International Journal of Foundations of Computer Science
Vol. 19, No. 3 (2008) 691–715

World Scientific
www.worldscientific.com

ON THE IMPLICATION OF MULTIVALUED DEPENDENCIES IN PARTIAL DATABASE RELATIONS*

SEBASTIAN LINK†
Information Science Research Centre
Massey University, New Zealand
s.link@massey.ac.nz

Received 1 March 2007
Accepted 11 November 2007
Communicated by Oscar H. Ibarra

The implication of multivalued dependencies (MVDs) in relational databases has originally and independently been defined in the context of some fixed finite universe by Delobel, Fagin, and Zaniolo. Biskup observed that the original axiomatisation for MVD implication does not reflect the fact that the complementation rule is merely a means to achieve database normalisation. He proposed two alternative ways to overcome this deficiency: i) an axiomatisation that does represent the role of the complementation rule adequately, and ii) a notion of MVD implication in which the underlying universe of attributes is left undetermined together with an axiomatisation of this notion.

In this paper we investigate multivalued dependencies with null values (NMVDs) as defined and axiomatised by Lien. We show that Lien's axiomatisation does not adequately reflect the role of the complementation rule, and extend Biskup's findings for MVDs in total database relations to NMVDs in partial database relations. Moreover, a correspondence between (minimal) axiomatisations in fixed universes that do reflect the property of complementation and (minimal) axiomatisations in undetermined universes is shown.

Keywords: Multivalued dependency; null value; implication; inference; minimality.

2000 Mathematics Subject Classification: 68P15

1. Introduction

Relational databases still form the core of most database management systems, even after more than three decades following their introduction (Codd 1970). The relational model organises data into a collection of relations. These structures permit the storage of inconsistent data, inconsistent in the semantic sense. Since this is not acceptable additional assertions, called dependencies, are formulated that every legal database is compelled to obey. There are many different classes of dependencies which can be utilised for improving the representation of the target

*This is an extended version of two conference papers presented at CATS 2006

†This research is supported by the Marsden Fund Council from Government Funding, administered by the Royal Society of New Zealand.

database. Excellent surveys on relational dependencies can be found in (Fagin & Vardi 1986, Thalheim 1991).

Multivalued dependencies (MVDs) (Delobel 1978, Fagin 1977, Zaniolo 1976) are an important class of dependencies. A relation exhibits an MVD precisely when it is decomposable into two of its projections without loss of information (Fagin 1977). This property is fundamental to relational database design, in particular 4NF (Fagin 1977), and a lot of research has therefore been devoted to studying the behaviour of these dependencies. Recently, extensions of multivalued dependencies have been found very useful for various design problems in advanced data models such as the nested relational data model (Fischer, Saxton, Thomas & Van Gucht 1985), the Entity-Relationship model (Thalheim 2003), data models that support nested lists (Hartmann, Link & Schewe 2006*b*) and XML (Vincent & Liu 2003, Vincent, Liu & Liu 2003).

The classical notion of an MVD (Fagin 1977) is dependent on the underlying universe R. This dependence is reflected syntactically by the R-complementation rule which is part of the axiomatisation for MVD implication in fixed universes R, from now on referred to as R-implication (Beeri, Fagin & Howard 1977). The complementation rule $\mathcal{C}_R$ is special in the sense that it is the only inference rule which is dependent on R. This observation has led to further research (Mendelzon 1979, Biskup 1978, Biskup 1980, Hartmann & Link 2006*b*, Hartmann et al. 2006*b*, Hartmann, Link & Köhler 2007, Link 2006*a*, Link 2006*b*) on the complementation rule. In particular, Biskup proposed an alternative notion of semantic implication in which the underlying universe is left undetermined (Biskup 1980). Biskup shows that this notion can be captured syntactically by a sound and complete set of inference rules, denoted by $\mathfrak{B}$. If $\mathfrak{B}_C$ results from adding the R-complementation rule $\mathcal{C}_R$ to $\mathfrak{B}$, then $\mathfrak{B}_C$ is R-sound and R-complete for the R-implication of MVDs for all relation schemata R. Moreover, every inference of an MVD by $\mathfrak{B}_C$ can be turned into an inference of the same MVD in which the R-complementation rule $\mathcal{C}_R$ is applied at most once, and if it is applied, then in the last step of the inference ($\mathfrak{B}_C$ is said to be R-complementary). This indicates that the R-complementation rule simply represents a part of the schema normalisation process, and does not necessarily infer semantically meaningful consequences. The R-complementary system $\mathfrak{B}_C$ reflects this property indicating that the underlying universe can be fixed in the very last step of an inference. Based on a common set of inference rules (Link 2006*a*) identifies all complementary axiomatisations for MVDs in fixed universes as well as all axiomatisations for MVDs in undetermined universes. Moreover, the time-complexity of the associated implication problem in undetermined universes is also studied. Apart from this work research has not been continued in this direction but focused on the original notion of R-implication. Since research on MVDs seems to experience a recent revival in the context of other data models (Fischer et al. 1985, Thalheim 2003, Hartmann & Link 2006*b*, Hartmann et al. 2006*b*, Hartmann et al. 2007, Link 2006*a*, Link 2006*b*, Vincent & Liu 2003, Vincent et al. 2003) it seems desirable to further extend the knowledge on relational MVDs. An advancement of such knowledge may also simplify the quest of finding suitable and comprehensible extensions of MVDs to currently popular data models. It is very rare in practice that the information in a database is complete. This observation has led to many extensions of the relational data model (Codd 1979, Lien 1982, Atzeni & Morfuni 1986, Levene & Loizou 1993, Levene & Loizou 1998, Johnson & Rosebrugh 2003) that can handle incomplete information. In particular, multivalued dependencies in the presence of null values (NMVDs) have been studied (Lien 1982). The original notion of an NMVD (Lien 1982) is again dependent on the underlying set R of attributes. This dependence is reflected syntactically by the R-complementation rule $\mathcal{C}_R$ which is part of the axiomatisation of NMVDs (Lien 1982). To the author's best knowledge no efforts have been made so far to study the role of the R-complementation rule for the implication of NMVDs. Moreover, the implication of NMVDs has not been investigated in undetermined universes.

Contributions and Organisation. In this article we investigate the role of the complementation rule $\mathcal{C}_R$ for MVD implication in the presence of null values. Section 2 repeats fundamental notions from the relational model of data and its extension to encompass incomplete information. In particular, the notions of implication for multivalued dependencies in fixed and undetermined universes are summarised, and sets of inference rules that capture these notions are repeated. It is demonstrated that Lien's axiomatisation $\mathfrak{K}$ for NMVDs is not R-complementary on all relation schemata R. Moreover, a notion of NMVD implication in undetermined universes is introduced. In Section 3 we propose a complementary axiomatisation $\mathcal{L}_C$ for NMVDs in fixed universes. Using this axiomatisation the underlying universe does not need to be fixed until the very last step of an inference, if at all. Subsequently, the notion of NMVD implication in undetermined universes is explored in Section 4. We show that this notion can be captured by a set $\mathcal{L}$ of inference rules where $\mathcal{L}_C = \mathcal{L} \cup \{\mathcal{C}_R\}$. Based on a set of common sound inference rules we prove in Section 5 that there is no other minimal axiomatisation for NMVD implication in undetermined universes. In Section 6 we extend the complementary axiomatisation from NMVDs to encompass functional dependencies (NFDs) as well. In sharp contrast to total database relations, NFDs and NMVDs can be dealt with separately, even if they are specified together. In Section 7 we show a general result proving that complementary axiomatiations in fixed universes result from axiomatisations in undetermined universes, and vice versa. As a corollary, $\mathcal{L}_C$ is the only minimal complementary axiomatisation based on those inference rules that we consider. Finally, we weaken the inference rules from the minimal axiomatisation $\mathcal{L}$ while still maintaining completeness in Section 8. This shows that the notion of minimality is not formalised in the strongest sense possible. We conclude the article in Section 9.

The problems studied in this article are not just of theoretical interest. In practice one does not necessarily want to generate all consequences of a given set of

(N)MVDs but only some of them. Such a task can be accomplished by using incomplete sets of inference rules. However, it is then essential to explore the power of such incomplete sets. Moreover, the notion of (N)MVD implication in undetermined universes may prove very useful in the context of views and distributed databases in which it appears to be rather difficult to speak about complements. Moreover, the long open problem of extending the synthesis approach from functional dependencies (Biskup, Dayal & Bernstein 1979) to MVDs may be tackled using this notion of MVD implication.

2. Multivalued Dependencies in Partial Relations

We use this section to introduce some notation and repeat notions and results for multivalued dependencies in the absence and presence of null values. Subsequently, we observe that the single existing axiomatisation for NMVDs is not adequate, and propose two alternative ways to overcome this insufficiency.

2.1. *Total and partial relations*

Let $\mathfrak{A} = \{A_1, A_2, \ldots\}$ be a (countably) infinite set of attributes. A *relation schema* is a finite set $R = \{A_1, \ldots, A_n\}$ of distinct symbols, called *attributes*, which represent column names of a relation. Each attribute A_i of a relation schema is associated an infinite domain $dom(A_i)$ which represents the set of possible values that can occur in the column named A_i. In order to encompass incomplete information it is assumed that every attribute may have a null value, denoted by $\nu \in dom(A_i)$. It may be noted that many kinds of null values have been proposed; for example, "missing" or "value unknown at present" (Codd 1975, Grant 1977, Grahne 1984), "non-existence" (Mikinouchi 1977), "inapplicable" (Grant 1977), "no information" (Zaniolo 1984) and "open" (Gottlob & Zicari 1988). The intention of the null value ν is to mean "no information". That is, the null value ν associated with an attribute in a tuple means that no information is available about that attribute for that tuple. This is the most primitive interpretation but can be used to model every kind of missing or incomplete information, and its semantics is certainly simple and well understood.

