_3) \wedge (x_2 \neq x_3) \rightarrow Q(x_2, x_3, x_2) \quad (1)$$

If the formula 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 const(x_2) or 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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ be schema mappings. Let us say that Σ_{21} is *too strong (for $\mathcal{M}_{12}$)* if there is a ground $\mathbf{S_1}$-instance I and a ground $\widehat{\mathbf{S_1}}$-instance J such that $\widehat{I} \subseteq J$ but $(I, J) \not\models \Sigma_{12} \circ \Sigma_{21}$. So Σ_{21} is not too strong precisely if whenever there is a ground $\mathbf{S_1}$-instance I and a ground $\widehat{\mathbf{S_1}}$-instance J such that $\widehat{I} \subseteq J$, then $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$. If Σ_{12} is a set of s-t tgds, and Σ_{21} is arbitrary, then it follows from Proposition 5.2 of Fagin [2007] that Σ_{21} is not too strong precisely if $(I, \widehat{I}) \models \Sigma_{12} \circ \Sigma_{21}$ for every ground $\mathbf{S_1}$-instance I.

Example 4.2. Let $\mathbf{S_1}$ consist of the binary relation symbol P and the unary relation symbol T. Let $\mathbf{S_2}$ consist of the binary relation symbol P' and the unary relation symbols 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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$. It was shown in Fagin et al. [2008] that $\mathcal{M}_{12}$ is invertible, and that the inverse "requires inequalities" (details are in Fagin et al. [2008]). Let Σ'_{21} consists only of the constraint $P'(x, y) \rightarrow P(x, y)$. We now show that Σ'_{21} is too strong for $\mathcal{M}_{12}$. Let $I = \{T(0)\}$, and let $J = \widehat{I}$. So, of course, $\widehat{I} \subseteq J$ (in fact, we have equality). However, we now show that $(I, J) \not\models \Sigma_{12} \circ \Sigma'_{21}$, which shows that Σ'_{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 $\widehat{P}(0, 0)$, which contradicts the fact that $J = \{\widehat{T}(0)\}$. This is the desired contradiction. So indeed, Σ'_{21} is too strong for $\mathcal{M}_{12}$. Now let $\Sigma''_{21} = \{P'(x, y) \wedge (x \neq y) \rightarrow P(x, y)\}$ (so that Σ''_{21} is obtained from Σ'_{21} by adding the inequality $x \neq y$ to the premise). Then (as we now show), Σ''_{21} is not too strong for $\mathcal{M}_{12}$. In fact, if I is a ground $\mathbf{S_1}$-instance I and J is a ground $\widehat{\mathbf{S_1}}$-instance such that $\widehat{I} \subseteq J$, and if $U = \text{chase}_{12}(I)$, 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.

Let us say that Σ_{21} is *too weak (for $\mathcal{M}_{12}$)* if there is a ground $\mathbf{S_1}$-instance I and a ground $\widehat{\mathbf{S_1}}$-instance J such that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ but $\widehat{I} \not\subseteq J$. So Σ_{21} is not too weak precisely if whenever there is a ground $\mathbf{S_1}$-instance I and a ground $\widehat{\mathbf{S_1}}$-instance J such that such that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$, then $\widehat{I} \subseteq J$.

Example 4.3. Let $\mathcal{M}_{12}$ be as in Example 4.2. Let Σ'''_{21} be the empty set. Then $(I, J) \models \Sigma_{12} \circ \Sigma'''_{21}$ for every choice of I and J, and it follows easily that Σ'''_{21} is too weak for $\mathcal{M}_{12}$. As a more interesting example, let Σ''_{21} be as in Example 4.2. We now show that Σ''_{21} is too weak for $\mathcal{M}_{12}$. 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, $\widehat{I} \not\subseteq J$. So indeed, Σ''_{21} is too weak for $\mathcal{M}_{12}$.

The following simple proposition follows immediately from our definitions.

PROPOSITION 4.4. *Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ be schema mappings. Then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ if and only if Σ_{21} is not too strong and not too weak for $\mathcal{M}_{12}$.*

Example 4.5. Let $\mathcal{M}_{12}$ be as in Examples 4.2 and 4.3. Let $\Sigma_{21} = \{P'(x, y) \wedge (x \neq y) \rightarrow \widehat{P}(x, y),\ Q(x) \rightarrow \widehat{P}(x, x),\ T'(x) \rightarrow \widehat{T}(x)\}$, and let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$. It can be shown that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, and so Σ_{21} is not too strong and not too weak for $\mathcal{M}_{12}$.

If Σ_{21} is not too strong, then for every ground $\mathbf{S_1}$-instance I and every ground $\widehat{\mathbf{S_1}}$-instance J where $\widehat{I} \subseteq J$, 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 $\mathcal{M}_{21}$ is a normal inverse of $\mathcal{M}_{12}$, 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 $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ is a normal inverse of the s-t tgd mapping $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$. Let I be a ground $\mathbf{S_1}$-instance, and let U be an arbitrary universal solution for I with respect to $\mathcal{M}_{12}$. Then, $(U, \widehat{I}) \models \Sigma_{21}$, and U witnesses $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ when $\widehat{I} \subseteq J$.*

PROOF. Since $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, we know that $(I, \widehat{I}) \models \Sigma_{12} \circ \Sigma_{21}$, and therefore there exists some K such that $(I, K) \models \Sigma_{12}$ and $(K, \widehat{I}) \models \Sigma_{21}$. Let U be an arbitrary universal solution for I with respect to $\mathcal{M}_{12}$. 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

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

Assume that U satisfies the premise of φ on $\bar{a}, \bar{b}$. Then $K \models \alpha(h(\bar{a}), h(\bar{b}))$.[2] 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 φ on $\bar{a}, h(\bar{b})$. Hence, since $(K, \widehat{I}) \models \Sigma_{21}$, we must have $\widehat{I} \models \widehat{A}(\bar{a})$. This shows that $(U, \widehat{I}) \models \varphi$. Since φ is an arbitrary member of Σ_{21}, it follows that $(U, \widehat{I}) \models \Sigma_{21}$, as desired. Since $(U, \widehat{I}) \models \Sigma_{21}$ and $\widehat{I} \subseteq J$, 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 $\mathcal{M}_{21}$ is not normal, then there may be no universal solution for I that witnesses $(I, \widehat{I}) \models \Sigma_{12} \circ \Sigma_{21}$.

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

Let I be an arbitrary nonempty ground instance, and let U be an arbitrary universal solution for I with respect to Σ_{12}. Assume that $S(c)$ is a fact of I. Then U must contain facts $T(c, n)$ and $T(n, c)$ for some null n. If $(U, J) \models \Sigma_{21}$, then necessarily J contains the fact $\widehat{S}(n)$, which is not a fact of $\widehat{I}$. So U does not witness $(I, \widehat{I}) \models \Sigma_{12} \circ \Sigma_{21}$. 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 $(I, \widehat{I}) \models \Sigma_{12} \circ \Sigma_{21}$. It is then straightforward to see that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$. Thus, $\mathcal{M}_{21}$ is a (non-normal) inverse of $\mathcal{M}_{12}$ where no universal solution witnesses $(I, \widehat{I}) \models \Sigma_{12} \circ \Sigma_{21}$.

Let $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$, where Σ'_{21} consists of the single normal constraint $T(x, y) \wedge T(y, x) \wedge \text{const}(x) \rightarrow \widehat{S}(x)$. Then $\mathcal{M}'_{21}$ is a normal inverse of $\mathcal{M}_{12}$, 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. It says that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ if and only if $\text{chase}_{21}(\text{chase}_{12}(I)) = \widehat{I}$ 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.

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

PROOF. Assume first that $\text{chase}_{21}(\text{chase}_{12}(I)) \subseteq \widehat{I}$ for every ground instance I. We must show that whenever there are ground instances I and J such that $\widehat{I} \subseteq J$, then $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$. Let I and J be ground instances such that $\widehat{I} \subseteq J$. 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, $U' \subseteq \widehat{I}$. Since also $\widehat{I} \subseteq J$, it follows that $U' \subseteq J$. Since $(U, U') \models \Sigma_{21}$ and $U' \subseteq J$, we see that $(U, J) \models \Sigma_{21}$. 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.

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

Assume now that Σ_{21} is not too strong. So $(I, \widehat{I}) \models \Sigma_{12} \circ \Sigma_{21}$. Let $U = \text{chase}_{12}(I)$, and let $U' = \text{chase}_{21}(U)$. We must show that $U' \subseteq \widehat{I}$. The argument in the proof of Proposition 4.6 shows that $(U, \widehat{I}) \models \Sigma_{21}$. Since $\mathcal{M}_{21}$ is a normal mapping, it is easy to see that the result of chasing an arbitrary instance with Σ_{21} is a ground instance. In particular, U' is a ground instance. Since U' is the result of chasing U with Σ_{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 $\widehat{I}$, then we see that $U' \subseteq \widehat{I}$, as desired. □

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

PROOF. Assume first that Σ_{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 $\widehat{I} \subseteq J$. Let J be $\text{chase}_{21}(\text{chase}_{12}(I))$. Then $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$. So $\widehat{I} \subseteq J$, that is, $\widehat{I} \subseteq \text{chase}_{21}(\text{chase}_{12}(I))$, as desired.

Assume now that $\widehat{I} \subseteq \text{chase}_{21}(\text{chase}_{12}(I))$ 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 $\widehat{I} \subseteq J$. 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 $\text{chase}_{12}(I) \to K$. Similarly, $\text{chase}_{21}(K) \to J$. Since $\text{chase}_{12}(I) \to K$, we see (by chasing both sides with Σ_{21}) that $\text{chase}_{21}(\text{chase}_{12}(I)) \to \text{chase}_{21}(K)$. Hence, since also $\text{chase}_{21}(K) \to J$, we have $\text{chase}_{21}(\text{chase}_{12}(I)) \to J$, by transitivity of homomorphism. Since also $\widehat{I} \subseteq \text{chase}_{21}(\text{chase}_{12}(I))$, it follows that $\widehat{I} \to J$. Therefore, since $\widehat{I}$ and J are both ground, we have $\widehat{I} \subseteq J$, as desired. □

We now discuss a nice property of normal inverses. Corollary 7.4 of Fagin [2007] says that if $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ are both full s-t tgd mappings, then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ if and only if $\text{chase}_{21}(\text{chase}_{12}(I)) = \widehat{I}$ for every ground instance I, The next theorem says that this strong property (that $\text{chase}_{21}(\text{chase}_{12}(I)) = \widehat{I}$ for every ground instance I) holds for normal inverses $\mathcal{M}_{21}$, even when $\mathcal{M}_{12}$ is not full.

THEOREM 4.10. *Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is an s-t tgd mapping, and assume that $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ is a normal mapping. Then, $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ if and only if $\text{chase}_{21}(\text{chase}_{12}(I)) = \widehat{I}$ for every ground instance I.*

PROOF. This follows easily from Proposition 4.4, along with Lemmas 4.8 and 4.9. □

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 $\text{chase}_{21}(\text{chase}_{12}(I)) \not\subseteq \widehat{I}$. It is more challenging to find an example where $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ but $\widehat{I} \not\subseteq \text{chase}_{21}(\text{chase}_{12}(I))$. Such an example was given in Fagin et al. [2008], as we now describe.

Example 4.11. We now give an example from Fagin et al. [2008] where $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ but where there is a ground instance I such that $\widehat{I} \not\subseteq \text{chase}_{21}(\text{chase}_{12}(I))$. Let $\mathbf{S_1}$ consist of the unary relation symbol P, and let $\mathbf{S_2}$ consist of the binary relation symbol Q. Let Σ_{12} consist of $P(x) \to \exists y Q(x, y)$, and let Σ_{21} consist of the constraints $Q(x, y) \rightarrow P(y)$ and $Q(x, y) \wedge \text{const}(y) \rightarrow P(x)$.[3] Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$. Note that $\mathcal{M}_{21}$ is not normal, because the second member of Σ_{21} has const(y) rather than const(x) in its premise.

We now show that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$. 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 $\widehat{I} \subseteq J$.

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 $(K, \widehat{I}) \models \Sigma_{21}$. Hence, $(I, \widehat{I}) \models \Sigma_{12} \circ \Sigma_{21}$, which implies that if $\widehat{I} \subseteq J$, 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 $\widehat{I} \subseteq J$. 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 Σ_{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 Σ_{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 $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$.

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

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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$) is a conjunction such that (a) the atoms in the conjunction arise in the chase of A with Σ_{12}, and (b) all of these atoms can arise together in a chase with Σ_{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). 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 the atoms in the conjunction can arise together in a chase with Σ_{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 δ 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 δ contain target atoms, we also allow δ to contain const formulas const(x) for some or all of the variables x in A, where the const formulas const(x) in δ tell us what variables must be mapped onto themselves in the homomorphism. This way, instead of simply requiring that the target atoms

[3] Note that our choice of Σ_{21}, although clearly peculiar (given what Σ_{12} is), is what we intended: it is not a typographical error!

in δ 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 δ 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 δ contains no const formulas, this is treated the same as having it contain all of the const formulas const(x) for every variable x of A, in that the homomorphism must map each variable onto itself. In addition to target atoms and const formulas, we also allow δ 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_δ and I_A, as defined earlier, rather than δ and A directly. We now give the formal definition of a relevant conjunction δ for a source atom A.

Definition 5.1. Let Σ_{12} be a finite set of s-t tgds. Assume that A is a relational atom, and δ is a conjunction $\alpha \wedge \chi \wedge \eta$, where α is a conjunction of relational atoms, χ is a conjunction (possibly empty) of *const* formulas const(x) for variables x in A, and η is a conjunction (possibly empty) of inequalities of the form $x \neq y$ for distinct variables x, y in A. Let us say that δ is *relevant for A (with respect to Σ_{12})* if $I_\delta \rightarrow$ *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 Σ_{12} is full. The const formulas play no role in inverses of full s-t tgd mappings, as shown in Fagin et al. [2008] (see Proposition 7.11 for a precise statement of what what we mean by "play no role"). Therefore, it is natural to assume in the full case that the relevant conjunction δ has no const formulas. Then I_δ 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, δ 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 δ is relevant for A if and only if "the atoms in δ 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 Σ_{12} consist of the full s-t tgds $P(x, y) \rightarrow (R(x, y) \wedge S(x, x) \wedge T(y))$ 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 δ_1 be $R(x, y) \wedge S(x, x)$, let δ_2 be $S(x, y)$, and let δ_3 be $S(x, x)$. Then, we have that I_{δ_1} is $\{R(c_x, c_y), S(c_x, c_x)\}$, that I_{δ_2} is $\{S(c_x, c_y)\}$, and that I_{δ_3} is $\{S(c_x, c_x)\}$. So δ_1 and δ_3 are both relevant for A, whereas δ_2 is not relevant for A.