If X and Y are sets of attributes, then we may write XY for $X \cup Y$. If $X = \{A_1, \ldots, A_m\}$, then we may write $A_1 \cdots A_m$ for X. In particular, we may write simply A to represent the singleton $\{A\}$. A *tuple* over $R = \{A_1, \ldots, A_n\}$ (R-tuple or simply tuple, if R is understood) is a function $t : R \to \bigcup_{i=1}^{n} dom(A_i)$ with $t(A_i) \in dom(A_i)$ for $i = 1, \ldots, n$. For $X \subseteq R$ let $t[X]$ denote the restriction of the tuple t over R on X, and $dom(X) = \prod_{A \in X} dom(A)$ the Cartesian product of the domains of attributes in X. A *relation* r over R is a finite set of tuples over R. The relation schema R is also called the domain $Dom(r)$ of the relation r over R. Suppose that t_1, t_2 are two tuples in the relation r over R. It is said that t_1 *subsumes* t_2 if for every attribute $A \in R$, either $t_1[A] = t_2[A]$ or $t_2[A] = \nu$ holds. For the remainder of this article, the following restriction will be imposed on the relations in a database: No relation in the database shall contain two tuples t_1 and t_2 such that t_1 subsumes t_2. When no null value is present, this restriction amounts to saying that no two tuples are identical, an explicit requirement for database relations.

In order to contrast relations with and without null values, several terms are introduced. A relation r over R is said to be a *total relation* or simply a *relation* if it contains no null values. That is, if for any tuple $t \in r$ and any attribute $A \in R$, $t[A] \neq \nu$. If r is not a total relation, it is a *partial relation*. For a tuple $t \in R$ and a set $X \subseteq R$, t is said to be X-total if for any $A \in X$, $t[A] \neq \nu$.

There are several operations on partial relations that are natural generalisations of their counterparts from total relations. These include projection and natural join. Let r be some relation over R. Let X be some attribute set of R. The *projection* of r on X, denoted by $r[X]$, is a set of tuples t for which (i) there is some $t_1 \in r$ such that $t = t_1[X]$ and (ii) there is no $t_2 \in r$ such that $t_2[X]$ subsumes t and $t_2[X] \neq t$. Let Y be some attribute set of R with $Y \subseteq X$. The *Y-total projection* of r on X, denoted by $r_Y[X]$, is the set $r_Y[X] = \{t \in r[X] \mid t \text{ is } Y\text{-total}\}$. Given an X-total relation r over R and an X-total relation s over S such that $X = R \cap S$ the *natural join* of r and s, denoted by $r \bowtie s$, is the relation over $R \cup S$ which contains exactly those tuples t such that there is some $t_1 \in r$ and some $t_2 \in s$ with $t_1 = t[R]$ and $t_2 = t[S]$.

2.2. *Dependencies*

Functional dependencies (FDs) between sets of attributes have always played a central role in the study of relational databases (Codd 1970, Codd 1972, Beeri & Bernstein 1979, Bernstein 1976, Bernstein & Goodman 1980), and seem to be central for the study of database design in other data models as well (Arenas & Libkin 2004, Hara & Davidson 1999, Hartmann, Link & Schewe 2006*a*, Hartmann & Link 2006*a*, Levene & Loizou 1998, Tari, Stokes & Spaccapietra 1997, Weddell 1992, Wijsen 1999). In relational databases the notion of a functional dependency is well-understood and the semantic interaction between these dependencies has been syntactically captured by Armstrong's axioms (Armstrong 1974, Armstrong, Nakamura & Rudnicki 2002).

A *functional dependency* (FD) (Codd 1972) on the relation schema R is an expression $X \to Y$ where $X, Y \subseteq R$. A total relation r over R *satisfies* the FD $X \to Y$, denoted by $\models_r X \to Y$, if and only if every pair of tuples in r that agrees on each of the attributes in X also agrees on the attributes in Y. That is, $\models_r X \to Y$ if and only if $t_1[Y] = t_2[Y]$ whenever $t_1[X] = t_2[X]$ holds for any $t_1, t_2 \in r$. A *functional dependency with nulls* on R, abbreviated NFD, is a statement $X \rightharpoonup Y$ where $X, Y \subseteq R$. The NFD $X \rightharpoonup Y$ on R is satisfied by a partial relation r over R, denoted by $\models_r X \rightharpoonup Y$, if and only if for all $t_1, t_2 \in r$ the following holds: if t_1 and t_2 are X-total and $t_1[X] = t_2[X]$, then $t_1[Y] = t_2[Y]$. Therefore, whenever two tuples agree on a non null X-value, they agree on the Y-value, which may be partial. Recall that $\nu \in dom(A)$ for every attribute A.

Functional dependencies are incapable of modelling many important properties that database users have in mind. Multivalued dependencies (MVDs, (Delobel 1978, Fagin 1977, Zaniolo 1976)) provide a more general notion and offer a response to the shortcomings of FDs. MVDs have also been studied in the presence of null values (Lien 1982).

A *multivalued dependency* (MVD) (Delobel 1978, Fagin 1977, Zaniolo 1976) on R is an expression $X \twoheadrightarrow Y$ where $X, Y \subseteq R$. A total relation r over R *satisfies* the MVD $X \twoheadrightarrow Y$, denoted by $\models_r X \twoheadrightarrow Y$, if and only if for all $t_1, t_2 \in r$ with $t_1[X] = t_2[X]$ there is some $t \in r$ with $t[XY] = t_1[XY]$ and $t[X(R-Y)] = t_2[X(R-Y)]$. Informally, the relation r satisfies $X \twoheadrightarrow Y$ when the value on X determines the set of values on Y independently of the set of values on $R - Y$. This actually suggests that the relation schema R is overloaded in the sense that it carries two independent facts XY and $X(R-Y)$. More precisely, it is shown in (Fagin 1977) that MVDs "provide a necessary and sufficient condition for a relation to be decomposable into two of its projections without loss of information (in the sense that the original relation is guaranteed to be the join of the two projections)". This means that $\models_r X \twoheadrightarrow Y$ if and only if $r = r[XY] \bowtie r[X(R-Y)]$. This characteristic of MVDs is fundamental to relational database design and 4NF (Fagin 1977). A lot of research has therefore been devoted to studying the behaviour of these dependencies.

A *multivalued dependency with nulls* on the relation schema R, abbreviated NMVD, is an expression $X \twoheadrightarrow Y$ where $X, Y \subseteq R$. A partial relation r over R *satisfies* the NMVD $X \twoheadrightarrow Y$ on R, denoted by $\models_r X \twoheadrightarrow Y$, if and only if for all $t_1, t_2 \in r$ the following holds: if t_1 and t_2 are X-total and $t_1[X] = t_2[X]$, then there is some $t \in r$ such that $t[XY] = t_1[XY]$ and $t[X(R-Y)] = t_2[X(R-Y)]$. Informally, the partial relation r satisfies $X \twoheadrightarrow Y$ when the total X-values determine the set of values on Y independently of the set of values on $R - Y$. It has been shown that NMVDs provide a necessary and sufficient condition for a X-total relation to be decomposable into two of its projections without loss of information (in the sense that the original X-total relation is guaranteed to be the natural join of the two projections) (Lien 1982). This means that $\models_r X \twoheadrightarrow Y$ if and only if $r_X[R] = r_X[XY] \bowtie r_X[X(R-Y)]$. Recently, extensions of multivalued dependencies have been found very useful for various design problems in advanced data models such as the nested relational data model (Fischer et al. 1985), the Entity-Relationship model (Thalheim 2003), data models that support nested lists (Hartmann et al. 2006*b*) and XML (Vincent & Liu 2003, Vincent et al. 2003).

2.3. *Inferences of MVDs and NMVDs in fixed Universes*

For the design of a relational database schema dependencies are normally specified as semantic constraints on the relations which are intended to be instances of the schema. During the design process one usually needs to determine further dependencies which are logically implied by the given ones. In order to emphasise the dependence of implication from the underlying relation schema R we refer to R-implication. In the following we define R-implication for two classes of dependencies, i.e., for MVDs and for NMVDs.

Definition 1. *Let R be a relation schema, and let $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ and $X \twoheadrightarrow Y$ be (N)MVDs on R, i.e., $X \cup Y \cup \bigcup_{i=1}^{k}(X_i \cup Y_i) \subseteq R$. Then Σ R-implies $X \twoheadrightarrow Y$ if and only if each (partial) relation r over R that satisfies all (N)MVDs in Σ also satisfies the (N)MVD $X \twoheadrightarrow Y$.*

In order to determine logical consequences of a finite set of (N)MVDs one can use the following inference rules. These inference rules have the form

$$\frac{\text{premise}}{\text{conclusion}}$$

and inference rules without a premise are called axioms. Note that we use the natural complementation rule (Biskup 1978) instead of the complementation rule that was originally proposed (Beeri et al. 1977).