Example 5.3. Assume that Σ_{12} consists of the s-t tgds $P(x) \rightarrow \exists z(R(x, z) \wedge S(x, x))$ and $Q(x, y) \rightarrow S(x, x)$. 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 $\{R(c_x, n), S(c_x, c_x)\}$ for a null n. Let δ_1 be $R(x, y) \wedge \text{const}(x)$. Then I_{δ_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, δ_1 is relevant for A. Now let δ_2 be $R(x, y)$. Then, I_{δ_2} is $\{R(c_x, c_y)\}$ for constants c_x, c_y. So δ_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 δ_3 be $S(x, y) \wedge \text{const}(x)$. Then, I_{δ_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, δ_3 is relevant for A.

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

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

Full Case. We now examine the notion of "demanding" when Σ_{12} is full. As in our discussion of the full case for the notion of "relevant", we assume that δ has no const formulas. As before, I_δ then contains only constants (no nulls). So we have that $I_\delta \rightarrow \text{chase}_{12}(I)$ holds if and only if $I_\delta \subseteq \text{chase}_{12}(I)$. Hence, δ is demanding for A if and only if whenever $I_\delta \subseteq \text{chase}_{12}(I)$, then $I_A \subseteq I$. This corresponds very well to the intuition we gave earlier that δ is demanding for A if "all of the atoms in δ 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 Σ_{12}, A, δ_1, δ_2, and δ_3 be as in Example 5.2. We now show that δ_1 is demanding for A. As before, I_{δ_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 $I_{\delta_1} \subseteq \text{chase}_{12}(I)$, that is, $\{R(c_x, c_y), S(c_x, c_x)\} \subseteq \text{chase}_{12}(I)$. In particular, $\{R(c_x, c_y)\} \subseteq \text{chase}_{12}(I)$. By looking at Σ_{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 $I_{\delta_2} \rightarrow \text{chase}_{12}(I)$. Therefore, it is automatically true that δ_2 is demanding for A. However, we now show that δ_3 is not demanding for A. Let I be $\{Q(c_x)\}$. It is easy to see that I_{δ_3} and $\text{chase}_{12}(I)$ 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 δ_3 is not demanding for A.

Example 5.6. Let Σ_{12}, A, δ_1, δ_2, and δ_3 be as in Example 5.3. We now show that δ_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_{δ_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, δ_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 δ_2 is demanding for A.

Finally, we show that δ_3 is not demanding for A. As in Example 5.3, we have that I_{δ_3} is 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, $I_A \nsubseteq I$. So δ_3 is not demanding for A.

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 Σ_{12}, A, and δ be as in Definition 5.1. We say that δ is *essential for A (with respect to* Σ_{12}*)* if δ is both relevant for A and demanding for A (with respect to Σ_{12}).

Example 5.8. Let Σ_{12}, A, δ_1, δ_2, and δ_2 be as in Examples 5.2 and 5.5. We see from these examples that δ_1 is both relevant and demanding for A, that δ_2 is demanding but not relevant for A, and that δ_3 is relevant but not demanding for A. So δ_1 is essential for A, but neither δ_2 nor δ_3 are essential for A.

Example 5.9. Let Σ_{12}, A, δ_1, δ_2, and δ_3 be as in Examples 5.3 and 5.6. We see from these examples that δ_1 is both relevant and demanding for A, that δ_2 is demanding but not relevant for A, and that δ_3 is relevant but not demanding for A. So δ_1 is essential for A, but neither δ_2 nor δ_3 are essential for A.

The s-t tgd mapping $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is said to have the *constant-propagation property* [Fagin 2007] if for every ground instance I, every member of the active domain of I is a member of the active domain of $\text{chase}_{12}(I)$ (i.e., $\text{dom}(I) \subseteq \text{dom}(\text{chase}_{12}(I))$). Later, we shall make use of the following proposition from [Fagin 2007].

PROPOSITION 5.10 [FAGIN 2007]. *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 δ is an essential conjunction for A with respect to the set Σ_{12} of s-t tgds. Then every variable in A appears in δ.*

PROOF. Assume that A is $P(v_1, \ldots, v_k)$, where $v_1, \ldots, v_k$ are variables, not necessarily distinct. Assume that the variable v_i does not appear in δ; 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 δ is relevant for A, we know that there is a homomorphism $h : I_\delta \to \text{chase}_{12}(I_A)$. So for the same homomorphism h, we have $h : I_\delta \to \text{chase}_{12}(I)$, since c_{v_i} does not appear in I_δ. Since $I_\delta \to \text{chase}_{12}(I)$, even though $I_A \not\subseteq I$, it follows that δ is not demanding for A, which contradicts the assumption that δ 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 φ is a formula, and f is a weak renaming, let φ^f be the result of replacing every variable x in φ by $f(x)$. We may refer to φ^f as a *weak renaming of* φ. If φ is a normal constraint with premise δ, then we say that f is *consistent with the inequalities of* δ if $f(x)$ and $f(y)$ are distinct for each inequality $x \neq y$ in δ. The intuition is that if f is consistent with the inequalities of the normal constraint φ, then φ^f is a special case of φ. For example, assume that φ 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, φ^f is

$$P(x_2, x_2, x_3, x_3) \wedge \text{const}(x_2) \wedge \text{const}(x_3) \wedge (x_2 \neq x_3) \to Q(x_2, x_3, x_2) \quad (2)$$

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

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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be an s-t tgd mapping and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ be a normal mapping. Then*

(1) *Σ_{21} is not too strong if and only if every constraint in Σ_{21} is of the form $\delta \to \widehat{A}$, where δ^f is demanding for A^f for every weak renaming f consistent with the inequalities of δ.*

(2) *Σ_{21} is not too weak if and only if for each source atom A, there is a relevant conjunction δ for A such that $\delta \to \widehat{A}$ is a weak renaming of a constraint in Σ_{21}.*

PROOF.
(1) Assume first that there is a constraint $\delta \to \widehat{A}$ of Σ_{21} and a weak renaming f consistent with the inequalities of δ such that δ^f is not demanding for A^f. Let δ' be δ^f, and let A' be A^f. Then, $\delta' \to \widehat{A'}$ is a normal constraint that is a logical consequence of Σ_{21}. Since δ' is not demanding for A', there is an instance I such that $I_{\delta'} \to \text{chase}_{12}(I)$, yet $I_{A'} \not\subseteq I$. Since $I_{\delta'} \to \text{chase}_{12}(I)$, it follows that $\widehat{I_{A'}}$ is the result of chasing $\text{chase}_{12}(I)$ with $\delta' \to \widehat{A'}$. So $\widehat{I_{A'}} \subseteq \text{chase}_{21}(\text{chase}_{12}(I))$ and therefore $\text{chase}_{21}(\text{chase}_{12}(I)) \not\subseteq \widehat{I}$. By Lemma 4.8, Σ_{21} is too strong.

Conversely, assume that Σ_{21} is too strong. Then, by Lemma 4.8, there is a ground instance I such that $\text{chase}_{21}(\text{chase}_{12}(I)) \not\subseteq \widehat{I}$. It follows that there must be a constraint of the form $\delta \to \widehat{A}$ in Σ_{21} such that the result of chasing $\text{chase}_{12}(I)$ with $\delta \to \widehat{A}$ 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 $I_{(\delta^f)} \to \text{chase}_{12}(I)$ and $I_{(A^f)} \not\subseteq I$. Hence, δ^f is not demanding for A^f.

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

Conversely, assume that for each source atom A, there is a relevant conjunction δ for A such that $\delta \to \widehat{A}$ is a weak renaming of a constraint $\varphi \in \Sigma_{21}$. Then, $I_\delta \to \text{chase}_{12}(I_A)$ because δ is relevant for A, and therefore we have $\widehat{I_A} \subseteq \text{chase}_{21}(\text{chase}_{12}(I_A))$ because φ fires on $\text{chase}_{12}(I_A)$ to introduce $\widehat{I_A}$. Since $\widehat{I_A} \subseteq \text{chase}_{21}(\text{chase}_{12}(I_A))$ for each source atom A, it follows easily that $\widehat{I} \subseteq \text{chase}_{21}(\text{chase}_{12}(I))$ for every ground instance I. By Lemma 4.9, Σ_{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 Σ_{21} to be not too strong, the members of Σ_{21} must be restricted. Let φ be a member of Σ_{21}. As we commented earlier, the formulas φ^f, where f is a weak renaming consistent with the inequalities of δ, can be viewed as all of the special cases of φ. So Part (1) of Theorem 5.12 says that Σ_{21} is not too strong if and only if for every special case φ^f of every member φ of Σ_{21}, if φ^f is $\delta^f \to \widehat{A^f}$, then δ^f is demanding for A^f. By the intuition we gave earlier for the notion of "demanding", δ^f being demanding for A^f means that, roughly speaking, "all of the atoms in δ^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 Σ_{21} to be not too weak, Σ_{21} must contain (or imply) certain constraints. If we again think of φ^f as a special case of a normal constraint φ, then Part (2) of Theorem 5.12 says that Σ_{21} is not too weak if and only if for each source atom A, there is a relevant conjunction δ for A such that $\delta \rightarrow \widehat{A}$ is a special case of a constraint in Σ_{21}. By the intuition we gave earlier for the notion of "relevant", δ being relevant for A means that, roughly speaking, "the atoms in δ 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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be an s-t tgd mapping and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ be a normal mapping. Then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ if and only if*

(1) *Every constraint in Σ_{21} is of the form $\delta \rightarrow \widehat{A}$, where δ^f is demanding for A^f for every weak renaming f consistent with the inequalities of δ.*
(2) *For each source atom A, there is a relevant conjunction δ for A such that $\delta \rightarrow \widehat{A}$ is a weak renaming of a constraint in Σ_{21}. (By Part (1), this relevant conjunction is essential.)*

PROOF. This follows immediately from Proposition 4.4 and Theorem 5.12. □

Definition 5.14. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ 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 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 Σ_{21}^e consist of all formulas $e(A) \wedge \eta_A \rightarrow \widehat{A}$, where A is a prime source atom and where $e(A)$ is defined, and η_A consists of all inequalities of the form $x \neq y$ where x and y are distinct variables of A.

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.*
(2) *$\mathcal{M}_{21}^e$ 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 δ is relevant for A and demanding for A'. Then A and A' are the same atom.*

PROOF. Since δ is relevant for A, we know that $I_\delta \to \text{chase}_{12}(I_A)$. Therefore, since δ 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. □

LEMMA 5.17. *Let $\mathcal{M}_{12}$ be a full s-t tgd mapping, and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ a normal inverse for $\mathcal{M}_{12}$. Let A be a source atom and B a target atom where $\widehat{I_A} \subseteq \text{chase}_{21}(I_B)$. Then B is demanding for A with respect to Σ_{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 $(U, \widehat{I}) \models \Sigma_{21}$. Since Σ_{21} is full, this implies further that $\text{chase}_{21}(U) \subseteq \widehat{I}$. Since $I_B \subseteq \text{chase}_{12}(I)$ and $\widehat{I_A} \subseteq \text{chase}_{21}(I_B)$, it follows that $\widehat{I_A} \subseteq \text{chase}_{21}(\text{chase}_{12}(I)) = \text{chase}_{21}(U)$. Since also $\text{chase}_{21}(U) \subseteq \widehat{I}$, we have that $\widehat{I_A} \subseteq \widehat{I}$, and so $I_A \subseteq I$, as desired. □

To prevent confusion, the reader should note that the inclusion $\widehat{I_A} \subseteq \text{chase}_{21}(I_B)$ in Lemma 5.17 uses chase_{21}, not 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 Σ_{12} is full, then we are interested in the situation where δ has no const formulas. Then I_δ contains only constants (no nulls). If δ is simply a relational atom, and if δ is demanding, then we call δ 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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be a full s-t tgd mapping. Assume that A is a source atom. Assume that the conjunction δ has no* const *formulas. If δ is demanding for A, then δ contains a demanding atom for A. If δ is essential for A, then δ contains an essential atom for A.*

PROOF. Assume by way of contradiction that δ is demanding for A, but that no atom B of δ is demanding for A. So for every atom B of δ, there is a ground instance J_B such that $I_B \to \text{chase}_{12}(J_B)$ and $I_A \not\subseteq J_B$. Since every member of I_B is a constant, and since $I_B \to \text{chase}_{12}(J_B)$, it follows that $I_B \subseteq \text{chase}_{12}(J_B)$. Let I be the union of the instances J_B. So $I_B \subseteq \text{chase}_{12}(I)$. From our assumptions about δ, it follows easily that I_δ is the union, over all atoms B in δ, of I_B. Therefore, since $I_B \subseteq \text{chase}_{12}(I)$ for every atom B of δ, it follows that $I_\delta \subseteq \text{chase}_{12}(I)$, and so $I_\delta \to \text{chase}_{12}(I)$. Since for every B we have that $I_A \not\subseteq J_B$, and I_A is a singleton set, it follows that $I_A \not\subseteq I$. So we have that $I_\delta \to \text{chase}_{12}(I)$ and $I_A \not\subseteq I$. This contradicts the assumption that δ is demanding for A. Therefore, δ contains a demanding atom for A, as desired.

Now assume that δ is essential for A. We have shown that δ contains a demanding atom B for A. Since δ is relevant for A, we have $I_\delta \to \text{chase}_{12}(I_A)$. But I_δ contains only constants (no nulls), and so $I_\delta \subseteq \text{chase}_{12}(I_A)$. As we noted, I_δ is the union, over all atoms B in δ, of I_B. Therefore, $I_B \subseteq I_\delta \subseteq \text{chase}_{12}(I_A)$, and so $I_B \to \text{chase}_{12}(I_A)$. 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 $\mathcal{M}_{12}$ be a full s-t tgd mapping. The following are equivalent.*

(1) *$\mathcal{M}_{12}$ 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 Σ_{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 Σ_{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 Fagin et al. [2008], 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 $\mathcal{M}$ has an inverse, then the canonical candidate inverse of $\mathcal{M}$ is also an inverse of $\mathcal{M}$. 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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ 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 Σ_{12}. Let ν_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 χ_A be the conjunction of the formulas const(x) for each variable x in A, and let ω_A be the conjunction of ν_A and χ_A. Let η_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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be an invertible s-t tgd mapping. Let A be a source atom. Then ω_A, as defined in Definition* 6.1*, is an essential conjunction for A. If $\mathcal{M}_{12}$ is full, then ν_A, as defined in Definition* 6.1*, is an essential conjunction for A.*

PROOF. It is clear that ω_A is relevant for A (as is ν_A, in the full case).

We now show that ω_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 ν_A is demanding for A in the full case. □

The *canonical candidate inverse* [Fagin et al. 2008] of an invertible s-t tgd mapping $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is the normal mapping $\mathcal{M}^c_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma^c_{21})$ where Σ^c_{21} contains, for every prime source atom A, the constraint $\nu_A \wedge \chi_A \wedge \eta_A \rightarrow \widehat{A}$. By Propositions 5.11 and 6.2, every variable in A appears in ν_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 Σ_{21} is played by Σ^e_{21}. Let $e(A) \wedge \eta_A \rightarrow \widehat{A}$ be an arbitrary member of Σ^e_{21}. Let δ be $e(A) \wedge \eta_A$. Then, δ is essential for A, and hence demanding for A and relevant for A. Let f be a weak renaming consistent with η_A. By construction of η_A, it follows that f is a one-to-one on the variables of A. Therefore, since δ is demanding for A, also δ^f is 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 Σ_{21} is played by Σ^e_{21}. 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, $\mathcal{M}^e_{21}$ is an inverse of $\mathcal{M}_{12}$, and so (3) of Theorem 5.15 holds, as desired. □

It is shown in Fagin et al. [2008] that if $\mathcal{M}_{12}$ is an invertible s-t tgd mapping, then the canonical candidate inverse of $\mathcal{M}_{12}$ is indeed an inverse of $\mathcal{M}_{12}$. The proof in Fagin et al. [2008] is quite complicated. We will now give a proof, based on the following proposition, that is much simpler (given our machinery).

THEOREM 6.3 [FAGIN, KOLAITIS, POPA, AND TAN 2008]. *Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be an invertible s-t tgd mapping. Then, the canonical candidate inverse of $\mathcal{M}_{12}$ is indeed an inverse of $\mathcal{M}_{12}$.*

PROOF. Let e be the function that assigns to each prime source atom A the formula ω_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 $\mathcal{M}^e_{21}$ is an inverse of $\mathcal{M}$. But $\mathcal{M}^e_{21}$ is the canonical candidate inverse of $\mathcal{M}_{12}$, and so the canonical candidate inverse of $\mathcal{M}_{12}$ is an inverse of $\mathcal{M}_{12}$, 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 $\mathcal{M}_{12}$ is the weakest normal inverse of $\mathcal{M}_{12}$. This proposition follows from results in Fagin et al. [2008] (although the notion of a normal inverse did not appear in Fagin et al. [2008]). 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 $\mathcal{M}_{12}$ be an s-t tgd mapping, let $\mathcal{M}^c_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma^c_{21})$ be the canonical candidate inverse of $\mathcal{M}_{12}$, and let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ be an arbitrary normal inverse of $\mathcal{M}_{12}$. Then Σ_{21} logically implies Σ^c_{21}.*

PROOF. Let σ be an arbitrary member of Σ^c_{21}; we must show that Σ_{21} logically implies σ. Assume not; we shall derive a contradiction.

Since Σ_{21} does not logically imply σ, 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 $\mathcal{M}_{21}$ that $(K, J') \models \Sigma_{21}$ (intuitively, the nonground facts of J have no effect in determining satisfaction of Σ_{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 σ as $\nu_A \wedge \chi_A \wedge \eta_A \rightarrow \widehat{A}$. 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 η_A), where $\widehat{I_A} \not\subseteq J'$. Since V_A is the result of chasing I_A with Σ_{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 $(I_A, J') \models \Sigma_{12} \circ \Sigma_{21}$. Since $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, and since I_A and J' are ground instances, it follows that $\widehat{I_A} \subseteq J'$. This is our desired contradiction. □

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 $\star$, and that I is the identity under this operation. Then, $X \star Y = I$. Apply Y' to both sides, and we get

$$Y' \star (X \star Y) = Y' \star I. \tag{3}$$

By associativity, the left-hand side of (3) is $(Y' \star X) \star Y = I \star Y = Y$, 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 [Fagin 1968]. 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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, where $\mathbf{S_1}$ consists of the unary relation symbol R, where $\mathbf{S_2}$ consists of the unary relation symbol S, and where Σ_{12} consists of the tgd $R(x) \to S(x)$. Let Σ_{21} consist of the normal constraint $S(x) \wedge \text{const}(x) \to \widehat{R}(x)$, and let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$. It is easy to see that $\mathcal{M}_{21}$ is a normal inverse of $\mathcal{M}_{12}$. A theorem we shall give later (Theorem 7.7) implies that $\mathcal{M}_{21}$ is in fact the unique normal inverse of $\mathcal{M}_{12}$.

Example 7.2. Let $\mathbf{S_1}$ consist of the unary relation symbol R, and let $\mathbf{S_2}$ consist of the binary relation symbol S. Let Σ_{12} consist of the tgd $R(x) \to S(x, x)$. Let Σ_{21} consist of the normal constraint $S(x, x) \wedge \text{const}(x) \to \widehat{R}(x)$, and let Σ'_{21} consist of the normal constraint $S(x, y) \wedge \text{const}(x) \to \widehat{R}(x)$. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$, and let $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$. It is straightforward to verify that $\mathcal{M}_{21}$ and $\mathcal{M}'_{21}$ are inequivalent normal inverses of $\mathcal{M}_{12}$.

Because of these two examples (but where the focus was on unique inverses specified by tgds), Fagin [2007] 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 Fagin [2007] commented that a reason why $\mathcal{M}_{12}$ in Example 7.2 does not have a unique inverse is that $\mathcal{M}_{12}$ 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 $\mathcal{C}$ 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 $\mathcal{M}$ to have a unique inverse in the class $\mathcal{C}$. We also show the surprising result that such mappings $\mathcal{M}$ are very special: they are very close to being copy mappings.

Say that two schema mappings $(\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $(\mathbf{S_1}, \mathbf{S_2}, \Sigma'_{12})$ are *equivalent* if Σ_{12} and Σ'_{12} are logically equivalent. It is useful to consider a weaker notion than equivalence. Recall that if $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ and $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ are schema mappings, then $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ 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 $\widehat{I} \subseteq J$. 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 Σ_{21} plays no role whatever in determining whether or not $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$. Based on this intuition, let us say that Σ_{21} and Σ'_{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 $(J_1, J_2) \models \Sigma'_{21}$.[4] We may also say that $(\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ and $(\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$ are then weakly equivalent. If $(\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ and $(\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$ 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

[4] This notion arises also in Fagin et al. [2008].

result of removing every nonground 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 $\mathcal{M}_{12}$ be a schema mapping, and let $\mathcal{M}_{21}$ and M'_{21} be weakly equivalent schema mappings. Then, $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$ if and only if $\mathcal{M}'_{21}$ is an inverse of $\mathcal{M}_{12}$.*

PROOF. By symmetry, we need only show that if $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, then $\mathcal{M}'_{21}$ is an inverse of $\mathcal{M}_{12}$. Let I, J be ground instances. Since $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, we know that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ if and only if $\widehat{I} \subseteq J$. To show that $\mathcal{M}'_{21}$ is an inverse of $\mathcal{M}_{12}$, 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 $\mathcal{M}_{21}$ 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 $(I, J) \models \Sigma_{12} \circ \Sigma'_{21}$. This was to be shown. □

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.

THEOREM 7.4. *No schema mapping has a unique inverse.*

PROOF. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be an invertible schema mapping. Assume that $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ is an inverse of $\mathcal{M}_{12}$. Define Σ'_{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 Σ''_{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 $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$ and let $\mathcal{M}''_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma''_{21})$. By construction, we see that Σ_{21}, Σ'_{21}, and Σ''_{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 Σ'_{21} and Σ''_{21} are not logically equivalent. So $\mathcal{M}'_{21}$ and $\mathcal{M}''_{21}$ are inverses of $\mathcal{M}_{12}$ that are not equivalent. □

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 tgd mapping has a normal inverse). In Example 7.1, we gave an s-t tgd 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 tgd mapping to have a unique normal inverse.

[5] Note that we define Σ'_{21} not by writing formulas, but simply by saying which pairs (J_1, J_2) satisfy Σ'_{21}.

THEOREM 7.5. *An invertible s-t tgd mapping has a unique normal inverse if and only if for every source atom A, if δ is an essential conjunction for A, and δ' is a demanding conjunction for A, both with formulas* 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 tgd mapping with a unique normal inverse. Let A be a source atom. Assume that δ is essential for A, and δ' is demanding for A, and both have formulas 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 δ' 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 $\mathcal{M}_{21}^e$ is an inverse of $\mathcal{M}_{12}$. Let σ' be $\delta' \wedge \eta_A \rightarrow \widehat{A}$, where η_A is a conjunction of the inequalities $x \neq y$ for distinct variables x, y of A. Let $\Sigma_{21} = \Sigma_{21}^e \cup \{\sigma'\}$. Let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$.

We now show that $\mathcal{M}_{21}$ is also an inverse of $\mathcal{M}_{12}$. Let δ'' be $\delta' \wedge \eta_A$. Let f be a weak renaming consistent with the inequalities of δ''. It follows easily from Theorem 5.13 that we need only show that $(\delta'')^f$ is demanding for A^f. Let I be a ground instance such that $I_{(\delta'')^f} \rightarrow \text{chase}_{12}(I)$; we must show that $I_{A^f} \subseteq I$. Since const(x) appears in δ'' only for variables x in A, and since f is consistent with the inequalities in η_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^f is simply A. Since const(x) appears in δ'' only for variables x where $f(x) = x$, it follows easily that f is a homomorphism from $I_{\delta''}$ to $I_{(\delta'')^f}$. Furthermore, $I_{\delta''} = I_{\delta'}$, 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 δ' is demanding for A, this implies that $I_A \subseteq I$. But $A = A^f$. Hence, $I_{A^f} \subseteq I$, as desired.

We now show that $\mathcal{M}_{21}^e$ 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 Σ_{21}^e. Clearly, $(J, I) \models \Sigma_{21}^e$. We now show that $(J, I) \not\models \sigma'$, and so $(J, I) \not\models \Sigma_{21}$. Note that because of the structure of Σ_{21}^e, it follows that I is a ground instance.

Let $\widehat{A}(\bar{c})$ be the result of chasing J with σ'. We need only show that $\widehat{A}(\bar{c})$ does not appear in I. Let σ be an arbitrary member of Σ_{21}^e. By construction of Σ_{21}^e, we know that σ is of the form $e(A) \wedge \eta_{A'} \rightarrow \widehat{A'}$, 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 σ does not produce $\widehat{A}(\bar{c})$. There are two cases.

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

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_δ to $I_{\delta'}$, that is, from I_δ to J, and since $e(A) = \delta$, it follows that σ 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 σ again does not produce $\widehat{A}(\bar{c})$.

We now prove the converse. Assume that for every source atom A, if δ is an essential conjunction for A, and δ' is a demanding conjunction for A, both with formulas 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 $\mathcal{M}_{21}^e = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21}^e)$ is an inverse of $\mathcal{M}_{12}$, by Theorem 5.15. Let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ be an arbitrary normal inverse of $\mathcal{M}_{12}$. We need only show that Σ_{21}^e and Σ_{21} are logically equivalent.

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

We now show that Σ_{21}^e logically implies Σ_{21}. Let σ be an arbitrary member of Σ_{21}. By part (1) of Theorem 5.13, we know that σ is of the form $\delta' \rightarrow \widehat{A}$, where A is a source atom and where $(\delta')^f$ is demanding for A^f for every weak renaming f consistent with the inequalities of δ'. For each weak renaming f consistent with the inequalities of δ', let τ_f be obtained from σ^f by adding to the premise of σ^f (if it is not already there) each inequality $x \neq y$ for every pair x, y of distinct variables in the conclusion of σ^f. It is fairly straightforward to see that σ is logically equivalent to the set of all such formulas τ_f. So to prove that Σ_{21}^e logically implies Σ_{21}, we need only show that Σ_{21}^e logically implies each such constraint τ_f.

Now τ_f is a normal constraint of the form $\delta'' \wedge \eta_{A'} \rightarrow \widehat{A'}$, where δ'' is demanding for A' (since as we said, $(\delta')^f$ is demanding for A^f), 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 δ for A' such that $\delta \wedge \eta_{A'} \rightarrow \widehat{A'}$ is a normal constraint in Σ_{21}^e. Let us denote this constraint by γ. Since δ is essential for A, and δ'' is demanding for A, it follows by assumption that $I_\delta \rightarrow I_{\delta''}$. It follows easily that γ logically implies τ_f. So Σ_{21}^e logically implies τ_f, as desired. □

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

PROPOSITION 7.6. *Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be an invertible s-t tgd mapping. Assume that whenever A is a source atom and δ' is a demanding conjunction for A with formulas* const(x) *precisely for the variables x of A, then* $\text{chase}_{12}(I_A) \rightarrow I_{\delta'}$. *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 δ is an essential conjunction for A, and δ' is a demanding conjunction for A, both with formulas const(x) for exactly the variables x that appear in A; we must show that $I_\delta \rightarrow I_{\delta'}$. Since δ is relevant for A, we have $I_\delta \rightarrow \text{chase}_{12}(I_A)$. By assumption, we have $\text{chase}_{12}(I_A) \rightarrow I_{\delta'}$. So $I_\delta \rightarrow I_{\delta'}$, as desired. □

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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is onto if for every target instance J there is a source instance I such that $\text{chase}_{12}(I) = J$. 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 $\mathcal{M}_{12}$ 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 $\mathbf{S_1}$ consist of four unary relation symbols P_i, for $1 \leq i \leq 4$, and let $\mathbf{S_2}$ consist of the four unary relation symbols Q_i, for $1 \leq i \leq 4$ and the unary relation symbol R. Let Σ_{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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$. The mapping $\mathcal{M}_{12}$ 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 $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$, where $\Sigma_{21} = \{Q_i(x) \wedge \text{const}(x) \rightarrow \widehat{P}_i(x) : 1 \leq i \leq 4\}$. Although $\mathcal{M}_{12}$ is not onto, we now show (by using Proposition 7.6) that $\mathcal{M}_{12}$ has a unique normal inverse, namely $\mathcal{M}_{21}$.

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 δ' be a demanding conjunction for A with const formula const(x) (and no other const formula). We now show that δ' must contain $Q_1(x)$. Assume not; we shall derive a contradiction.

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

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 $\mathbf{S_1}$ consist of two unary relation symbols P and Q, and let $\mathbf{S_2}$ consist of the binary relation symbol R. Let Σ_{12} consist of the s-t tgds $P(x) \rightarrow \exists y R(x, y)$, $Q(y) \rightarrow \exists x R(x, y)$, 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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, and let $\mathcal{M}^c_{21}$ be the canonical candidate inverse of $\mathcal{M}_{12}$. Thus, $\mathcal{M}^c_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma^c_{21})$, where Σ^c_{21} consists of the normal constraints $R(x, y) \wedge \text{const}(x) \rightarrow \widehat{P}(x)$ and $R(x, y) \wedge \text{const}(y) \rightarrow \widehat{Q}(y)$.