$\dfrac{}{X \twoheadrightarrow A} A \in X$ (membership, $\mathcal{M}$)	$\dfrac{}{X \twoheadrightarrow Y} Y \subseteq X$ (reflexivity, $\mathcal{R}$)	$\dfrac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV} V \subseteq U$ (augmentation, $\mathcal{A}$)
$\dfrac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$ (pseudo-transitivity, $\mathcal{T}$)	$\dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}$ (R-complementation, $\mathcal{C}_R$)	$\dfrac{}{\emptyset \twoheadrightarrow R}$ (R-axiom, $\mathcal{C}.1$)
$\dfrac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ}$ (union, $\mathcal{U}$)	$\dfrac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$ (difference, $\mathcal{D}$)	$\dfrac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}$ (intersection, $\mathcal{I}$)

Let R be some arbitrary relation schema. The set

$$\mathfrak{F} = \{\mathcal{R}, \mathcal{A}, \mathcal{T}, \mathcal{C}_R, \mathcal{U}, \mathcal{D}, \mathcal{I}\}$$

is both R-sound and R-complete for the R-implication of MVDs, for each relation schema R (Beeri et al. 1977). Let $\mathfrak{S}$ denote some set of inference rules, and let $\Sigma \cup \{\sigma\}$ be a set of (N)MVDs on the relation schema R. Let $\Sigma \vdash_{\mathfrak{S}} \sigma$ denote the inference of σ from Σ with respect to $\mathfrak{S}$. Let $\Sigma^+_{\mathfrak{S}} = \{\sigma \mid \Sigma \vdash_{\mathfrak{S}} \sigma\}$ denote the *syntactic closure* of Σ under inference using only rules from $\mathfrak{S}$. An inference rule is called *R-sound* if the set of dependencies in the premise of the rule R-implies the dependency in the conclusion. The set $\mathfrak{S}$ is called *R-sound* for the R-implication of (N)MVDs if and only if for every set Σ of (N)MVDs on the relation schema R we have $\Sigma^+_{\mathfrak{S}} \subseteq \Sigma^*_R = \{\sigma \mid \Sigma\ R\text{-implies } \sigma\}$. The set $\mathfrak{S}$ is called *R-complete* for the R-implication of (N)MVDs if and only if for every set Σ of (N)MVDs on R we have $\Sigma^*_R \subseteq \Sigma^+_{\mathfrak{S}}$.

An interesting question is now whether all the rules of a certain set are really necessary to capture the R-implication of (N)MVDs for every R. More precisely, an inference rule $\mathfrak{R}$ is said to be *independent* of $\mathfrak{S}$ if and only if there is some relation schema R and some finite set $\Sigma \cup \{\sigma\}$ of (N)MVDs on R such that $\sigma \notin \Sigma^+_{\mathfrak{S}}$, but $\sigma \in \Sigma^+_{\mathfrak{S}\cup\{\mathfrak{R}\}}$. Finally, let $\mathfrak{S}$ denote a set of inference rules that is R-complete for the R-implication of (N)MVDs for all relation schemata R. Then $\mathfrak{S}$ is said to be *minimal* for the R-implication of (N)MVDs if and only if every inference rule $\mathfrak{R} \in \mathfrak{S}$ is independent of $\mathfrak{S} - \{\mathfrak{R}\}$. This means that no proper subset of $\mathfrak{S}$ is still R-complete for all R. Mendelzon (Mendelzon 1979) shows that $\{\mathcal{C}_R, \mathcal{R}, \mathcal{T}\}$ is minimal for the R-implication of MVDs. Moreover, Biskup (Biskup 1978) shows that $\{\mathcal{C}.1, \mathcal{A}, \mathcal{T}\}$ is also minimal for the R-implication of MVDs, and Hartmann and Link (Hartmann & Link 2006*b*) show that $\{\mathcal{C}.1, \mathcal{M}, \mathcal{T}\}$ together with exactly one element of $\{\mathcal{U}, \mathcal{D}, \mathcal{I}\}$ form a minimal axiomatisation for the R-implication of MVDs.

In the presence of null values the pseudo-transitivity rule $\mathcal{T}$ is no longer R-sound for all relation schemata R (Lien 1982). However, Lien (Lien 1982) proves that $\mathfrak{K} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{C}_R\}$ is minimal for the R-implication of NMVDs. Apart from $\mathfrak{K}$ the sets $\mathfrak{K}_1 = \{\mathcal{R}, \mathcal{A}, \mathcal{I}, \mathcal{C}_R\}$ and $\mathfrak{K}_2 = \{\mathcal{R}, \mathcal{A}, \mathcal{D}, \mathcal{C}_R\}$ are also minimal for the R-implication of NMVDs. This fact was not noticed in (Lien 1982), but is an easy consequence of the DeMorgan rules.

2.4. *MVDs in Undetermined Universes*

Consider the classical example (Fagin 1977) in which the MVD *Employee* $\twoheadrightarrow$ *Child* is specified, i.e., the set of children is completely determined by an employee, independently of the rest of the information in any schema. If the relation schema R consists of the attributes *Employee*, *Child* and *Salary*, then we may infer the MVD *Employee* $\twoheadrightarrow$ *Salary* by means of the complementation rule. However, if the underlying relation schema R consists of the four attributes *Employee*, *Child*, *Salary* and *Year*, then the MVD *Employee* $\twoheadrightarrow$ *Salary* is no longer R-implied. Note the fundamental difference of the MVDs

$$\textit{Employee} \twoheadrightarrow \textit{Child} \qquad \text{and} \qquad \textit{Employee} \twoheadrightarrow \textit{Salary}.$$

The first MVD has been specified to establish the relationship of employees and their children as a fact due to a set-valued correspondence. The second MVD does not necessarily correspond to any semantic information, but simply results from the context in which *Employee* and *Child* are considered. If the context changes, the MVD disappears.

It may therefore be argued that consequences which are dependent on the underlying relation schema are in fact no consequences. This implies, however, that the notion of R-implication is not suitable. Biskup introduced the following notion of implication (Biskup 1980). An MVD is a syntactic expression $X \twoheadrightarrow Y$ with finite $X, Y \subseteq \mathfrak{A}$. The MVD $X \twoheadrightarrow Y$ is satisfied by some relation r if and only if $X \cup Y \subseteq Dom(r)$ and $r = r[XY] \bowtie r[X \cup (Dom(r) - Y)]$.

Definition 2. *The set* $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ *of MVDs implies the single MVD* $X \twoheadrightarrow Y$ *if and only if for each relation* r *with* $X \cup Y \cup \bigcup_{i=1}^{k} (X_i \cup Y_i) \subseteq Dom(r)$ *the MVD* $X \twoheadrightarrow Y$ *is satisfied by* r *whenever* r *already satisfies all MVDs in* Σ.

In this definition, the underlying relation schema is left undetermined. The only requirement is that the MVDs must apply to the relations. If $X \cup Y \cup \bigcup_{i=1}^{k} (X_i \cup Y_i) \subseteq R$, then it follows immediately that $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ R-implies $X \twoheadrightarrow Y$ whenever Σ implies $X \twoheadrightarrow Y$. The converse, however, is false (Biskup 1980).

The notions of soundness and completeness with respect to the notion of implication from Definition 2 are simply adapted from the corresponding notions in the context of fixed universes by dropping the reference to the underlying relation schema R.

While the singletons $\mathcal{M}, \mathcal{R}, \mathcal{A}, \mathcal{T}, \mathcal{U}, \mathcal{D}, \mathcal{I}$ are all sound, the R-complementation rule and R-axiom are R-sound, but not sound (Biskup 1980). In fact, the main result of (Biskup 1980) shows that the following set $\mathfrak{B}$ (denoted by $\mathfrak{S}_0$ in (Biskup 1980)) of inference rules

$$\frac{}{\emptyset \twoheadrightarrow \emptyset} \qquad \frac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\, V \subseteq U \qquad \frac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$

(empty-set-axiom, $\mathcal{R}_\emptyset$) (augmentation, $\mathcal{A}$) (pseudo-transitivity, $\mathcal{T}$)

$$\frac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow YZ} \qquad \frac{X \twoheadrightarrow Y, W \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}\, Y \cap W = \emptyset$$

(additive transitivity, $\mathcal{T}^*$) (subset, $\mathcal{S}$)

is sound and complete for the implication of MVDs. The major proof argument shows that the set $\mathfrak{B}_C = \{\mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{S}, \mathcal{C}_R\}$ is R-complementary for the R-implication of MVDs for all relation schemata R. An R-complete set $\mathfrak{S}$ of inference rules is said to be *R-complementary* for the R-implication of (N)MVDs if and only if for every set $\Sigma \cup \{\sigma\}$ of (N)MVDs on R the inference of σ from Σ using $\mathfrak{S}$ can be turned into an inference of σ from Σ using $\mathfrak{S}$ in which the R-complementation rule $\mathcal{C}_R$ is applied at most once, and if it is applied, then it is applied in the last step of the inference. This clarifies the role of $\mathcal{C}_R$ as a mere means of database normalisation. More precisely, for a set Σ of MVDs on the relation schema R the set $Dep_{\mathfrak{B}_C}(X) = \{Y \mid \Sigma \models X \twoheadrightarrow Y\}$ can be syntactically determined in two successive stages: i) determine the set $Dep_{\mathfrak{B}}(X) = \{Y \mid \Sigma \vdash_{\mathfrak{B}} X \twoheadrightarrow Y\}$ using $\mathfrak{B}$, and then add the complements $R - Y$ to $Dep_{\mathfrak{B}}(X)$ for all $Y \in Dep_{\mathfrak{B}}(X)$. This shows that

$$X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{B}_C} \quad \text{iff} \quad X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{B}} \text{ or } X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{B}}$$

where $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ and $X \cup Y \cup \bigcup_{i=1}^{k} (X_i \cup Y_i) \subseteq R$. We would now like to extend these findings to encompass incomplete information.