It is straightforward to verify that $\mathcal{M}^c_{21}$ is an inverse of $\mathcal{M}_{12}$. We now show that $\mathcal{M}^c_{21}$ is the unique normal inverse of $\mathcal{M}_{12}$. Let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ be another normal inverse of $\mathcal{M}_{12}$. By Proposition 6.4, we know that Σ_{21} logically implies Σ^c_{21}. So we need only show that Σ^c_{21} logically implies Σ_{21}. Let τ be an arbitrary member $\delta \rightarrow A$ of Σ_{21}; we must show that Σ^c_{21} logically implies τ. By symmetry in the roles of P and Q, we can assume without loss of generality that the conclusion A of τ is of the form $\widehat{P}(x)$, where x is a variable. By normality of τ, we know that const(x) appears in δ. We now show that there is a variable y such that the relational atom $R(x, y)$ appears in δ. Assume not; we shall derive a contradiction.

We now define a ground instance I. For each relational atom $R(v, w)$ in δ, 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 Σ_{12}, we see that the fact $R(c_v, c_w)$ is in chase$_{12}(I)$ for each relational atom $R(v, w)$ in δ. It follows easily that $I_\delta \rightarrow \text{chase}_{12}(I)$. Since by assumption, there is no variable y such that the relational atom $R(x, y)$ appears in δ, it follows by construction of I that $I_A \not\subseteq I$. Because $I_\delta \rightarrow \text{chase}_{12}(I)$ and $I_A \not\subseteq I$, we know that δ is not demanding for A. But $\delta \rightarrow A$ is in Σ_{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 δ.

Since δ contains $R(x, y)$ and const(x), it follows easily that τ is a logical consequence of the tgd $R(x, y) \wedge \text{const}(x) \rightarrow \widehat{P}(x)$, which is in Σ^c_{21}. So Σ^c_{21} logically implies τ, as desired. This completes the proof that $\mathcal{M}^c_{21}$ is the unique normal inverse of $\mathcal{M}_{12}$.

Let Σ'_{12} consist of the first two s-t tgds in Σ_{12}; that is, Σ'_{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 $\mathcal{M}'_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma'_{12})$. 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 Σ'_{21} consist of the normal constraints $R(x, y) \wedge \text{const}(x) \rightarrow \widehat{P}(x)$ and $R(x, y) \wedge \text{const}(y) \rightarrow \widehat{Q}(y)$, as before, along with a third normal constraint $R(x, x) \wedge R(x, y) \wedge \text{const}(y) \rightarrow \widehat{P}(y)$. Let $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$. Since the conclusion of the third constraint is $\widehat{P}(y)$ rather than $\widehat{Q}(y)$, 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 $\mathcal{M}_{12}$.

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 $\alpha \wedge \eta \rightarrow \beta$, where α is a conjunction of source atoms, β is a disjunction of formulas of the form $\exists \bar{z} \beta'$ with β' a conjunction of target atoms, and η is a conjunction (possibly empty) of inequalities of the form $x \neq y$ for distinct variables x, y of β that are not existentially quantified. Further, there is the safety condition that every variable in β that is not existentially quantified must appear in α. Again, we have suppressed writing the leading universal quantifiers.

Disjunctive tgds with inequalities were defined in Fagin et al. [2008], 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 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 const formulas dropped. The reason it is all right to drop the const formulas is because of a simple result in Fagin et al. [2008], which we now state, which says that const formulas play no role for inverses of full s-t tgd mappings.

PROPOSITION 7.11 [FAGIN, KOLAITIA, POPA, AND TAN 2008]. *Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be a full s-t tgd mapping. Let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$, where Σ_{21} is a set of disjunctive s-t tgds with constants and inequalities. Let $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$, where Σ'_{21} is obtained from Σ_{21} by removing every* 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.

PROPOSITION 7.12. *Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is a full s-t tgd mapping, $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ is an inverse of $\mathcal{M}_{12}$, and Σ_{21} is a set of disjunctive tgds with inequalities. Let I be a ground instance, and let $U =$* chase$_{12}(I)$. *Then, $(U, \widehat{I}) \models \Sigma_{21}$, and U witnesses $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$ when $\widehat{I} \subseteq J$.*

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

$$\alpha(\bar{x}, \bar{y}) \wedge \eta(\bar{x}) \to \psi(\bar{x}),$$

where $\eta(\bar{x})$ is a conjunction of inequalities (possibly empty) among the variables in $\bar{x}$, and where $\psi(\bar{x})$ is a disjunction of existentially-quantified conjunctions, and where the variables in $\psi(\bar{x})$ that are not existentially quantified are exactly

[6] In the definition of disjunctive tgds with inequalities that is given in Fagin et al. [2008], 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.

the members of $\bar{x}$. Assume that U satisfies the premise of φ on $\bar{a}, \bar{b}$. That is, $U \models \alpha(\bar{a}, \bar{b}) \wedge \eta(\bar{a})$. 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 Σ_{12} is full. Therefore $h(\bar{a}) = \bar{a}$ and $h(\bar{b}) = \bar{b}$, and so $K \models \alpha(\bar{a}, \bar{b}) \wedge \eta(\bar{a})$. Since $(K, \widehat{I}) \models \Sigma_{21}$, we must have $\widehat{I} \models \psi(\bar{a})$. This shows that $(U, \widehat{I}) \models \varphi$. Since φ is an arbitrary member of Σ_{21}, it follows that $(U, \widehat{I}) \models \Sigma_{21}$, as desired. Since $\widehat{I} \subseteq J$, 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. □

Recall that the *copy mapping*, which is used to define the inverse, is the schema mapping $\text{Id} = (\mathbf{S}, \widehat{\mathbf{S}}, \Sigma_{\text{Id}})$, where Σ_{Id} consists of the s-t tgds $R(\bar{x}) \rightarrow \widehat{R}(\bar{x})$ 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 Σ 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 Σ.

(3) Every target relation symbol appears in exactly one conclusion of Σ.

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 Σ_{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) ⇒ (2). From (3), we know that there is a schema mapping $\mathcal{M}'_{12}$ that is equivalent to $\mathcal{M}_{12}$ and that is a p-copy mapping. Clearly, $\mathcal{M}'_{12}$ is invertible and onto. It follows easily that $\mathcal{M}_{12}$ is invertible and onto, as desired.

We now show that (2) ⇒ (1). Assume that (2) holds. Since $\mathcal{M}_{12}$ is invertible, Theorem 6.3 tells us that the canonical candidate inverse is indeed an inverse of $\mathcal{M}_{12}$, so $\mathcal{M}_{12}$ has a normal inverse. By Proposition 7.11, the const formulas are irrelevant, and so $\mathcal{M}_{12}$ has an inverse specified by tgds with inequalities, and hence by disjunctive tgds with inequalities. Now assume that $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ and $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$ are both inverses of $\mathcal{M}_{12}$, where Σ_{21} and Σ'_{21} are disjunctive tgds with inequalities. We must show that Σ_{21} and Σ'_{21} are logically equivalent. We now show that Σ_{21} logically implies Σ'_{21}. By symmetry, we have that Σ'_{21} logically implies Σ_{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 Σ_{21} has no const formulas, it follows easily that $(J', K') \models \Sigma_{21}$. Since $\mathcal{M}_{12}$ 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 $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, we have that $\widehat{I} \subseteq K'$. Hence, since $\mathcal{M}'_{21}$ is an inverse of $\mathcal{M}_{12}$, 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 Σ'_{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 $\mathcal{M}_{12}$ 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 $(\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma^{e}_{21})$ is an inverse of A. Let Σ^{E}_{21} be the result of removing every const formula from every member of Σ^{e}_{21}, and let $\mathcal{M}^{E}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma^{E}_{21})$. Note that every member of Σ^{E}_{21} is of the form $B \wedge \eta_A \rightarrow \widehat{A}$, where B is an essential atom for A, and where η_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 A and B are the same, so η_A has inequalities among all distinct variables of B also. By Proposition 7.11, we know that $\mathcal{M}^{E}_{21}$ is an inverse of $\mathcal{M}_{12}$. Since (1) holds, $\mathcal{M}^{E}_{21}$ is the unique inverse that is specified by disjunctive tgds 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 Σ^{E}_{21} contains the formula $B \wedge \eta_A \rightarrow \widehat{A}$. Let ν_A be as in Definition 6.1. In particular, ν_A contains B and at least one more distinct relational atom B'. Form Σ_{21} from Σ^{E}_{21} by replacing $B \wedge \eta_A \rightarrow \widehat{A}$ by $\nu_A \wedge \eta_A \rightarrow \widehat{A}$, and let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$. Since B is demanding for A, it is easy to see that ν_A is demanding for A. It follows from Theorem 5.13 (and the fact that every weak renaming consistent with η_A is a strict renaming, since A and ν_A have the same variables) that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$. We now show that Σ_{21} is not logically equivalent to Σ^{E}_{21}.

Assume that B is an atom $Q(x_1, \ldots, x_t)$, and that F is a fact $Q(a_1, \ldots, a_t)$. 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 $(J, \emptyset) \not\models \Sigma^{E}_{21}$, but $(J, \emptyset) \models \Sigma_{21}$. So Σ_{21} is not logically equivalent to Σ^{E}_{21}, as desired. The fact that $(J, \emptyset) \not\models \Sigma^{E}_{21}$ follows from the fact that Σ^{E}_{21} contains the formula $B \wedge \eta_A \rightarrow \widehat{A}$, and

J contains a fact that is an exact match for B. It remains to show that $(J, \emptyset) \models \Sigma_{21}$. Let σ be a member of Σ_{21} except for $\nu_A \wedge \eta_A \rightarrow \widehat{A}$. Since J consists of a single fact that is an exact match for B, it follows that σ does not fire on J, because otherwise F would be an exact match for the atom B' in the premise of σ, 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 $(J, \emptyset) \models \nu_A \wedge \eta_A \rightarrow \widehat{A}$. To show this, we must show that $\nu_A \wedge \eta_A \rightarrow \widehat{A}$ does not fire on J. If it were to fire on J, then there would be a mapping h on the variables in ν_A that maps each atom in ν_A onto F and (because of η_A) is one-to-one on the variables in ν_A. Recall that B' is a relational atom in ν_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 $\nu_A \wedge \eta_A \rightarrow \widehat{A}$ does not fire on J, as desired. This concludes the proof that Σ_{21} is not logically equivalent to Σ^E_{21}.

Since $\mathcal{M}_{21}$ and $\mathcal{M}^E_{21}$ are both inverses of $\mathcal{M}_{12}$, specified by tgds with inequalities, even though Σ_{21} and Σ^E_{21} are not logically equivalent, this contradicts our assumption that (1) holds. This proves Claim 1.

Define Σ'_{12} to consist of all s-t tgds of the form $A \rightarrow e(A)$, where A is a prime atom with all variables distinct, and where $e(A)$ is as before (in particular, $e(A)$ is an essential atom for A). Note for later use that the variables in the premise A and conclusion $e(A)$ are the same by Proposition 5.20. Since $e(A)$ is essential for A, it follows that $e(A)$ is relevant 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 Σ_{12} logically implies Σ'_{12}. Later, we shall show that Σ_{12} is logically equivalent to Σ'_{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 Σ_{12}. Then, Σ'_{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 ν'_A like ν_A, except that the chase is with Σ'_{12} instead of Σ_{12}. Let B' be ν'_A. Note that B' is a singleton atom, because it arises only by firing the s-t tgd in Σ'_{12} whose premise is the P-atom with all variables distinct. Since Σ'_{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 Σ'_{12}, and since Σ_{12} logically implies Σ'_{12}, it follows that B' is in ν_A. So ν_A contains at least the two distinct atoms B and B'. This contradicts Claim 1, which is our desired contradiction.

Claim 3. Σ_{12} is logically equivalent to Σ'_{12}.

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

Assume that B is the atom $Q(x_1, \ldots, x_t)$ where $x_1, \ldots, x_t$ are variables (not necessarily distinct). Let τ be an arbitrary member of Σ_{12} of the form $\delta \rightarrow Q(z_1, \ldots, z_t)$, where $z_1, \ldots, z_t$ are variables, not necessarily distinct, and where x_i and x_j are the same variable whenever z_i and z_j are the same variable. Let h_τ be a function with domain the variables in τ such that $h_\tau(z_i) = x_i$ for each i (this is well defined, since x_i and x_j are the same variable whenever z_i and z_j are the same variable), and where h_τ maps each variable in the premise δ of τ that is not in the conclusion $Q(z_1, \ldots, z_t)$ of τ onto a new variable. Let τ' be the image of τ under h_τ. Thus, τ' is a weak renaming of τ. Note that τ' is not necessary a strict renaming of τ, 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 τ produces a fact F that is an exact match for B, then the chase of I with τ' also produces F.

By construction, the conclusion of τ' is B. Assume that the premise of τ' is δ'. Let ψ_τ be the formula $\exists \mathbf{y}\widehat{\delta'}$, where $\mathbf{y}$ consists of the variables in τ' that are not in B. Let Z consist of all such formulas ψ_τ. In particular, Z contains $\exists \mathbf{y}\widehat{\alpha}$, where $\mathbf{y}$ consists of all variables in α that are not in B. Now Z is finite, since its size is at most the number of members of Σ_{12}. Let γ be the formula $B \wedge \eta \rightarrow \zeta$, where η 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 α), and where ζ is the disjunction of the members of Z. Let $\Sigma_{21} = \Sigma^E_{21} \cup \{\gamma\}$, and let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$.