2.5. *The problem definition for NMVDs*

Recall that $\mathfrak{R} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{C}_R\}$ is R-sound and R-complete for the R-implication of NMVDs for all relation schemata R. The question is whether $\mathfrak{R}$ is also R-complementary for all R, i.e. whether $\mathfrak{R}$ allows one to fix the underlying universe within the very last step of an inference. The following example shows that this is not the case.

Example 3. *Let $\Sigma = \{A \twoheadrightarrow BC, A \twoheadrightarrow B\}$. The following table represents the syntactic closure $\Sigma^+_{\{\mathcal{R},\mathcal{A},\mathcal{U}\}}$ of Σ under inferences using $\{\mathcal{R}, \mathcal{A}, \mathcal{U}\}$. More precisely, an NMVD $X \twoheadrightarrow Y$ is in $\Sigma^+_{\{\mathcal{R},\mathcal{A},\mathcal{U}\}}$ if and only if there is a cross $\times$ in the row labelled X and column labelled Y. The closure can be obtained as follows. First, the NMVDs from Σ are entered followed by applications of the reflexivity axiom $\mathcal{R}$. Subsequently, the union rule $\mathcal{U}$ is applied twice to gain the A-row below. Finally, several applications of the augmentation rule $\mathcal{A}$ result in the table below. Further applications of any of the inference rules in $\{\mathcal{R}, \mathcal{A}, \mathcal{U}\}$ do not result in any new NMVDs.*

	$\emptyset$	A	B	C	AB	AC	BC	ABC
$\emptyset$	×							
A	×	×	×		×		×	×
B	×		×					
C	×			×				
AB	×	×	×		×		×	×
AC	×	×	×	×	×	×	×	×
BC	×		×	×			×	
ABC	×	×	×	×	×	×	×	×

It shows in particular that $A \twoheadrightarrow C \notin \Sigma^+_{\{\mathcal{R},\mathcal{A},\mathcal{U}\}}$. Moreover, Lemma 10 shows that $A \twoheadrightarrow Y \notin \Sigma^+_{\{\mathcal{R},\mathcal{A},\mathcal{U}\}}$ for all Y such that $Y - \{A, B, C\} \neq \emptyset$. However, for $R = \{A, B, C, D\}$ we have $A \twoheadrightarrow C \in \Sigma^+_{\mathfrak{R}}$, say by

$$\mathcal{C}_R: \cfrac{\mathcal{U}: \cfrac{\mathcal{C}_R: \cfrac{A \twoheadrightarrow BC}{A \twoheadrightarrow AD} \qquad A \twoheadrightarrow B}{A \twoheadrightarrow ABD}}{A \twoheadrightarrow C} .$$

Hence, in any such inference the rule $\mathcal{C}_R$ must be used at least once, but since $R - \{C\} = \{A, B, D\}$ it is not only used as the last rule.

Example 3 brings up the question whether there is any axiomatisation for the R-implication of NMVDs that is also R-complementary for all relation schemata R. We will give an affirmative answer to this question in Section 3.

As it was the case in the absence of null values it may be argued that consequences that depend on the underlying universe are in fact not consequences at all.

Consequently, the notion of R-implication is not suitable anymore. We will therefore generalise the notion of MVDs in undetermined universes to the presence of null values. An NMVD is a syntactic expression $X \twoheadrightarrow Y$ with finite $X, Y \subseteq \mathfrak{A}$. The NMVD $X \twoheadrightarrow Y$ is satisfied by some partial relation r if and only if $X \cup Y \subseteq Dom(r)$ and $r_X[Dom(r)] = r_X[XY] \bowtie r_X[X \cup (Dom(r) - Y)]$.

Definition 4. *The set* $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ *of NMVDs implies the single NMVD* $X \twoheadrightarrow Y$ *if and only if for each partial relation* r *with* $X \cup Y \cup \bigcup_{i=1}^{k} (X_i \cup Y_i) \subseteq Dom(r)$ *the NMVD* $X \twoheadrightarrow Y$ *is satisfied by* r *whenever* r *already satisfies all NMVDs in* Σ.

In this definition, the underlying relation schema is left undetermined. The only requirement is that the NMVDs must apply to the partial relations. The following fact is immediate and generalises a result from (Biskup 1980).

Theorem 5. *Let* $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ *be a set of NMVDs, and* $X \cup Y \cup \bigcup_{i=1}^{k} (X_i \cup Y_i) \subseteq R$. *If* Σ *implies* $X \twoheadrightarrow Y$, *then* Σ *R-implies* $X \twoheadrightarrow Y$.

The following example shows that the converse of Theorem 5 is false.

Example 6. *For* $R = \{Employee, Child, Salary\}$ *and* $\Sigma = \{Employee \twoheadrightarrow Child\}$ *we have that* Σ *R-implies Employee* $\twoheadrightarrow$ *Salary. However,* Σ *does not imply Employee* $\twoheadrightarrow$ *Salary. Consider for instance the following partial relation* r *with domain* $\{Employee, Child, Salary, Year\}$.

Employee	*Child*	*Salary*	*Year*
Don Juan	ν	*4000*	*2004*
Don Juan	ν	*5000*	*2005*

The two relations $r_{Employee}[Employee, Child]$ *and* $r_{Employee}[Employee, Salary, Year]$

Employee	*Child*
Don Juan	ν

Employee	*Salary*	*Year*
Don Juan	*4000*	*2004*
Don Juan	*5000*	*2005*

show that r *satisfies the NMVD Employee* $\twoheadrightarrow$ *Child. However, the two relations* $r_{Employee}[Employee, Salary]$ *and* $r_{Employee}[Employee, Child, Year]$

Employee	*Salary*
Don Juan	*4000*
Don Juan	*5000*

Employee	*Child*	*Year*
Don Juan	ν	*2004*
Don Juan	ν	*2005*

indicate that r *does not satisfy Employee* $\twoheadrightarrow$ *Salary. Consequently,* Σ *does not imply Employee* $\twoheadrightarrow$ *Salary.*

Notice that the singletons $\mathcal{M}, \mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}, \mathcal{I}$ are all sound, and the R-complementation rule $\mathcal{C}_R$ and R-axiom $\mathcal{C}.1$ are both R-sound, but not sound with respect to Definition 4.

It is our second objective to identify a set of inference rules that is sound and complete for the implication of NMVDs in undetermined universes. We will pursue this goal in Section 4.

3. NMVDs in Fixed Universes

The goal of this section is to identify a set $\mathfrak{L}_C$ of inference rules which is R-sound, R-complete and R-complementary for the R-implication of NMVDs for all relation schemata R. Biskup has successfully provided a solution to this problem for MVDs, i.e., in the absence of null values (Biskup 1980). One may hope that the inclusion of the additive transitivity rule $\mathcal{T}^*$ and/or subset rule $\mathcal{S}$ into $\mathfrak{K}$ result in a complementary axiomatisation for NMVDs. The following example demonstrates that neither of the rules is R-sound for all relation schemata R.

Example 7. *Consider the following partial relation r:*

A	B	C	D
a	b_1	c_1	ν
a	b_2	c_2	ν

For $X = A$, $Y = BC$, $W = D$ and $Z = B$ we see that $\models_r X \twoheadrightarrow Y$ and $\models_r W \twoheadrightarrow Z$ with $Y \cap W = \emptyset$. However, $\not\models_r X \twoheadrightarrow Y \cap Z$. Note that r satisfies $W \twoheadrightarrow Z$ since the two tuples are not total on W. This shows that the subset rule for NMVDs is not R-sound for $R = \{A, B, C, D\}$.

Similarly, the R-soundness of the additive transitivity rule fails as well. Consider the partial relation r:

A	B	C	D
a	ν	c_1	d_1
a	ν	c_2	d_2

For $X = A$, $Y = B$ and $Z = C$ we see that $\models_r X \twoheadrightarrow Y$ and $\models_r Y \twoheadrightarrow Z$. However, $\not\models_r X \twoheadrightarrow YZ$. Note that r satisfies $Y \twoheadrightarrow Z$ since the two tuples are not total on Y.

Our first theorem shows that there are indeed axiomatisations for NMVDs which are R-complementary for all relation schemata R. In order to be precise, we give the following definition. Let Σ be a finite set of NMVDs, and let $\mathfrak{S}$ be a set of inference rules. A finite sequence of NMVDs $\gamma = [\sigma_1, \ldots, \sigma_k]$ is called an *inference from Σ by $\mathfrak{S}$* if and only if each σ_i is either an element of Σ or is obtained by applying one of the rules of $\mathfrak{S}$ to appropriate elements of $\{\sigma_1, \ldots, \sigma_{i-1}\}$. We say that the inference γ infers σ_k (the last element of the sequence γ). The syntactic closure $\Sigma^+_{\mathfrak{S}}$ is the set of all NMVDs which can be inferred by some inference from Σ by $\mathfrak{S}$.