We now show that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, and that Σ_{21} is not logically equivalent to Σ^E_{21}. To show that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, we must show that for all ground instances I and J:

$$(I, J) \models \Sigma_{12} \circ \Sigma_{21} \text{ if and only if } \widehat{I} \subseteq J. \tag{4}$$

Since $\mathcal{M}^E_{21}$ is an inverse of $\mathcal{M}_{12}$, we know that for all ground instances I and J:

$$(I, J) \models \Sigma_{12} \circ \Sigma^E_{21} \text{ if and only if } \widehat{I} \subseteq J. \tag{5}$$

Now Σ_{21} logically implies Σ^E_{21}, since Σ_{21} is a superset of Σ^E_{21}. It follows easily that $\Sigma_{12} \circ \Sigma_{21}$ logically implies $\Sigma_{12} \circ \Sigma^E_{21}$. So if $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$, then $(I, J) \models \Sigma_{12} \circ \Sigma^E_{21}$, which, by (5), implies that $\widehat{I} \subseteq J$. Assume now that $\widehat{I} \subseteq J$; we must show that $(I, J) \models \Sigma_{12} \circ \Sigma_{21}$. Let $J^* = \text{chase}_{12}(I)$. Since $(I, J^*) \models \Sigma_{12}$, we need only show that $(J^*, J) \models \Sigma_{21}$. Now $(J^*, \widehat{I}) \models \Sigma^E_{21}$, by Proposition 7.12. and so $(J^*, J) \models \Sigma^E_{21}$ since $\widehat{I} \subseteq J$. Therefore, since $\Sigma_{21} = \Sigma^E_{21} \cup \{\gamma\}$, we need only show that $(J^*, J) \models \gamma$. Assume that γ fires on J^*. Then, there is a one-to-one mapping h (one-to-one because of η) 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 τ of Σ_{12} that generates F in the chase of I with Σ_{12}. Let τ' and δ' be as before. Since F is an exact match for B, it follows from an earlier comment that the chase of I with τ' generates F. It follows fairly easily that $\exists \mathbf{y}\delta'$ is satisfied in I under h, so $\exists \mathbf{y}\widehat{\delta'}$ is satisfied in $\widehat{I}$ under h. Since $\widehat{I} \subseteq J$, it follows that $\exists \mathbf{y}\widehat{\delta'}$ is satisfied in J under h. But $\exists \mathbf{y}\widehat{\delta'}$ is a disjunct in the conclusion of γ. Therefore, $(J^*, J) \models \gamma$, as desired. This concludes the proof that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$.

We now show that Σ_{21} is not logically equivalent to Σ_{21}^E. It is clear that $(I_B, \emptyset) \not\models \gamma$, and so $(I_B, \emptyset) \not\models \Sigma_{21}$. We now show that $(I_B, \emptyset) \models \Sigma_{12}^E$. Since, as we showed, there is no atom A such that B is essential for A with respect to Σ_{12}, no member of Σ_{21}^E has a strict renaming of B in its premise. So for each member $B' \wedge \eta_{A'} \rightarrow A'$ of Σ_{21}^E, there is no mapping h that maps B' onto B and satisfies $\eta_{A'}$. It follows that $(I_B, \emptyset) \models \Sigma_{12}^E$, as desired.

We have shown that $\mathcal{M}_{12}$ has two distinct, inequivalent inverses given by disjunctive tgds with inequalities, namely $\mathcal{M}_{21}^E$ and $\mathcal{M}_{21}$. 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 Σ'_{12} 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 Σ'_{12} whose conclusion is a Q-atom, where no two variables appearing in this Q-atom are the same. Assume not; we shall derive a contradiction. Let γ be the formula $Q(x_1, \ldots, x_t) \wedge \eta \rightarrow \beta$, where the variables $x_1, \ldots, x_t$ are distinct, where η is a conjunction of the inequalities $x_i \neq x_j$ whenever $i \neq j$, and where β is an arbitrary disjunction of source atoms whose variables altogether are exactly $x_1, \ldots, x_t$.[7] Let $\Sigma_{21} = \Sigma_{21}^E \cup \{\gamma\}$, and let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$.

We now show that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, and that Σ_{21} is not logically equivalent to Σ_{21}^E. Let $J^* = \text{chase}_{12}(I)$. Let $K = \text{chase}'_{12}(I)$, the result of chasing I with Σ'_{12}. Since (by Claim 3) Σ_{12} is logically equivalent to Σ'_{12}, and since both Σ_{12} and Σ'_{12} 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 $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, we need only show that $(J^*, J) \models \gamma$. Since by assumption there is no member of Σ'_{12} whose conclusion is a Q-atom with no two variables appearing in this Q-atom the same, and since, as we showed, $J^* = \text{chase}'_{12}(I)$ 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 γ does not fire on J^*, and so $(J^*, J) \models \gamma$, as desired.

We now show that Σ_{21} is not logically equivalent to Σ_{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 $(J, \emptyset) \not\models \gamma$, and so $(J, \emptyset) \not\models \Sigma_{21}$. We now show that $(J, \emptyset) \models \Sigma_{21}^E$. Each member of Σ_{21}^E is of the form $e(A) \wedge \eta_A \rightarrow A$, Since, by assumption, there is no member of Σ'_{12} 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 Σ_{21}^E is a conclusion of a member of Σ'_{12}, it follows that no atom in a premise of a member of Σ_{21}^E is a Q-atom with no two variables appearing in this Q-atom the same. Therefore, no member of Σ_{21}^E fires on J. So $(J, \emptyset) \models \Sigma_{21}^E$, as desired.

We have shown that $\mathcal{M}_{12}$ has two distinct, inequivalent inverses given by disjunctive tgds with inequalities, namely $\mathcal{M}_{21}^E$ and $\mathcal{M}_{21}$. This contradiction shows that there must be at least one member σ of Σ'_{12} whose conclusion is a Q-atom with no two variables appearing in this Q-atom the same.

[7] A disjunction is required if no source atom has arity at least t.

We now show that there can be no other member σ' of Σ'_{12} whose conclusion is a Q-atom. Assume not; we shall derive a contradiction. Assume that the premise of σ is a P-atom and the premise of σ' is a P'-atom. Since σ and σ' are different, we know that P and P' are different, by construction of Σ'_{12}. Since no two variables appearing in the conclusion of σ are the same, there is a mapping h that maps the variables in σ to the variables in σ' that maps the conclusion of σ onto the conclusion of σ'. Let A be the P-atom that is the result of applying h to the premise of σ. So the chase of I_A with Σ'_{12} is $I_{B'}$, where B' is the conclusion of σ'. This contradicts the fact that conclusion of σ' is essential for the premise of σ'. This contradiction shows that the only member of Σ_{12} whose conclusion is a Q-atom is σ, 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 Σ'_{12} are the same by Proposition 5.20, since the target is essential for the source with respect to Σ_{12}. By construction, for every source relation symbol P, there is exactly one member of Σ'_{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 Σ'_{12} whose conclusion is a Q-atom, and no two variables appearing in this Q-atom are the same. It follows that $\mathcal{M}_{12}$ is a p-copy mapping. Also, by Claim 3, Σ_{12} is logically equivalent to Σ'_{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 "$\mathcal{M}_{12}$ 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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is a full s-t tgd mapping that is invertible and onto. By Theorem 7.14, we know that $\mathcal{M}_{12}$ is equivalent to a p-copy mapping. We can assume without loss of generality that $\mathcal{M}_{12}$ itself is a p-copy mapping. Since $\mathcal{M}_{12}$ is invertible, we know that the canonical candidate inverse is a normal inverse of $\mathcal{M}_{12}$. We now use Proposition 7.6 to show that $\mathcal{M}_{12}$ has a unique normal inverse. To apply Proposition 7.6, assume that A is a source atom and δ' is a demanding conjunction for A with formulas const(x) precisely for the variables x of A; we must show that $\text{chase}_{12}(I_A) \subseteq I_{\delta'}$.

If Σ_{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 call the atoms $P(y_1, \ldots, y_k)$ and $Q(y_{f(1)}, \ldots, y_{f(k)})$ *buddies*. Let χ_A be the conjunction of the formulas const(x) for the variables x of A. Let γ be the conjunction of the buddies of the relational atoms in δ', and let γ' be $\gamma' \wedge \chi_A$. Since δ' 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 $\mathcal{M}_{12}$ from Example 7.8. It has a unique normal inverse, but since $\mathcal{M}_{12}$ is not equivalent to a p-copy mapping, it follows from Theorem 7.14 that $\mathcal{M}_{12}$ does not have a unique inverse specified by disjunctive tgds with inequalities. In addition to $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ from Example 7.8, another inverse is specified by Σ_{21} along with the disjunctive tgd $R(x) \rightarrow (\widehat{P}_1(x) \vee \widehat{P}_3(x))$.

Define a *near p-copy mapping* to be a full s-t tgd mapping $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ where (i) for each member σ of Σ, the premise and conclusion of σ 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 $\mathbf{T}$ appears in the conclusion of exactly one member of Σ, and every member of $\mathbf{S}$ appears in the premise of at most one member of Σ. 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 $\mathbf{S}$ may fail to appear in Σ. 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 const(x) for every variable x that appears in the conclusion. Returning again to Example 7.8, we see that the unique normal inverse $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ is the constants-added version of a near p-copy mapping (it is only "near", because the relation symbol R does not appear in Σ_{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 σ_1 be the constraint $S(x) \wedge \text{const}(x) \rightarrow \widehat{R}(x)$, and let σ_2 be the constraint $S(x) \rightarrow \exists y \widehat{R}(y)$. In addition to the inverse $\mathcal{M}_{21}$ given in Example 7.1, which is specified by σ_1, another inverse is specified by $\{\sigma_1, \sigma_2\}$. The constraint σ_2 is not logically implied by the constraint σ_1, because of the const formula in the premise of σ_1 but not σ_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 $\mathcal{M}_{12}$.

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 Fagin [2007].

LEMMA 8.2 [FAGIN 2007]. *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.*

PROOF. As we noted earlier, it is straightforward to see that every p-copy mapping $\mathcal{M}_{12}$ has an invertible normal inverse (in fact, $\mathcal{M}_{12}$ itself is an inverse of its normal inverse). We now prove the converse for full s-t tgd mappings. Thus, let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ be a full s-t tgd mapping, and let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$ be an invertible normal inverse of $\mathcal{M}_{12}$. We shall show that $\mathcal{M}_{12}$ is equivalent to a p-copy mapping.

Whenever we speak of relevant, demanding, or essential atoms in this proof, we mean with respect to Σ_{12}. We shall reserve A and A' for source atoms (with relation symbols in $\mathbf{S_1}$), and B and B' for target atoms (with relation symbols in $\mathbf{S_2}$).

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 $\mathcal{M}_{21}$ plays the role of $\mathcal{M}_{12}$). Since $\mathcal{M}_{21}$ 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 B, and B is essential for A.

We now prove Claim 2. Since $\mathcal{M}_{12}$ 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 $\mathcal{M}_{21}$ plays the role of $\mathcal{M}_{12}$).

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

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

We now prove Claim 3. Assume that $A^f = A'$. 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, $(B_A)^f = B_{A'} = B_{A^f}$, 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 $\mathcal{M}_{21}$ 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 $\mathcal{M}_{21}$ plays the role of $\mathcal{M}_{12}$). So there is some member $\delta \rightarrow \widehat{A}$ of Σ_{21} that fires on I_B. Hence, $\text{chase}_{21}(I_\delta)$ includes $\widehat{I_A}$. By Claim 1, we have that $\text{chase}_{21}(I_{B_A}) = \widehat{I_A}$. So $\text{chase}_{21}(I_{B_A}) \subseteq \text{chase}_{21}(I_\delta)$. Since $\mathcal{M}_{21}$ 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, δ has B_A as a conjunct. Since the constraint $\delta \rightarrow \widehat{A}$ of Σ_{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^f}. So B' is relevant for A^f.

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

Claim 6. Σ_{12} and Σ'_{12} are logically equivalent.

We now prove Claim 6. Clearly, Σ_{12} logically implies Σ'_{12}. We now show that Σ'_{12} logically implies Σ_{12}. We first show that each of the constraints $A' \to B_{A'}$ is a logical consequence of Σ'_{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 $A' \to B_{A'}$ is $(A \to B_A)^f$. Therefore, $A' \to B_{A'}$ is a logical consequence of $A \to B_A$, and so of Σ'_{12}, as desired.

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

We conclude by showing that $\mathcal{M}'_{12}$ is a p-copy mapping. Let $A \to B_A$ be a member of Σ'_{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 Σ'_{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 Σ'_{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 Σ'_{12}.

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

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) *$\mathcal{M}_{12}$ is equivalent to a p-copy mapping.*
(4) *$\mathcal{M}_{12}$ 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 $\mathcal{M}_{12}$ be full.

Example 8.5. Let $\mathbf{S_1}$ consist of the unary relation symbols P_1 and P_2, and let $\mathbf{S_2}$ consist of the unary relation symbols Q_1 and Q_2. Let Σ_{12} consist of the s-t tgds $P_1(x) \rightarrow Q_1(x)$ and $P_2(x) \rightarrow \exists y(Q_2(x) \wedge Q_1(y))$. Let Σ_{12} consist of the s-t tgds $P_1(x) to Q_1(x)$ and $P_2(x) \rightarrow Q_2(x)$. Let Σ_{21} consist of the normal constraints $Q_1(x) \wedge \text{const}(x) \rightarrow \widehat{P_1}(x)$ and $Q_2(x) \wedge \text{const}(x) \rightarrow \widehat{P_2}(x)$. Let Σ'_{21} consist of the normal constraints $Q_1(x) \wedge \text{const}(x) \rightarrow \widehat{P_1}(x)$ and $Q_2(x) \wedge Q_1(y) \wedge \text{const}(x) \rightarrow \widehat{P_2}(x)$. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, let $\mathcal{M}'_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma'_{12})$, let $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$, and let $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$. It is straightforward to verify that $\mathcal{M}_{21}$ and $\mathcal{M}'_{21}$ are inequivalent normal inverses of $\mathcal{M}_{12}$, and $\mathcal{M}'_{12}$ is an inverse of $\mathcal{M}_{21}$. So condition (2) of Theorem 8.4 holds, since $\mathcal{M}_{21}$ is a normal inverse of $\mathcal{M}_{12}$, and $\mathcal{M}'_{12}$ is an inverse of $\mathcal{M}_{21}$. However, condition (1) of Theorem 8.4 fails, since $\mathcal{M}_{12}$ has two inequivalent normal inverses, namely $\mathcal{M}_{21}$ and $\mathcal{M}'_{21}$. 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 $(\mathbf{S}, \mathbf{T}, \Sigma)$ to be the sum of the sizes of the members of Σ. We show that there is a family of invertible, full s-t tgd mappings $\mathcal{M}$ where the minimal number of constraints in a normal inverse of $\mathcal{M}$ is exponential in the size of $\mathcal{M}$ (and so, since the size of a mapping is at least equal to the number of constraints in it, the normal inverse of $\mathcal{M}$ with the smallest size has size exponential in $\mathcal{M}$). 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 $\mathcal{M}$ is invertible, then the output of this polynomial-time algorithm is an inverse of $\mathcal{M}$. 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 $\mathbf{S_1} = \{P_0, \ldots, P_k\}$, and let $\mathbf{S_2} = \{P'_0, \ldots, P'_k, Q_0, \ldots, Q_{k-1}\}$. Assume that all of the relation symbols in $\mathbf{S_1}$ and $\mathbf{S_2}$ are $4k$-ary.

Let $x_1, \ldots, x_{4k}$ be distinct variables. Let S_1 consist of the s-t tgds $P_i(x_1, x_2, \ldots, x_{4k}) \rightarrow P'_i(x_1, x_2, \ldots, x_{4k})$, for $0 \leq i \leq k$. Define $\bar{x}^i$, for $0 \leq i \leq k-1$, by letting $x^i_{4i+2} = x_{4i+1}$, $x^i_{4i+4} = x_{4i+3}$, and $x^i_j = x_j$ 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

$$(x_1, x_2, x_3, x_4, x_5, x_5, x_7, x_7, x_9, x_{10}, \ldots, x_{4k-1}, x_{4k}).$$

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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$.

We first show that $\mathcal{M}_{12}$ is invertible. Let T_1 consist of the s-t tgds $P'_j(x_1, x_2, \ldots, x_{4k}) \rightarrow \widehat{P}_j(x_1, x_2, \ldots, x_{4k})$, for $1 \leq j \leq k$ (note that we do not include the case $j = 0$). Let T_2 consist of the formula

$$\begin{aligned} P'_0(x_1, x_2, \ldots, x_{4k}) \;&\wedge\; ((x_1 \neq x_2) \vee (x_3 \neq x_4)) \\ &\wedge\; ((x_5 \neq x_6) \vee (x_7 \neq x_8)) \\ &\wedge\; \ldots \\ &\wedge\; ((x_{4k-3} \neq x_{4k-2}) \vee (x_{4k-1} \neq x_{4k})) \\ &\rightarrow\; \widehat{P}_0(x_1, x_2, \ldots, x_{4k}). \end{aligned}$$

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

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

To show that $\mathcal{M}_{21}$ is an inverse of $\mathcal{M}_{12}$, it is sufficient to show $\widehat{I} = \text{chase}_{21}(\text{chase}_{12}(I))$ 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 $\widehat{P}_i(\bar{a})$ is a fact of $\widehat{I}$ (and so $P_i(\bar{a})$ 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 $\widehat{P}_i(\bar{a})$ is in $\text{chase}_{21}(\text{chase}_{12}(I))$. So assume that $\widehat{P}_0(\bar{a})$ is a fact of $\widehat{I}$ (and so $P_0(\bar{a})$ 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 $Q_i(\bar{x}^i) \rightarrow \widehat{P}_0(\bar{x}^i)$ in T_3 guarantee that $\widehat{P}_0(\bar{a})$ is a fact of $\text{chase}_{21}(\text{chase}_{12}(I))$.

Case 2. There is no i with $0 \le i \le k-1$ such that $a_{4i+2} = a_{4i+1}$ and $a_{4i+4} = a_{4i+3}$. Then, the s-t tgd $P_0(x_1, x_2, \ldots, x_{4k}) \to P'_0(x_1, x_2, \ldots, x_{4k})$ in S_1 and the formula in T_2 guarantee that $\widehat{P}_0(\bar{a})$ 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 of Σ_{12} and of Σ_{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 $P_i(a_1, a_2, \ldots, a_{4k})$ of I with a single member σ_1 of Σ_{12}, and then chasing the single tuple that results from this chase by a single member σ_2 of Σ_{21}. We must show that the result of this second chase is either the empty set, or is the fact $\widehat{P}_i(a_1, a_2, \ldots, a_{4k})$.

We now consider cases.

Case 1. σ_1 is the s-t tgd

$$P_0(x_1, x_2, \ldots, x_{4k}) \to P'_0(x_1, x_2, \ldots, x_{4k})$$

of S_1. Assume σ_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 Σ_{21} whose premise contains P'_0 is the formula in T_2, and so we may assume that σ_2 is this formula. Since the conclusion of σ_2 is $\widehat{P}_0(x_1, x_2, \ldots, x_{4k})$, it follows that chasing $P'_0(a_1, a_2, \ldots, a_{4k})$ with σ_2 gives either the empty set or the fact $\widehat{P}_0(a_1, a_2, \ldots, a_{4k})$, as desired.

Case 2. σ_1 is the s-t tgd

$$P_i(x_1, x_2, \ldots, x_{4k}) \to P'_i(x_1, x_2, \ldots, x_{4k})$$

of S_1, for some i ($1 \le i \le k$). Assume σ_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 Σ_{21} whose premise contains P'_i is the tgd $P'_i(x_1, x_2, \ldots, x_{4k}) \to \widehat{P}_i(x_1, x_2, \ldots, x_{4k})$, and so we may assume that σ_2 is this tgd. Clearly, chasing $P'_i(a_1, a_2, \ldots, a_{4k})$ with σ_2 gives the fact $\widehat{P}_i(a_1, a_2, \ldots, a_{4k})$, as desired.

Case 3. σ_1 is the s-t tgd

$$P_{i+1}(\bar{x}^i) \to P'_0(\bar{x}^i)$$

of S_2, for some i ($0 \le i \le k-1$). Assume σ_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 Σ_{21} whose premise contains P'_0 is the formula in T_2, and so we may assume that σ_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. σ_1 is the s-t tgd

$$P_0(\bar{x}^i) \to Q_i(\bar{x}^i)$$

of S_3, for some i ($0 \le i \le k-1$). Assume σ_1 was applied to the fact $P_0(a_1, a_2, \ldots, a_{4k})$ of I to obtain the fact $Q_i(a_1, a_2, \ldots, a_{4k})$. The only member of Σ_{21} whose premise contains Q_i is the s-t tgd $Q_i(\bar{x}^i) \to \widehat{P}_0(\bar{x}^i)$ of T_3, so we may assume that σ_2 is this s-t tgd. Clearly, the result of chasing $Q_i(a_1, a_2, \ldots, a_{4k})$ with σ_2 is either the empty set or the fact $\widehat{P}_0(a_1, a_2, \ldots, a_{4k})$, as desired.

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

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

$$\widehat{I} = \text{chase}'_{21}(\text{chase}_{12}(I)). \tag{6}$$

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 $\bar{a}$ (i.e., 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 $(\rho_1, \rho_3, \rho_5, \ldots, \rho_{4k-1})$ 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 ρ_1 or ρ_3 is 0, exactly one of ρ_5 or ρ_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 σ be a member of Σ'_{21} whose conclusion is of the form $\widehat{P_0}(\bar{x})$, 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* σ 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 σ. If i is not of type 1 with respect to σ, then let us say that it is of *type 0 with respect to* σ. 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 σ.

Let $\boldsymbol{\rho} = (\rho_1, \rho_3, \rho_5, \ldots, \rho_{4k-1})$ be a special equality profile, and let $\bar{a}$ be a special $4k$-tuple of constants with equality profile $\boldsymbol{\rho}$. Let I be a ground instance whose only fact is $P_0(\bar{a})$. Now chase$_{12}(I)$ consists of the single fact $P'_0(\bar{a})$ (the s-t tgds in S_3 cannot be applied in the chase since $\bar{a}$ is special). It follows from (6) that there must be a member $\sigma_{\boldsymbol{\rho}}$ of Σ'_{21} such that the chase of $P'_0(\bar{a})$ with $\sigma_{\boldsymbol{\rho}}$ produces $P_0(\bar{a})$. It is clear that $\sigma_{\boldsymbol{\rho}}$ must have the following properties:

(1) The conclusion of $\sigma_{\boldsymbol{\rho}}$ is of the form $\widehat{P_0}(\bar{x})$, 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 $\sigma_{\boldsymbol{\rho}}$ for each i where $\rho_i = 0$; and
(4) the only relation symbol that appears in the premise of $\sigma_{\boldsymbol{\rho}}$ 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 ρ_i with respect to $\sigma_{\boldsymbol{\rho}}$. We already have that if $\rho_i = 0$, then i is of type 0 with respect to $\sigma_{\boldsymbol{\rho}}$ (this follows from the third condition above for $\sigma_{\boldsymbol{\rho}}$). So we need only show that if $\rho_i = 1$, then i is of type 1 with respect to $\sigma_{\boldsymbol{\rho}}$. 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 $\bar{a}$ is special, and since $\rho_{i_1} = 1$, it follows that $\rho_{i_2} = 0$. Therefore, as noted before, i_2 is of type 0 with respect to $\sigma_{\boldsymbol{\rho}}$. Assume that i_1 is also of type 0 with respect to $\sigma_{\boldsymbol{\rho}}$; we shall derive a contradiction.

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

We now show that h respects each of the inequalities of $\sigma_{\boldsymbol{\rho}}$, that is, that if $y \ne y'$ is an inequality that appears in the premise of $\sigma_{\boldsymbol{\rho}}$, then $h(y) \ne h(y')$. Assume that $y \ne y'$ is an inequality that appears in the premise of $\sigma_{\boldsymbol{\rho}}$, 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 $\boldsymbol{\rho}$, 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 $\sigma_{\boldsymbol{\rho}}$. Hence, the inequality $x_{m_j} \ne x_{m_j+1}$, that is, the inequality $y \ne y'$, does not appear in the premise of $\sigma_{\boldsymbol{\rho}}$. This is our desired contradiction.

Let J_1 be the target instance that consists of all 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 $\sigma_{\boldsymbol{\rho}}$. Obtain J_2 from J_1 by replacing each occurrence of a_{i_1+1} by a_{i_1}. Define $\bar{a}' = (a_1', \ldots, a_{4k}')$ by letting $a_{i_1+1}' = a_{i_1}$, and letting $a_j' = a_j$ if $j \ne i_1 + 1$. Since $a_{i_2+1} = a_{i_2}$ (because $\rho_{i_2} = 0$), we have $a_{i_2+1}' = a_{i_2+1} = a_{i_2} = a_{i_2}'$. Thus, $a_{i_2+1}' = a_{i_2}'$.

Define h' by letting $h'(y) = h(y)$ if y is not x_{i_1+1}, and letting $h'(x_{i_1+1}) = h(x_{i_1})$, that is, $h'(x_{i_1+1}) = a_{i_1}$. So J_2 consists of all 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 $\sigma_{\boldsymbol{\rho}}$.

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

Case 1. $\{y, y'\}$ does not contain x_{i_1+1}. Then $h'(y) = h(y)$ and $h'(y') = h(y)$. Now $h(y) \ne h(y')$, since h respects the inequalities of $\sigma_{\boldsymbol{\rho}}$. Therefore, $h'(y) \ne h'(y')$, as desired.

Case 2. $\{y, y'\}$ contains x_{i_1+1} but not x_{i_1}. In particular, either y or y' is x_{i_1+1}; assume without loss of generality that y is x_{i_1+1}. Then $h'(y) = h(x_{i_1})$ and $h'(y') = h(y')$. Since y' is not x_{i_1} or x_{i_1+1}, and i_1 is odd, we know by the fact that $\bar{a}$ is a special $4k$-tuple that $h(y') \ne h(x_{i_1})$. So $h'(y') = h(y') \ne h(x_{i_1}) = h'(y)$. Therefore, $h'(y) \ne h'(y')$, as desired.

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

Since J_2 consists of all 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 $\sigma_{\boldsymbol{\rho}}$, and since h' respects each of the inequalities of $\sigma_{\boldsymbol{\rho}}$, it follows that the chase of J_2 with $\sigma_{\boldsymbol{\rho}}$ contains $\widehat{P}_0(\bar{a}')$. Form the source instance I_2 from the target instance J_2 by replacing each fact $P_0'(\bar{b})$ by $P_0(\bar{b})$. Let τ_1 be the s-t tgd $P_0(x_1, x_2, \ldots, x_{4k}) \to P_0'(x_1, x_2, \ldots, x_{4k})$. Clearly, the chase of I_2 with τ_1 is J_2.

Since i_1 and i_2 are buddies, there is s with $0 \le s \le k-1$ such that $\{i_1, i_2\} = \{4s+1, 4s+3\}$. Let I_2' be the set difference $I_2 \setminus P_0(\bar{a}')$, and let J_2' be the set difference $J_2 \setminus P_0'(\bar{a}')$. Then, the chase of I_2' with τ_1 is J_2'. Let I_2'' consist of the fact $P_{s+1}(\bar{a}')$. Let τ_2 be the s-t tgd $P_{s+1}(\bar{x}^s) \to P_0'(\bar{x}^s)$. Since $a_{i_1}' = a_{i_1+1}'$ (by construction) and $a_{i_2}' = a_{i_2+1}'$ (as noted earlier), the chase of I_2'' with τ_2 contains the fact $P_0'(\bar{a}')$. Let $I_3 = I_2' \cup I_2''$. So the chase of I_3 with $\{\tau_1, \tau_2\}$ contains $J_2' \cup \{P_0'(\bar{a}')\}$, which contains J_2. Since τ_1 and τ_2 are members of Σ_{12}, it follows that the chase of I_3 with Σ_{12}

contains J_2. Since the chase of I_3 with Σ_{12} contains J_2, and the chase of J_2 with $\sigma_{\boldsymbol{\rho}}$ contains $\widehat{P}_0(\bar{a}')$, it follows that $\text{chase}'_{21}(\text{chase}_{12}(I_3))$ contains $\widehat{P}_0(\bar{a}')$, which is not in $\widehat{I}_3$. Therefore, $\widehat{I}_3 \neq \text{chase}'_{21}(\text{chase}_{12}(I_3))$, which contradicts (6) when I is I_3. This is our desired contradiction.