Theorem 8. *Let Σ be a set of NMVDs on the relation schema R. For each inference γ from Σ by the set $\mathfrak{K} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{C}_R\}$ there is an inference ξ from Σ by the set $\mathfrak{L}_C = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}, \mathcal{C}_R\}$ with the following properties:*

(1) γ and ξ infer the same NMVD.
(2) In ξ the R-complementation rule $\mathcal{C}_R$ is applied at most once.
(3) If $\mathcal{C}_R$ is applied in ξ, then $\mathcal{C}_R$ is applied as the last rule.

Proof. The proof is by induction on the length l of the inference $\gamma = [\sigma_1, \ldots, \sigma_l]$. If $l = 1$, then $\xi := \gamma$ has the desired properties. Let $l > 1$, and $\gamma = [\sigma_1, \ldots, \sigma_l]$ be an inference from Σ by $\mathfrak{K}$ which has length l. We consider four cases according to how σ_l was obtained from $[\sigma_1, \ldots, \sigma_{l-1}]$.

Case 1. σ_l is either an element of Σ or was obtained using the reflexivity axiom $\mathcal{R}$. Then $\xi = [\sigma_l]$ has the desired properties.

Case 2. We obtain σ_l by applying the augmentation rule $\mathcal{A}$ to the premise σ_i with $i < l$. Let ξ_i be obtained by using the induction hypothesis for $\gamma_i := [\sigma_1, \ldots, \sigma_i]$. Consider the inference $\xi := [\xi_i, \sigma_l]$. If in ξ_i the rule $\mathcal{C}_R$ is not applied, then ξ has the desired properties. If in ξ_i the rule $\mathcal{C}_R$ is applied (as last rule), then the last two steps of ξ are of the form:

$$\mathcal{A}: \cfrac{\mathcal{C}_R: \cfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}}{XU \twoheadrightarrow (R-Y)V} V \subseteq U.$$

However, these two steps can be replaced as follows:

$$\mathcal{C}_R: \cfrac{\mathcal{D}: \cfrac{\mathcal{A}: \cfrac{X \twoheadrightarrow Y}{XU \twoheadrightarrow Y} \qquad \mathcal{R}: \cfrac{}{XU \twoheadrightarrow V} V \subseteq U \subseteq XU}{XU \twoheadrightarrow Y - V}}{XU \twoheadrightarrow \underbrace{R - (Y - V)}_{=(R-Y)V}}$$

The result of this replacement is an inference with the desired properties.

Case 3. We obtain σ_l by applying the union rule $\mathcal{U}$ to the premises σ_i and σ_j with $i, j < l$. Let ξ_i and ξ_j be obtained by using the induction hypothesis for $\gamma_i = [\sigma_1, \ldots, \sigma_i]$ and $\gamma_j = [\sigma_1, \ldots, \sigma_j]$, respectively. Consider the inference $\xi := [\xi_i, \xi_j, \sigma_l]$. We distinguish four cases according to the occurrences of the R-complementation rule $\mathcal{C}_R$ in ξ_i and ξ_j.

Case 3.1. If $\mathcal{C}_R$ is applied neither in ξ_i nor in ξ_j, then ξ has the desired properties.

Case 3.2. If $\mathcal{C}_R$ is applied in ξ_i (as last rule), but not in ξ_j, then the last step of ξ_i and the last step of ξ are of the following form:

$$\mathcal{U}: \cfrac{\mathcal{C}_R: \cfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y} \qquad X \twoheadrightarrow Z}{X \twoheadrightarrow (R-Y)Z}.$$

However, these steps can be replaced as follows:

$$\mathcal{D}: \dfrac{X \twoheadrightarrow Z \qquad X \twoheadrightarrow Y}{\mathcal{C}_R: \dfrac{X \twoheadrightarrow Y - Z}{X \twoheadrightarrow \underbrace{R - (Y - Z)}_{=(R-Y)Z}}} .$$

The result of this replacement is an inference with the desired properties.

Case 3.3. If $\mathcal{C}_R$ is applied in ξ_j (as last rule), but not in ξ_i, then the last step of ξ_j and the last step of ξ are of the following form:

$$\mathcal{U}: \dfrac{X \twoheadrightarrow Y \qquad \mathcal{C}_R: \dfrac{X \twoheadrightarrow Z}{X \twoheadrightarrow R - Z}}{X \twoheadrightarrow Y(R - Z)} .$$

However, these steps can be replaced as follows:

$$\mathcal{D}: \dfrac{X \twoheadrightarrow Y \qquad X \twoheadrightarrow Z}{\mathcal{C}_R: \dfrac{X \twoheadrightarrow Z - Y}{X \twoheadrightarrow \underbrace{R - (Z - Y)}_{=Y(R-Z)}}} .$$

The result of this replacement is an inference with the desired properties.

Case 3.4. If $\mathcal{C}_R$ is applied both in ξ_i and ξ_j (as last rule), then the last steps of ξ_i and ξ_j and the last step of ξ are of the following form:

$$\mathcal{U}: \dfrac{\mathcal{C}_R: \dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y} \qquad \mathcal{C}_R: \dfrac{X \twoheadrightarrow Z}{X \twoheadrightarrow R - Z}}{X \twoheadrightarrow (R - Y) \cup (R - Z)} .$$

However, these steps can be replaced as follows:

$$\mathcal{D}: \dfrac{\mathcal{D}: \dfrac{X \twoheadrightarrow Y \qquad X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y} \qquad X \twoheadrightarrow Z}{\mathcal{C}_R: \dfrac{X \twoheadrightarrow \underbrace{Z - (Z - Y)}_{=Y \cap Z}}{X \twoheadrightarrow \underbrace{R - (Y \cap Z)}_{=(R-Y)\cup(R-Z)}}} .$$

The result of this replacement is an inference with the desired properties.

Case 4. We obtain σ_l by applying the R-complementation rule $\mathcal{C}_R$ to the premise σ_i with $i < l$. Let ξ be obtained by using the induction hypothesis for $\gamma_i := [\sigma_1, \ldots, \sigma_i]$. Consider the inference $\xi := [\xi_i, \sigma_i]$. If in ξ_i the rule $\mathcal{C}_R$ is not applied, then ξ has the desired properties. If in ξ_i the rule $\mathcal{C}_R$ is applied (as last rule), then the last two steps of ξ are of the following form:

$$\mathcal{C}_R: \dfrac{\mathcal{C}_R: \dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}}{X \twoheadrightarrow \underbrace{R - (R - Y)}_{=Y}} .$$

Hence, the inference obtained by removing these two steps from ξ has the desired properties. □

The set $\mathfrak{L}_C$ is R-complete for the R-implication of NMVDs since $\mathfrak{L}_C$ is an extension of the R-complete set $\mathfrak{K}$ (Lien 1982). While $\mathfrak{K}$ is minimal the set $\mathfrak{L}_C$ is not (the pseudo-difference rule $\mathcal{D}$ can be omitted). However, $\mathfrak{L}_C$ is R-complementary while $\mathfrak{K}$ is not. A reasonable question is whether there is any minimal set $\mathfrak{S}$ which is also R-complementary. This warrants future research.

4. NMVDs in Undetermined Universes

Now, our objective is to identify a set $\mathfrak{L}$ of inference rules which is sound and complete for the implication of NMVDs according to Definition 4. We therefore explore the power of the common part of the sets $\mathfrak{L}_C$, namely $\mathfrak{L} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}\}$, which can be obtained from any of the sets $\mathfrak{L}_C$ by removing the R-complementation rule $\mathcal{C}_R$. Hence, $\mathfrak{L}$ does not permit the possibly semantically meaningless inference of complementation.

Theorem 8 states that for all relation schemata R the set $\mathfrak{L} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}\}$ is nearly R-complete. More precisely, we can formulate the following corollary.

Corollary 9. *Let $R \subseteq \mathfrak{A}$ be a finite set of attributes. Then for all finite sets $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ of NMVDs, for all NMVDs $X \twoheadrightarrow Y$ such that $X \cup Y \cup \bigcup_{i=1}^{k} (X_i \cup Y_i) \subseteq R$ we have that*

$$X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{L}_C} \quad \textit{if and only if} \quad X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{L}} \textit{ or } X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{L}}.$$

Corollary 9 indicates that by the set $\mathfrak{L}$ we can infer those consequences of a given set of NMVDs which are independent of the underlying relation schema R.

We shall prove now that the set $\mathfrak{L}$ is actually sound and complete for the implication of NMVDs, in the sense of Definition 4, that is by $\mathfrak{L}$ we can generate exactly all implications in an undetermined universe. We shall prove two lemmata in preparation.

The correctness of the first lemma can easily be observed by inspecting the syntactic definitions of the inference rules in $\mathfrak{L}$. For each of the rules, the right-hand side of the conclusion does not contain any attribute that did not already occur in the right-hand side of at least one of the premises.