We just showed that if $\boldsymbol{\rho} = (\rho_1, \rho_2, \rho_5, \ldots, \rho_{4k-1})$ is a special equality profile, then there is a member $\sigma_{\boldsymbol{\rho}}$ of Σ'_{21} such that for each odd i with $1 \leq i \leq 4k-1$, we have that i is of type ρ_i with respect to $\sigma_{\boldsymbol{\rho}}$. Therefore, $\sigma_{\boldsymbol{\rho}}$ and $\sigma_{\boldsymbol{\rho}'}$ are different when $\boldsymbol{\rho} \neq \boldsymbol{\rho}'$. Hence, Σ'_{21} has at least as many members as there are special equality profiles. Clearly, there are 2^k distinct special equality profiles. So Σ'_{21} has at least 2^k members. Since the size of the schema mapping $\mathcal{M}_{12}$ is linear in k, it follows that the number of constraints in $\mathcal{M}'_{21}$ is exponential in the size of $\mathcal{M}_{12}$. 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 $\alpha \wedge \chi_A \wedge \theta \rightarrow A$, where α is a conjunction of source atoms, A is a target atom, χ_A is the conjunction of the formulas $const(x)$ for every variable x of A, and θ 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 α. 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 σ be an s-t tgd whose premise consists only of P-atoms for some single relational symbol P. Define an equivalence relation $\mathcal{E}_\sigma$ on the variables that appear in σ 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 ith position of some atom in the premise of σ. Let $\mathcal{E}_\sigma$ be the most refined equivalence relation (largest number of equivalence classes) such that each Y_i is a subset of an equivalence class of $\mathcal{E}_\sigma$. It is easy to see that each equivalence class of $\mathcal{E}_\sigma$ 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 σ' from σ by replacing each variable x by $[x]$. It is easy to see that each of the atoms in the premise of σ' is the same atom (call it A). Intuitively, we have done a minimum unification of the atoms in the premise of σ, so that they all "unify" to A. Form $\sigma^\dagger$ from σ' by replacing the premise of σ' by A. Now $\sigma^\dagger$ is a "special case" of σ (thus, $\sigma^\dagger$ is obtained from σ by identifying some variables and then replacing the conjunction $A \wedge \cdots \wedge A$ by simply A), and so $\sigma^\dagger$ is a logical consequence of σ. As an example, let σ be $P(x, y, z) \wedge P(y, w, z) \rightarrow Q(z, w)$ Then $Y_1 = \{x, y\}$, $Y_2 = \{y, w\}$, and $Y_3 = \{z\}$. There are two equivalence classes of $\mathcal{E}_\sigma$, 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 σ' is $P(x, x, z) \wedge P(x, x, z) \rightarrow Q(z, x)$, and $\sigma^\dagger$ is $P(x, x, z) \rightarrow Q(z, x)$. Note that if σ has a singleton premise, then $\sigma^\dagger$ is simply σ 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 σ 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 σ equals the chase of I with $\sigma^\dagger$.*

PROOF. Since σ logically implies $\sigma^\dagger$, it follows easily that the chase of I with $\sigma^\dagger$ is contained in the chase of I with σ. We now show the opposite inclusion. Assume that σ fires on I. Then there is a homomorphism h from the the premise of σ 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 $\mathcal{E}_\sigma$. 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 $\mathcal{E}_\sigma$. 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 σ under h equals the homomorphic image of the conclusion of $\sigma^\dagger$ under h'. It follows that the chase of I with σ 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 $\mathcal{M}_{12}$ 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 $\mathcal{M}_{12}$ if $\mathcal{M}_{12}$ has an inverse.*[8]

PROOF. Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, where Σ_{12} is a finite set of full s-t tgds. For each member $\varphi(\bar{x}) \rightarrow (A_1 \wedge \cdots \wedge A_r)$ of Σ_{12}, where each A_i is an atom, let Σ'_{12} contain the s-t tgds $\varphi(\bar{x}) \rightarrow A_1, \ldots, \varphi(\bar{x}) \rightarrow A_r$. Thus, Σ'_{12} is a finite set of full s-t tgds, each with a singleton conclusion, that is logically equivalent to Σ_{12}.

We now give a procedure to augment Σ'_{12} to a set Σ''_{12}. Let U be the set of all of the s-t tgds $\sigma^\dagger$, as defined earlier, and let $\Sigma''_{12} = \Sigma'_{12} \cup U$. Since Σ''_{12} consists of Σ'_{12} along with some logical consequences of Σ'_{12}, it follows that Σ''_{12} is logically equivalent to Σ'_{12}. Since also Σ'_{12} is logically equivalent to Σ_{12}, it follows that Σ''_{12} is logically equivalent to Σ_{12}. By renaming variables if needed, we can assume that no two distinct members of Σ''_{12} have a variable in common. Furthermore, we find

[8] 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 $\mathcal{M}_{12}$ has no inverse.

it convenient to assume that for each member σ of Σ''_{12}, there is another member $\sigma^\diamond$ of Σ''_{12} that is obtained from σ by renaming the variables in a one-to-one manner and with a disjoint set of variables from σ (we add $\sigma^\diamond$ to Σ''_{12} if needed). It is easy to see that there is a polynomial-time procedure for generating Σ''_{12} from Σ_{12}.

Let us say that a member σ of Σ''_{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 σ be a special member of Σ''_{12}. Assume that σ is $P(x_{z_1}, \ldots, x_{z_t}) \to Q(x_{i_1}, \ldots, x_{i_k})$. So the conclusion of σ is a Q-atom. Let τ be an arbitrary member of Σ''_{12}, other than σ, such that the conclusion of τ is a Q-atom. Assume that the conclusion of τ is $Q(x_{j_1}, \ldots, x_{j_k})$. Recall that σ and τ have no variables in common. Let $\mathcal{E}^\tau$ be the most refined equivalence relation (largest number of equivalence classes) on the variables in σ and τ such that x_{i_ℓ} and x_{j_ℓ} are in the same equivalence class, for $1 \le \ell \le k$. Let θ^1_τ be a conjunction of equalities among the variables in σ, where the equality $x_{i_r} = x_{i_s}$ is an atom in θ^1_τ precisely if x_{i_r} and x_{i_s} are in the same equivalence class of $\mathcal{E}^\tau$. Intuitively, θ^1_τ tells us how to equate variables to obtain a minimum unification of the conclusions of σ and τ. For each equivalence class E of $\mathcal{E}^\tau$, select a unique representative. If this equivalence class E contains a variable in σ, then choose the representative of E to be a variable in σ. (The only times that the equivalence class E does not contain a variable in σ is when E consists of a variable in the premise of τ but not in the conclusion of τ. This is possible, since we are not assuming that τ is special.) Let $[x]^\tau$ denote the representative of the equivalence class of $\mathcal{E}^\tau$ containing x. Let us refer to the variables $[x_{i_1}]^\tau, \ldots, [x_{i_k}]^\tau$ as *distinguished*. Let us say that a P-atom $P(x_{w_1}, \ldots, x_{w_t})$ in the premise of τ is *distinguished* if $[x_{w_\ell}]^\tau$ is distinguished for $1 \le \ell \le t$. If A is the distinguished P-atom $P(x_{w_1}, \ldots, x_{w_t})$, define γ_A to be the conjunction of the equalities $[x_{w_\ell}]^\tau = [x_{z_\ell}]^\tau$ for $1 \le \ell \le t$. Intuitively, γ_A tells us how to equate variables to obtain a minimum unification of A and the premise of σ. Define θ^2_τ to be the disjunction of the formulas γ_A for each distinguished P-atom A of τ. If this disjunction is empty (because τ has no distinguished P-atom), then θ^2_τ is the empty disjunction, which is logically equivalent to *False*. Intuitively, θ^2_τ tells us what possible collections of equalities among variables are needed to provide a minimum unification of some atom in the premise of τ with the premise of σ. Let θ_τ be the formula $\theta^1_\tau \to \theta^2_\tau$. Intuitively, θ^2_τ says that if the conclusions of τ and σ can unify to the same atom, then some atom in the premise of τ can unify to the same atom as the premise of σ. Note that if τ has no distinguished P-atom, then θ_τ is logically equivalent to $\neg\theta^1_\tau$. Let θ be the conjunction of the formulas θ_τ, over all members τ of Σ_{12} other than σ, where the conclusion of τ is a Q-atom. We now define σ^* to be $Q(x_{i_1}, \ldots, x_{i_k}) \wedge \theta \to \widehat{P}(x_{z_1}, \ldots, x_{z_t})$. Note that (the hatted version of) the premise of σ is the conclusion of σ^*, and the conclusion of σ is the relational atom in the premise of σ^*. Let Σ_{21} consist of all of the formulas σ^*, where σ is a special member of Σ_{12}. Obtain Σ'_{21} from Σ_{21} by adding to the premise of every member τ of Σ_{21} the conjuncts const(x) where x is a variable that appears in τ. Let $\mathcal{M}'_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma'_{21})$. The output of the algorithm is $\mathcal{M}'_{21}$. It is clear that our algorithm runs in polynomial time (and so, of course, the output $\mathcal{M}'_{21}$ is of polynomial size). Clearly, $\mathcal{M}'_{21}$ is a Boolean normal schema mapping. We shall show that $\mathcal{M}'_{21}$ is an inverse of $\mathcal{M}_{12}$ if $\mathcal{M}_{12}$ has an inverse.

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

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

Note that σ^* is derived from σ in Σ''_{12}, where σ is $\alpha \rightarrow \beta$. Assume that α is a P-atom and β is a Q-atom. We now prove the Claim. Assume first that β^f is essential for α^f; we wish to show that θ holds under f. To show this, we must show that if τ is a member of Σ''_{12} other than σ, and the conclusion of τ is a Q-atom, then θ_τ holds under f (this is because θ is the conjunction of such formulas θ_τ). So assume that θ^1_τ holds under f; we must show that θ^2_τ holds under f. Now the conclusion of σ^f is β^f. Since θ^1_τ holds under f, it follows that σ^f and τ^f have the same conclusion. So the conclusion of τ^f is β^f. Let g be a function that maps variables into constants and that agrees with f on equalities. So the conclusion of of τ^g is β^g. Let I be an instance whose facts are the facts A^g for each atom A in the premise of τ. So the chase of I with τ includes β^g. Since β^f is essential for α^f, it follows that α^g is a fact in I. So α^g is A^g for some atom A in the premise of τ. It follows that γ_A, as defined earlier, holds under g, and so θ^2_τ holds under g. Since f and g agree on equalities, this implies that θ^2_τ holds under f, as desired.

Assume now that θ holds under f; we must show that β^f is an essential atom for α^f. Certainly β^f is relevant for α^f, since σ is in Σ_{12}. So we must show that β^f is demanding for α^f. Let g be a function that maps variables into constants and that agrees with f on equalities. So θ holds under g. Let I be an instance where $\text{chase}_{12}(I)$ contains β^g; we need only show that α^g is a fact in I. It is easy to see that the result of chasing with Σ_{12} and Σ''_{12} are the same. So the result of chasing I with Σ''_{12} contains β^g. Hence, there is a constraint τ in Σ''_{12} that fires on I and produces β^g. If τ is σ, then it is not hard to see that this implies that α^g is in I, as desired. If τ is not σ, it is straightforward to verify that θ^1_τ holds under g. Since also θ holds under g, this implies that θ^2_τ holds under g. So there is some distinguished atom A in the premise of τ such that γ_A holds under g. Hence, A^g and α^g are the same fact. Let us denote the conclusion of τ by C. Since θ^1_τ holds under g, we have that $\beta^g = C^g$. So C^g arises by chasing I with τ. Since also τ has A in its premise, we have that A^g is in I. But we showed that A^g and α^g are the same fact. So α^g is in I, as desired. This concludes the proof of the Claim.

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

Assume that σ^* is a member of Σ_{21}, and σ^* is $\beta \wedge \theta \rightarrow \widehat{\alpha}$. Let k be the number of variables that appear in σ^*. Define the set T_{σ^*} as follows. For each weak renaming f of the variables in σ^* such that the range of f is in $\{x_1, \ldots, x_k\}$ and such that θ holds for f, let T_{σ^*} contain the constraints $\beta^f \wedge \eta_f \rightarrow \widehat{\alpha^f}$, where η_f is the conjunction of the inequalities $f(x) \neq f(y)$ where x and y are variables of σ^* 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_{σ^*} be finite.) It is straightforward to see that σ^* is logically equivalent to T_{σ^*}, and similarly the constants-added version of σ^* is logically equivalent to the constants-added version of T_{σ^*} (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 const(x) for every variable x that appears in the conclusion). Let Σ''_{21} be the union of the constants-added version of the sets T_{σ^*} over all σ^* in Σ_{21}, and let $\mathcal{M}''_{21} = (\mathbf{S_2}, \mathbf{S_1}, \Sigma''_{21})$. By Proposition 7.11, we know that $\mathcal{M}'_{21}$ is an inverse of $\mathcal{M}_{12}$ if and only if $\mathcal{M}''_{21}$ is an inverse of $\mathcal{M}_{12}$. So we need only show that $\mathcal{M}''_{21}$ is an inverse of $\mathcal{M}_{12}$.

We now use Theorem 5.13 to show that $\mathcal{M}''_{21}$ is an inverse of $\mathcal{M}_{12}$. Since each T_{σ^*} was obtained by considering weak renamings f such that θ holds for f, it follows easily from the Claim that for every member φ of Σ''_{21}, the premise of φ is essential for the conclusion of φ. Hence, the first condition of Theorem 5.13 holds (when Σ''_{21} plays the role of Σ_{21}). We now show that the second condition also holds. Let A be a source atom. Since $\mathcal{M}_{12}$ is invertible, we know by Proposition 6.2 that ω_A is essential for A, and so contains an atom B that is essential for A (with respect to Σ_{12}).

It follows from Lemma 9.3 that there is a member σ of Σ''_{12} with a singleton premise (and a singleton conclusion) such that the chase of I_A is the same with σ as it is with Σ_{12}. Write σ as $\alpha \rightarrow \beta$. So there is a weak renaming f such that α^f is A and β^f is B. We now show that every variable in α appears in β, and so σ is special. Assume that some variable x appears in α but not in β; 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 β^f, that is, B. So chase$_{12}(I_{\alpha^{f'}})$ contains I_B, even though $I_A \not\subseteq I_{\alpha^{f'}}$. This contradicts the fact that B is essential for A. Hence, σ is special, as desired.

So there is θ such that σ^* is $\beta \wedge \theta \rightarrow \widehat{\alpha}$, and σ^* is in Σ_{21}. Since β^f (namely, B) is essential for α^f (namely, A), it follows from the Claim that θ holds under f. Therefore, if χ_A is the conjunction of the formulas const(x) for every variable x of A, and if η_f is as above, then $B \wedge \chi_A \wedge \eta_f \rightarrow \widehat{A}$ is a weak renaming of a constraint in Σ''_{21}. Hence, the second condition of Theorem 5.13 holds (when Σ''_{21} plays the role of Σ_{21}), as desired. This completes the proof that $\mathcal{M}''_{21}$ is an inverse of $\mathcal{M}_{12}$.