Lemma 10. *Let $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ be a finite set of NMVDs. If $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{L}}$, then $Y \subseteq X \cup \bigcup_{i=1}^{k} Y_i$.*

Proof. We show that if $\gamma = [\sigma_1, \ldots, \sigma_l]$ is an inference from Σ by $\mathfrak{L}$ such that γ infers the NMVD $\sigma_l = X \twoheadrightarrow Y$, then $Y \subseteq X \cup \bigcup_{i=1}^{k} Y_i$. The proof is by induction on the length l of γ. If $l = 1$, then $X \twoheadrightarrow Y$ was obtained either by application of the

reflexivity axiom, i.e. $Y \subseteq X$, or it is an element of Σ. Thus we have $Y \subseteq X \cup \bigcup_{i=1}^{k} Y_i$ in any case.
Let $l > 1$. We consider four cases according to how σ_l was obtained from $[\sigma_1, \ldots, \sigma_{l-1}]$.

Case 1. σ_l was obtained by application of the reflexivity axiom or it is an element of Σ. This is the same situation as for $l = 1$.

Case 2. σ_l was obtained by application of the augmentation rule $\mathcal{A}$ to the premise σ_i with $i < l$. Then the last step of γ has the form

$$\frac{R \twoheadrightarrow S}{RU \twoheadrightarrow SV} V \subseteq U$$

where $\sigma_i = R \twoheadrightarrow S$ and $S \subseteq R \cup \bigcup_{i=1}^{k} Y_i$ by induction hypothesis, and $\sigma_l = RU \twoheadrightarrow SV$. Consequently, we have

$$SV \subseteq RU \cup \bigcup_{i=1}^{k} Y_i \ .$$

Case 3. σ_l was obtained by application of the union rule $\mathcal{U}$ to the premises σ_i and σ_j with $i, j < l$. Then the last step of γ has the form

$$\frac{R \twoheadrightarrow S, R \twoheadrightarrow T}{R \twoheadrightarrow ST}$$

where $\sigma_i = R \twoheadrightarrow S$ and $S \subseteq R \cup \bigcup_{i=1}^{k} Y_i$ by induction hypothesis, $\sigma_j = R \twoheadrightarrow T$ and $T \subseteq R \cup \bigcup_{i=1}^{k} Y_i$ by induction hypothesis, and $\sigma_l = R \twoheadrightarrow ST$. Consequently, we have

$$ST \subseteq R \cup \bigcup_{i=1}^{k} Y_i \ .$$

Case 4. σ_l was obtained by application of the difference rule $\mathcal{D}$ to the premises σ_i and σ_j with $i, j < l$. Then the last step of γ has the form

$$\frac{R \twoheadrightarrow S, R \twoheadrightarrow T}{R \twoheadrightarrow T - S}$$

where $\sigma_i = R \twoheadrightarrow S$ and $S \subseteq R \cup \bigcup_{i=1}^{k} Y_i$ by induction hypothesis, $\sigma_j = R \twoheadrightarrow T$ and $T \subseteq R \cup \bigcup_{i=1}^{k} Y_i$ by induction hypothesis, and $\sigma_l = R \twoheadrightarrow T - S$. Consequently, we have

$$T - S \subseteq R \cup \bigcup_{i=1}^{k} Y_i \ .$$

This concludes the proof. □

For the next lemma one may notice that attributes outside of W can always be introduced only in the last step of the inference utilising the *augmentation rule* $\mathcal{A}$.

Lemma 11. *Let* $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ *be a finite set of NMVDs. Let* $W := \bigcup_{i=1}^{k} (X_i \cup Y_i)$*. If* $X \twoheadrightarrow Y \in \Sigma_{\mathfrak{L}}^{+}$*, then there is an inference* $\gamma = [\sigma_1, \ldots, \sigma_l]$ *of* $X \twoheadrightarrow Y$ *from* Σ *by* $\mathfrak{L}$ *such that any attribute occurring in* $\sigma_1, \ldots, \sigma_{l-1}$ *is an element of* W*.*

Proof. Let $\bar{\xi} = [R_1 \twoheadrightarrow S_1, \ldots, R_{l-1} \twoheadrightarrow S_{l-1}]$ be any inference of $X \twoheadrightarrow Y$ from Σ by $\mathfrak{L}$. Consider the sequence

$$\xi := [R_1 \cap W \twoheadrightarrow S_1 \cap W, \ldots, R_{l-1} \cap W \twoheadrightarrow S_{l-1} \cap W].$$

We claim that ξ is an inference of $X \cap W \twoheadrightarrow Y \cap W$ from Σ by $\mathfrak{L}$. For if $R_i \twoheadrightarrow S_i$ is an element of Σ or was obtained by application of the reflexivity axiom $\mathcal{R}$, then $R_i \cap W \twoheadrightarrow S_i \cap W = R_i \twoheadrightarrow S_i$. Moreover, one can verify that if $R_i \twoheadrightarrow S_i$ is the result of applying one of the rules $\mathcal{A}, \mathcal{U}, \mathcal{D}$ in $\bar{\xi}$, then $R_i \cap W \twoheadrightarrow S_i \cap W$ is the result of the same rule applied to the corresponding premises in ξ.

Now by Lemma 10 we know that $Y \subseteq X \cup \bigcup_{i=1}^{k} Y_i \subseteq X \cup W$, hence $Y - W \subseteq X$. However, this implies that we can infer $X \twoheadrightarrow Y$ from $X \cap W \twoheadrightarrow Y \cap W$ by the augmentation rule $\mathcal{A}$:

$$\frac{X \cap W \twoheadrightarrow Y \cap W}{\underbrace{(X \cap W) \cup X}_{=X} \twoheadrightarrow \underbrace{(Y \cap W) \cup (Y - W)}_{=Y}}\ .$$

Hence the inference $[\xi, X \twoheadrightarrow Y]$ has the desired properties. □

Theorem 12. *The set* $\mathfrak{L} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}\}$ *is sound and complete for the implication of multivalued dependencies with null values.*

Proof. Let $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ be a finite set of NMVDs, and let $X \twoheadrightarrow Y$ be an NMVD. We have to prove that

$$\Sigma \text{ implies } X \twoheadrightarrow Y \quad \text{if and only if} \quad X \twoheadrightarrow Y \in \Sigma_{\mathfrak{L}}^{+} \quad . \tag{1}$$

Let $T := X \cup Y \cup \bigcup_{i=1}^{k} (X_i \cup Y_i)$. In order to prove the soundness of $\mathfrak{L}$ (if-part of (1)) we assume that $X \twoheadrightarrow Y \in \Sigma_{\mathfrak{L}}^{+}$ holds. Let r be any partial relation such that $T \subseteq Dom(r)$ and such that r satisfies $X_i \twoheadrightarrow Y_i \in \Sigma$ for all $i = 1, \ldots, k$. We must show that r also satisfies $X \twoheadrightarrow Y$. According to Lemma 11 there is an inference γ of $X \twoheadrightarrow Y$ from Σ by $\mathfrak{L}$ such that $R \cup S \subseteq T \subseteq Dom(r)$ holds for each NMVD $R \twoheadrightarrow S$ occurring in γ. Since each rule of $\mathfrak{L}$ is sound we can conclude (by induction) that each NMVD occurring in γ is satisfied by r. Hence, r satisfies $X \twoheadrightarrow Y$ in particular.

In order to prove the completeness of $\mathfrak{L}$ (only if-part of (1)) we assume $X \twoheadrightarrow Y \notin \Sigma^+_{\mathfrak{L}}$. Let $R \subseteq \mathfrak{A}$ be a finite set of attributes such that T is a proper subset of R, that is $T \subset R$. Consequently, $R - Y$ is not a subset of T. Hence, by Lemma 10, $X \twoheadrightarrow (R - Y) \notin \Sigma^+_{\mathfrak{L}}$. Now from $X \twoheadrightarrow Y \notin \Sigma^+_{\mathfrak{L}}$ and $X \twoheadrightarrow (R - Y) \notin \Sigma^+_{\mathfrak{L}}$ we conclude that $X \twoheadrightarrow Y \notin \Sigma^+_{\mathfrak{L}_C}$ by Corollary 9. Since $\mathfrak{L}_C$ is R-complete for the R-implication of NMVDs it follows that Σ does not R-imply $X \twoheadrightarrow Y$. Hence, Σ does not imply $X \twoheadrightarrow Y$ by Theorem 5. □

5. All Axiomatisations of NMVDs in Undetermined Universes

One may ask whether there are any other subsets of $\{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I}, \mathcal{D}\}$ which are also complete for the implication of NMVDs.

Theorem 13. *$\mathfrak{L}$ is the only minimal complete subset of $\{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I}, \mathcal{D}\}$ which is sound and complete for the implication of NMVDs.*

Proof. The completeness of $\mathfrak{L}$ has been established in Theorem 12. That is, for all sets Σ of NMVDs the four inference rules of $\mathfrak{L}$ suffice to infer all the NMVDs that are implied by Σ. We will show now that the four inference rules in $\mathfrak{L}$ are also necessary for gaining a complete subset of $\{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I}, \mathcal{D}\}$. In fact, we will establish that each inference rule in $\mathcal{L}$ is independent of the remaining inference rules in $\{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I}, \mathcal{D}\}$. Hence, if a subset of $\{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I}, \mathcal{D}\}$ is complete then it contains $\mathcal{L}$. Consequently, $\mathcal{L}$ is the only minimal complete subset of $\{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I}, \mathcal{D}\}$.

- The reflexivity axiom $\mathcal{R}$ is independent of $\mathfrak{S} = \{\mathcal{A}, \mathcal{I}, \mathcal{D}, \mathcal{U}\}$. Let $\Sigma = \emptyset$, and $\sigma = \emptyset \twoheadrightarrow \emptyset$. Since $\sigma \notin \Sigma^+_{\mathfrak{S}}$, but $\sigma \in \Sigma^+_{\mathfrak{S} \cup \{\mathcal{R}\}}$ we have found witnesses Σ and σ for the independence of $\mathcal{R}$ from $\mathfrak{S}$.
- The augmentation rule $\mathcal{A}$ is independent of $\mathfrak{S} = \{\mathcal{R}, \mathcal{I}, \mathcal{D}, \mathcal{U}\}$. Let $\Sigma = \{A \twoheadrightarrow B\}$, and $\sigma = AC \twoheadrightarrow B$. The following table represents the closure $\Sigma^+_{\mathfrak{S}}$ of Σ under $\mathfrak{S}$ neglecting all remaining trivial NMVDs $X \twoheadrightarrow Y$ with $Y \subseteq X$ and $X, Y \subseteq \mathfrak{A}$. It can be obtained as follows. First, we enter $A \twoheadrightarrow B$ from Σ and then apply the reflexivity axiom $\mathcal{R}$. Finally, we can apply the union rule $\mathcal{U}$ once to obtain $A \twoheadrightarrow AB$ and the table below. Further inferences using $\mathfrak{S}$ do not result in any new NMVDs.

	$\emptyset$	A	B	C	AB	AC	BC	ABC
$\emptyset$	×							
A	×	×	×		×			
B	×		×					
C	×			×				
AB	×	×	×		×			
AC	×	×		×		×		
BC	×		×	×			×	
ABC	×	×	×	×	×	×	×	×

Since $\sigma \notin \Sigma^+_{\mathfrak{S}}$, but $\sigma \in \Sigma^+_{\mathfrak{S}\cup\{\mathcal{S}\}}$ we have found witnesses Σ and σ for the independence of $\mathcal{A}$ from $\mathfrak{S}$.

- The union rule $\mathcal{U}$ is independent of $\mathfrak{S} = \{\mathcal{R}, \mathcal{A}, \mathcal{I}, \mathcal{D}\}$. Let $\Sigma = \{A \twoheadrightarrow B, A \twoheadrightarrow C\}$, and $\sigma = A \twoheadrightarrow BC$. The following table represents the closure $\Sigma^+_{\mathfrak{S}}$ of Σ under $\mathfrak{S}$ neglecting all remaining trivial NMVDs $X \twoheadrightarrow Y$ with $Y \subseteq X$ and $X, Y \subseteq \mathfrak{A}$. It can be obtained as follows. First, we enter the NMVDs from Σ and apply the reflexivity axiom $\mathcal{R}$ as often as possible. The table below is now the result of several applications of the augmentation rule $\mathcal{A}$. Further inferences using $\mathfrak{S}$ do not result in any new NMVDs.

	$\emptyset$	A	B	C	AB	AC	BC	ABC
$\emptyset$	×							
A	×	×	×	×	×	×		
B	×		×					
C	×			×				
AB	×	×	×	×	×	×	×	×
AC	×	×	×	×	×	×	×	×
BC	×		×	×			×	
ABC	×	×	×	×	×	×	×	×

Since $\sigma \notin \Sigma^+_{\mathfrak{S}}$, but $\sigma \in \Sigma^+_{\mathfrak{S}\cup\{\mathcal{U}\}}$ we have found witnesses Σ and σ for the independence of $\mathcal{U}$ from $\mathfrak{S}$.

- The difference rule $\mathcal{D}$ is independent of $\mathfrak{S} = \{\mathcal{R}, \mathcal{A}, \mathcal{I}, \mathcal{U}\}$. Let $\Sigma = \{A \twoheadrightarrow BC, A \twoheadrightarrow B\}$, and $\sigma = A \twoheadrightarrow C$. The following table represents the closure $\Sigma^+_{\mathfrak{S}}$ of Σ under $\mathfrak{S}$ neglecting all remaining trivial NMVDs $X \twoheadrightarrow Y$ with $Y \subseteq X$ and $X, Y \subseteq \mathfrak{A}$. It can be obtained as follows. First, we enter the NMVDs from Σ and apply the reflexivity axiom $\mathcal{R}$ as often as possible. The table below is now the result of several applications of the augmentation rule $\mathcal{A}$. Further inferences using $\mathfrak{S}$ do not result in any new NMVDs.

	$\emptyset$	A	B	C	AB	AC	BC	ABC
$\emptyset$	×							
A	×	×	×		×		×	×
B	×		×					
C	×			×				
AB	×	×	×		×		×	×
AC	×	×	×	×	×	×	×	×
BC	×		×	×			×	
ABC	×	×	×	×	×	×	×	×

Since $\sigma \notin \Sigma^+_{\mathfrak{S}}$, but $\sigma \in \Sigma^+_{\mathfrak{S}\cup\{\mathcal{D}\}}$ we have found witnesses Σ and σ for the independence of $\mathcal{D}$ from $\mathfrak{S}$. □

6. NFDs and NMVDs in Undetermined Universes

We apply Theorem 12 and the results from (Lien 1982) to obtain an axiomatisation for NFDs and NMVDs in undetermined universes. Note that an axiomatisation of FDs and MVDs in undetermined universes has already been proposed (Biskup 1980).

Theorem 14. *The following set of inference rules*

$$\frac{}{X \to Y}\,Y\subseteq X \qquad \frac{X \to Y}{XU \to YV}\,V\subseteq U \qquad \frac{X \to Y, X \to Z}{X \to YZ}$$

$$\frac{X \to Y}{X \to Z}\,Z\subseteq Y \qquad \frac{X \to Y}{X \twoheadrightarrow Y} \qquad \frac{}{X \twoheadrightarrow Y}\,Y\subseteq X$$

$$\frac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\,V\subseteq U \qquad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ} \qquad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$

is sound and complete for the implication of functional and multivalued dependencies with null values.

The reflexivity rule for NMVDs is certainly redundant in this set of inference rules. It has been included to emphasise the fact that NFDs and NMVDs can be dealt with separately (even when they are specified together). This is entirely different from traditional relational databases without null values where FDs and MVDs have been shown to interact non-trivially (Beeri et al. 1977).

7. All Axiomatisations of NMVDs in Fixed Universes

Recall that $\mathfrak{L}_\mathcal{C} = \mathfrak{L} \cup \{\mathcal{C}_R\}$, i.e., we obtain the (minimal) complementary axiomatisation $\mathfrak{L}_\mathcal{C}$ in fixed universes from the (minimal) axiomatisation $\mathcal{L}$ in undetermined universes by adding the complementation rule $\mathcal{C}_R$. We will show in this section that this correspondence is not just limited to these particular (minimal) axiomatisations.

Theorem 15. *Let $\mathfrak{S}$ be a sound set of inference rules for the implication of NMVDs. The set $\mathfrak{S}$ is complete for the implication of NMVDs if and only if for all relation schemata R the set $\mathfrak{S}_\mathcal{C} = \mathfrak{S} \cup \{\mathcal{C}_R\}$ is R-complete and R-complementary for the R-implication of NMVDs.*

Proof. We show first that if $\mathfrak{S}$ is complete for the implication of NMVDs, then for each relation schema R the set $\mathfrak{S}_\mathcal{C} = \mathfrak{S} \cup \{\mathcal{C}_R\}$ is both R-complete and R-complementary for the R-implication of NMVDs.

Let R be arbitrary. We know that $\mathfrak{L}_\mathcal{C}$ is R-complete, i.e., $\Sigma^*_R \subseteq \Sigma^+_{\mathfrak{L}_\mathcal{C}}$. Moreover, $\mathfrak{S}$ and $\mathfrak{L}$ are both sound and complete, i.e., $\Sigma^+_\mathfrak{L} = \Sigma^* = \Sigma^+_\mathfrak{S}$. Let $X, Y \subseteq R$ and $X \twoheadrightarrow Y \in \Sigma^*_R$. Since $\mathfrak{L}_\mathcal{C}$ is R-complete it follows that $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{L}_\mathcal{C}}$. Corollary 9 shows that $X \twoheadrightarrow Y \in \Sigma^+_\mathfrak{L}$ or $X \twoheadrightarrow (R - Y) \in \Sigma^+_\mathfrak{L}$ holds. Since $\Sigma^+_\mathfrak{L} = \Sigma^+_\mathfrak{S}$ this is equivalent to $X \twoheadrightarrow Y \in \Sigma_{\mathfrak{S}}^{+}$ or $X \twoheadrightarrow (R-Y) \in \Sigma_{\mathfrak{S}}^{+}$. However, $Y = R-(R-Y)$ and therefore $X \twoheadrightarrow Y \in \Sigma_{\mathfrak{S}_{\mathcal{C}}}^{+}$. This shows that $\Sigma_R^{*} \subseteq \Sigma_{\mathfrak{S}_{\mathcal{C}}}^{+}$, i.e., $\mathfrak{S}_{\mathcal{C}}$ is R-complete. Moreover, $\Sigma_{\mathfrak{S}_{\mathcal{C}}}^{+} = \Sigma_{\mathfrak{L}_{\mathcal{C}}}^{+}$ and Corollary 9 imply that

$$X \twoheadrightarrow Y \in \Sigma_{\mathfrak{S}_{\mathcal{C}}}^{+} \quad \text{if and only if} \quad X \twoheadrightarrow Y \in \Sigma_{\mathfrak{S}}^{+} \text{ or } X \twoheadrightarrow (R-Y) \in \Sigma_{\mathfrak{S}}^{+}$$

whenever $X, Y \subseteq R$ and Σ is a set of NMVDs on R. That is, every inference of an NMVD $X \twoheadrightarrow Y$ using $\mathfrak{S}_{\mathcal{C}}$ can be turned into an inference of $X \twoheadrightarrow Y$ in which the R-complementation rule $\mathcal{C}_R$ is applied at most once, and if it is applied, then as the last rule of the inference. Since R was arbitrary the set $\mathfrak{S}_{\mathcal{C}}$ is R-complete and R-complementary for all relation schemata R.