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 $\mathcal{M}_{12}$, there is a relationship between the number of normal inverses of $\mathcal{M}_{12}$ and the minimal number of constraints in a Boolean inverse for $\mathcal{M}_{12}$. 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 $\mathbf{S_1}$ consist of the unary relation symbol P, and let $\mathbf{S_2}$ consist of the binary relation symbol Q. Let Σ_{12} consist of the s-t tgd $P(x) \rightarrow Q(x, x)$. Let Σ_{21}^k consist of the normal constraint

$$Q(x, y_1) \wedge Q(y_1, y_2) \wedge \cdots \wedge Q(y_{k-1}, y_k) \wedge Q(y_k, x) \wedge \text{const}(x) \rightarrow P(x).$$

Let $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$, and let $\mathcal{M}_{21}^k = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21}^k)$. It is straightforward to verify that for every ground instance I and for each $k \geq 1$ we have $\text{chase}_{21}^k(\text{chase}_{12}(I)) = \widehat{I}$ (where $\text{chase}_{21}^k(J)$ is the result of chasing J with Σ_{21}^k). It therefore follows from Theorem 4.10 that $\mathcal{M}_{21}^k$ is an inverse of $\mathcal{M}_{12}$ for every k. It is also straightforward to verify that Σ_{21}^k and $\Sigma_{21}^{k'}$ are not logically equivalent if $k \neq k'$. So $\mathcal{M}_{12}$ 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 $\mathcal{M}_{12}$ be a full s-t tgd mapping, with k source relation symbols. Assume that $\mathcal{M}_{12}$ has exactly $m \geq 1$ inequivalent normal inverses. Then $\mathcal{M}_{12}$ has a Boolean normal inverse with at most $k + \log_2(m)$ constraints.*

PROOF. Assume that $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$. Let us say that the source atom B is *good* if $\text{chase}_{12}(I_B)$ 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 $\mathcal{M}$) and that $\mathcal{M}$ 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 φ_B be the formula that is the conjunction of (1) the equalities $x_i = x_j$ 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, φ_B completely describes the equality pattern of the variables in B. Let θ_P be the disjunction of the formulas φ_B where B is a good P-atom. Let σ_P be the formula $\omega_{A_P} \wedge \theta_P \rightarrow \widehat{A_P}$, where ω_{A_P} 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 ω_{B_i} is essential for B_i with respect to Σ_{12}, for each i. Since B_i is bad, it follows that ω_{B_i} has more than one atom. By the construction in Definition 6.1, we see that ω_{B_i} contains the formulas const(x) for every variable x of B_i. So by Proposition 5.18, we know that some atom C_i in ω_{B_i} is essential for B_i with respect to Σ_{12}, for each i. If B_i is $P(y_1, \ldots, y_r)$, define η_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 ψ_i^0 be the constraint $C_i \wedge \eta_i \rightarrow \widehat{B_i}$, and let ψ_i^1 be the constraint $\omega_{B_i} \wedge \eta_i \rightarrow \widehat{B_i}$.

Let $\mathbf{v} = (v_1, \ldots, v_b)$ be an arbitrary $\{0, 1\}$-vector of length b. Define $\Sigma_{21}^{\mathbf{v}}$ to consist of the k formulas σ_P (one for each source relation symbol P) along with the b constraints $\psi_i^{v_i}$ for $1 \leq i \leq b$. Let $\mathcal{M}_{21}^{\mathbf{v}} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21}^{\mathbf{v}})$. We now show that each $\mathcal{M}_{21}^{\mathbf{v}}$ is an inverse of $\mathcal{M}_{12}$, and that $\mathcal{M}_{21}^{\mathbf{v}}$ and $\mathcal{M}_{21}^{\mathbf{v}'}$ are not equivalent if $\mathbf{v} \neq \mathbf{v}'$. Since the number of vectors $\mathbf{v}$ is 2^b, this shows that $2^b \leq m$. Further, since each $\mathcal{M}_{21}^{\mathbf{v}}$ is a Boolean normal inverse with $k + b$ constraints, this shows that $\mathcal{M}$ 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 $\mathbf{v} = (v_1, \ldots, v_b)$. We begin by showing that $\mathcal{M}_{21}^{\mathbf{v}}$ is an inverse of $\mathcal{M}_{12}$. We now define a function e that maps each prime source atom B to an essential conjunction $e(B)$ with respect to Σ_{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, $\mathcal{M}_{21}^{e}$ is an inverse of $\mathcal{M}$. We now show that $\mathcal{M}_{21}^{e}$ is equivalent to $\mathcal{M}_{21}^{\mathbf{v}}$, which completes the proof that $\mathcal{M}_{21}^{\mathbf{v}}$ is an inverse of $\mathcal{M}_{12}$.

For each prime source atom B where B is bad, $\mathcal{M}_{21}^{e}$ and $\mathcal{M}_{21}^{\mathbf{v}}$ contain the same constraint with conclusion $\widehat{B}$. Let us now consider the good prime source atoms B. The formula σ_P is logically equivalent to the set consisting of all of the formulas $\omega_{A_P} \wedge \varphi_B \rightarrow \widehat{A_P}$, where B is a good prime P-atom. Assume that B is a good prime source atom. Let σ_1 be the formula $\omega_{A_P} \wedge \varphi_B \rightarrow \widehat{A_P}$, and let σ_2 be the formula $\omega_B \wedge \eta_B \rightarrow \widehat{B}$, where as before η_B is the conjunction of all inequalities of the form $x \neq y$ where x and y are distinct variables in B. By construction, σ_2 is the unique member of $\mathcal{M}_{21}^{e}$ with conclusion $\widehat{B}$. So to complete the proof that $\mathcal{M}_{21}^{e}$ is equivalent to $\mathcal{M}_{21}^{\mathbf{v}}$, we need only show that the formula σ_1 is logically equivalent to the formula σ_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 ψ_B be the formula obtained from σ_1 by replacing the variable x_i by y_i, for $1 \leq i \leq r$. We now show that ψ_B is logically equivalent to both σ_1 and σ_2, which implies that σ_1 and σ_2 are logically equivalent, as desired. In forming ψ_B, two variables x_i and x_j in σ_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 φ_B. It follows easily that ψ_B is logically equivalent to σ_1. We now show that ψ_B is logically equivalent to σ_2.

It is easy to see that the conclusions of ψ_B and σ_2 are the same, and that the result of replacing the variable x_i by y_i, for $1 \leq i \leq r$, in φ_B is equivalent to η_B. Let τ_B be the result of replacing the variable x_i by y_i in ω_{A_P}, for $1 \leq i \leq r$. So we need only show that τ_B is equivalent to ω_B. Now the conjunct(s) of τ_B must be in ω_B, by properties of the chase with s-t tgds. Since ω_B is a singleton (because B is good), it follows easily that τ_B is the same as ω_B. This concludes the proof that $\mathcal{M}_{21}^{e}$ is equivalent to $\mathcal{M}_{21}^{\mathbf{v}}$,

We conclude the proof by showing that $\mathcal{M}_{21}^{\mathbf{v}}$ and $\mathcal{M}_{21}^{\mathbf{v}'}$ are not equivalent if $\mathbf{v} \neq \mathbf{v}'$. Say $\mathbf{v} \neq \mathbf{v}'$, and that $\mathbf{v} = (v_1, \ldots, v_b)$ and $\mathbf{v}' = (v_1', \ldots, v_b')$. 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 $\Sigma_{21}^{\mathbf{v}'}$ but not $\Sigma_{21}^{\mathbf{v}}$. This, of course, shows that $\mathcal{M}_{21}^{\mathbf{v}}$ and $\mathcal{M}_{21}^{\mathbf{v}'}$ are not equivalent. Clearly ψ_i^0 fires on I_{C_i}, and so $(I_{C_i}, \emptyset)$ does not satisfy ψ_i^0. Hence, $(I_{C_i}, \emptyset)$ does not satisfy $\Sigma_{21}^{\mathbf{v}}$, because $\Sigma_{21}^{\mathbf{v}}$ contains ψ_i^0. We now show that no member of $\Sigma_{21}^{\mathbf{v}'}$ fires on I_{C_i}. Since B_i is bad, we know that ω_{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 Σ_{12} is full, every variable in A is in B_i, and hence in C_i. Assume that C_i is $Q(y_1, \ldots, y_m)$. Then I_{C_i} consists of the fact $Q(c_{y_1}, \ldots, c_{y_m})$. If ψ_i^1 were to fire on I_{C_i}, then there would be a homomorphism h from the premise of ψ_i^1 to I_{C_i}. Since C_i is part of the premise of ψ_i^1, we must have $h(y_i) = c_{y_i}$ for $1 \leq i \leq m$. Since h must map A onto $Q(c_{y_1}, \ldots, c_{y_m})$, 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 ψ_i^1 does not fire on I_{C_i}. We now show that no other member of $\Sigma_{21}^{\mathbf{v}'}$ fires on I_{C_i}. If some member $\psi_j^{v'_j}$ were to fire on I_{C_i} where $j \neq i$, then because of the inequalities in $\psi_j^{v'_j}$, it would follow that some (and in fact, every) member of chase$_{12}(I_{B_j})$ is of the form $Q(c_1, \ldots, c_m)$, where $c_k = c_\ell$ if and only if $y_k = y_\ell$ (intuitively, each member of the prefix of $\psi_j^{v'_j}$ is a Q-atom with the same equality pattern of variables as C_i). So there would be a homomorphism $I_{C_i} \to$ chase$_{12}(I_{B_j})$. Since C_i is demanding for B_i, it follows that $I_{B_i} \subseteq I_{B_j}$. But this is impossible, since $i \neq j$. So no member $\psi_j^{v'_j}$ fires on I_{C_i} where $j \neq i$. A similar argument shows that no σ_P fires on I_{C_i}. So no member of $\Sigma_{21}^{\mathbf{v}'}$ fires on I_{C_i}, and hence $(I_{C_i}, \emptyset)$ satisfies $\Sigma_{21}^{\mathbf{v}'}$, as desired.

It is an 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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is a full s-t tgd mapping with a unique normal inverse $\mathcal{M}_{21} = (\mathbf{S_2}, \widehat{\mathbf{S_1}}, \Sigma_{21})$, and $\mathcal{M}_{21}$ is not equivalent to the constants-added version of a near p-copy mapping. Assume without loss of generality that Σ_{21} has the minimal number of constraints among the various sets of normal constraints logically equivalent to Σ_{21}, and each constraint σ in Σ_{21} has the minimal size among normal constraints logically equivalent to σ. 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 $\widehat{P}$ is in the conclusion of no member of Σ_{21}, then chase$_{21}$(chase$_{12}(I_{A_P})$) contains no $\widehat{P}$-fact, so Σ_{21} is too weak. Therefore, $\widehat{P}$ is in the conclusion of some member of Σ_{21}. Since also Σ_{21} has at most k constraints, it follows that $\widehat{P}$ is in the conclusion of exactly one member of Σ_{21}. Let σ_P be the member of Σ_{21} whose conclusion is a $\widehat{P}$-atom. Every variable in the conclusion of σ_P is distinct, or else chase$_{21}$(chase$_{12}(I_{A_P})$) does not contain $\widehat{I_{A_P}}$, so Σ_{21} is too weak. Let B_P be a P-atom with all variables the same. Now σ_P has no inequalities, or else chase$_{21}$(chase$_{12}(I_{B_P})$) does not contain $\widehat{I_{B_P}}$, so Σ_{21} is too weak.

The proof of Theorem 10.2 shows that A_P is good, that is, that chase$_{12}(I_{A_P})$ is a singleton. This singleton is the only relevant atom (with respect to Σ_{12}) for A_P, so it follows fairly easily from part (2) of Theorem 5.13 (and the assumption that constraints in Σ_{21} are of minimal size) that the premise of σ_P contains (along with const formulas) only a single relational atom. This relevant atom is also essential for P, as noted in Theorem 5.13. So the premise of σ_P has (along with const formulas) a single relational atom, that is essential for the conclusion of σ_P. Note that this is also true about each weak renaming of σ_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 $\mathbf{S_2}$. We now show that at most one member of Σ_{21} can have Q appear in its premise. Assume that σ_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 essential for both B_P and $B_{P'}$ (this follows from our earlier comment about weak renamings of σ_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 σ_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 [2007] showed that the unique-solutions property is a necessary condition for a schema mapping to have an inverse. Fagin [2007] 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 chase$_{12}(I) \leftrightarrow$ chase$_{12}(I')$. The reason that the unique-solutions property is equivalent to its homomorphic version is because of the fact, shown in Fagin et al. [2005], 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 [FAGIN 2007]. *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 $\mathcal{M}_{12} = (\mathbf{S_1}, \mathbf{S_2}, \Sigma_{12})$ is a LAV mapping that satisfies the unique-solutions property. We now show that $\mathcal{M}_{12}$ 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 $\mathcal{M}_{12}$ 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 $\mathcal{M}_{12}$ satisfies the homomorphic version of the subset property, and so is invertible, as desired. □

12. *Concluding Remarks and Open Problems*

In addition to resolving the key problem left open in Fagin [2007] 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

ARENAS, M., PÉREZ, J., AND RIVEROS, C. 2009. The recovery of a schema mapping: Bringing exchanged data back. *ACM Trans. Datab. Syst. 34*, 4.

BERNSTEIN, P. A. 2003. Applying model management to classical meta-data problems. In *Proceedings of the Conference on Innovative Data Systems Research*. 209–220.

BERNSTEIN, P. A., AND MELNIK, S. 2007. Model management 2.0: Manipulating richer mappings. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*. 1–12.

FAGIN, R. 1968. Representation theory for a class of denumerable markov chains. *J. Math. Anal. Appl. 23*, 500–530.

FAGIN, R. 2007. Inverting schema mappings. *ACM Trans. Datab. Syst. 32*, 4.

FAGIN, R., KOLAITIS, P. G., MILLER, R. J., AND POPA, L. 2005a. Data exchange: Semantics and query answering. *Theoret. Comput. Sci. 336*, 1, 89–124.

FAGIN, R., KOLAITIS, P. G., POPA, L., AND TAN, W.-C. 2005b. Composing schema mappings: Second-order dependencies to the rescue. *ACM Trans. Datab. Syst. 30*, 4, 994–1055.

FAGIN, R., KOLAITIS, P. G., POPA, L., AND TAN, W.-C. 2008. Quasi-inverses of schema mappings. *ACM Trans. Datab. Syst. 33*, 2.

FAGIN, R., KOLAITIS, P. G., POPA, L., AND TAN, W.-C. 2009. Reverse data exchange: Coping with nulls. In *Proceedings of the ACM Symposium on Principles of Database Systems*. 23–32.

KOLAITIS, P. G. 2005. Schema mappings, data exchange, and metadata management. In *Proceedings of the ACM Symposium on Principles of Database Systems*. 61–75.

LENZERINI, M. 2002. Data integration: A theoretical perspective. In *Proceedings of the ACM Symposium on Principles of Database Systems*. 233–246.

MADHAVAN, J., AND HALEVY, A. Y. 2003. Composing mappings among data sources. In *Proceedings of the International Conference on Very Large Data Bases*. 572–583.

MAIER, D., MENDELZON, A. O., AND SAGIV, Y. 1979. Testing implications of data dependencies. *ACM Trans. Datab. Syst. 4*, 4, 455–469.

MELNIK, S. 2004. *Generic Model Management: Concepts and Algorithms*. Lecture Notes in Computer Science, vol. 2967. Springer.

NASH, A., BERNSTEIN, P. A., AND MELNIK, S. 2007. Composition of mappings given by embedded dependencies. *ACM Trans. Datab. Syst. 32*, 1.

RECEIVED APRIL 2008; REVISED DECEMBER 2009 AND JUNE 2010; ACCEPTED JUNE 2010