It remains to show that $\mathfrak{S}$ is complete for the implication of NMVDs whenever for all relation schemata R the set $\mathfrak{S}_{\mathcal{C}} = \mathfrak{S} \cup \{\mathcal{C}_R\}$ is both R-complete and R-complementary for the R-implication of NMVDs. We need to show that $\Sigma^{*} \subseteq \Sigma_{\mathfrak{S}}^{+}$ holds for every finite set Σ of NMVDs. Let $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ and $X \twoheadrightarrow Y \in \Sigma^{*} = \Sigma_{\mathfrak{L}}^{+}$. Let $T := X \cup Y \cup \bigcup_{i=1}^{k}(X_i \cup Y_i)$ and R be some relation schema such that T is properly contained in R, i.e., $T \subset R$. Theorem 5 shows that $X \twoheadrightarrow Y \in \Sigma_R^{*}$. The R-completeness of $\mathfrak{S}_{\mathcal{C}}$ implies further that $X \twoheadrightarrow Y \in \Sigma_{\mathfrak{S}_{\mathcal{C}}}^{+}$. Since $\mathfrak{S}_{\mathcal{C}}$ is also R-complementary we must have $X \twoheadrightarrow Y \in \Sigma_{\mathfrak{S}}^{+}$ or $X \twoheadrightarrow (R-Y) \in \Sigma_{\mathfrak{S}}^{+}$. Assume that $X \twoheadrightarrow (R-Y) \in \Sigma_{\mathfrak{S}}^{+}$. The soundness of $\mathfrak{S}$ implies that $X \twoheadrightarrow (R-Y) \in \Sigma^{*} = \Sigma_{\mathfrak{L}}^{+}$. Furthermore, since the union rule belongs to $\mathfrak{L}$ it follows that $X \twoheadrightarrow R \in \Sigma_{\mathfrak{L}}^{+}$. However, we obtain the contradiction $R \subseteq T \subset R$ by Lemma 10. Consequently, $X \twoheadrightarrow Y \in \Sigma_{\mathfrak{S}}^{+}$ must hold, and this shows the completeness of $\mathfrak{S}$. □

Theorem 15 extends to minimal complete sets of inference rules.

Corollary 16. *Let $\mathfrak{S}$ be a sound set of inference rules for the implication of NMVDs. The set $\mathfrak{S}$ is minimal and complete for the implication of NMVDs if and only if for all relation schemata R the set $\mathfrak{S}_{\mathcal{C}} = \mathfrak{S} \cup \{\mathcal{C}_R\}$ is R-complete and R-complementary for the R-implication of NMVDs, and there is no inference rule $\mathfrak{R} \in \mathfrak{S}_{\mathcal{C}}$ such that for all relation schemata R the set $\mathfrak{S}_{\mathcal{C}} - \{\mathfrak{R}\}$ is still both R-complete and R-complementary for the R-implication of NMVDs.*

The last corollary helps us finding all subsets of $\{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I}, \mathcal{D}, \mathcal{C}_R\}$ that are R-complete and R-complementary for the R-implication of NMVDs. The next result is a consequence of Theorem 13, Corollary 16 and the fact that the R-complementation rule $\mathcal{C}_R$ is independent of the rules in $\mathcal{L}$.

Corollary 17. *There are no proper subsets of $\mathfrak{L}_{\mathcal{C}}$ which are both R-complete and R-complementary for the R-implication of NMVDs for all relation schemata R.*

The trade-off between complementarity and minimality seems to be also present in the case of NMVDs. The set $\mathfrak{L}_{\mathcal{C}} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}, \mathcal{C}_R\}$ is not minimal because $\mathcal{D}$ is derivable from $\{\mathcal{U}, \mathcal{C}_R\}$ since $Z - Y = Z \cap (R-Y) = R - ((R-Z) \cup Y)$.

8. Minimising Minimality

Recall that a complete set $\mathfrak{S}$ of inference rules is said to be minimal iff none of the rules in $\mathfrak{S}$ can be omitted from $\mathfrak{S}$ without losing completeness. In this sense the set $\mathfrak{L} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}\}$ is minimal for the implication of NMVDs. A stricter version of minimality would include that the side conditions of all inference rules cannot be weakened. For instance, since both the reflexivity axiom $\mathcal{R}$ and the augmentation rule $\mathcal{A}$ are present in $\mathfrak{L}$ one may replace $\mathcal{R}$ by the empty-set-axiom $\mathcal{R}_\emptyset$ and still maintain completeness. In fact, $\mathcal{R}_\emptyset$ is a very weak form of the reflexivity axiom $\mathcal{R}$ representing just the single instance of $\mathcal{R}$ where $X = Y = \emptyset$. However, $\mathcal{R}$ is derivable from $\{\mathcal{R}_\emptyset, \mathcal{A}\}$:

$$\begin{array}{rl} \mathcal{R}_\emptyset : & \overline{\emptyset \twoheadrightarrow \emptyset} \\ \mathcal{A} : & \overline{X \twoheadrightarrow Y}\; Y \subseteq X \end{array} .$$

Lemma 18. *The set $\{\mathcal{R}_\emptyset, \mathcal{A}, \mathcal{U}, \mathcal{D}\}$ is sound and complete for the implication of multivalued dependencies with null values.*

Instead of weakening the reflexivity axiom, one may replace the augmentation rule $\mathcal{A}$ by the weak augmentation rule $\mathcal{W}$: $\dfrac{X \twoheadrightarrow Y}{XA \twoheadrightarrow Y}$ which is a very restricted form of augmentation in which $V = \emptyset$ and $U = A$ is a singleton. However, $\mathcal{A}$ can be derived from $\{\mathcal{R}, \mathcal{W}, \mathcal{U}\}$ as follows (suppose $U = \{A_1, \ldots, A_k\}$):

$$\begin{array}{ll} \mathcal{W} : & \dfrac{X \twoheadrightarrow Y}{XA_1 \twoheadrightarrow Y} \\ & \quad\vdots \\ \mathcal{W} : & \overline{XA_1 \cdots A_k \twoheadrightarrow Y} \qquad \mathcal{R} : \overline{XU \twoheadrightarrow V}\; V \subseteq U \subseteq XU \\ \mathcal{U} : & \overline{\qquad\qquad XU \twoheadrightarrow YV \qquad\qquad} \end{array} .$$

The reflexivity axiom $\mathcal{R}$ may also be replaced by the empty-set-axiom $\mathcal{R}_\emptyset$ and the attribute axiom $\mathcal{A}t$: $\overline{A \twoheadrightarrow A}$. In fact, $\mathcal{R}$ can be derived from $\{\mathcal{R}_\emptyset, \mathcal{A}t, \mathcal{W}, \mathcal{U}\}$. If $Y = \emptyset$ and X consists of k attributes, then we apply the empty-set-axiom $\mathcal{R}_\emptyset$ first to derive $\emptyset \twoheadrightarrow \emptyset$. Subsequently, the weak augmentation rule $\mathcal{W}$ is applied k times to derive $X \twoheadrightarrow \emptyset$. In case that $Y = \{B_1, \ldots, B_l\}$ and X has k attributes, $k \geq l$, we derive $B_1 \twoheadrightarrow B_1, \ldots, B_l \twoheadrightarrow B_l$ by l applications of the attribute axiom $\mathcal{A}t$. Subsequently, we apply the weak augmentation rule $\mathcal{W}$ to each of these NMVDs k times to derive $X \twoheadrightarrow B_1, \ldots, X \twoheadrightarrow B_l$. Finally, the union rule $\mathcal{U}$ is applied $l-1$ times to derive $X \twoheadrightarrow Y$.

Theorem 19. *The set $\{\mathcal{R}_\emptyset, \mathcal{A}t, \mathcal{W}, \mathcal{U}, \mathcal{D}\}$ is sound and complete for the implication of multivalued dependencies with null values.*

9. Conclusion and Future Work

We have explored the notion of MVD implication in the presence of null values (NMVDs) with meaning "no information". We observed that Lien's original axiomatisation of NMVDs (Lien 1982) does not adequately reflect the role of the R-complementation rule as a mere means of database normalisation. This observation is analagous to Biskup's findings for total database relations. We have then proposed sound and complete sets of inference rules for the R-implication of NMVDs that are indeed adequate. Moreover, Biskup's alternative notion of MVD implication, in which the underlying universe is left undetermined, was extended to the presence of null values. Several sound and complete sets of inference rules for NMVD implication in undetermined universes have been proposed, which were also extended to cover both functional and multivalued dependencies in the presence of null values. The results clarify the role of the R-complementation rule for NMVDs, and may simplify the quest of finding suitable and comprehensible notions of multivalued dependencies in the context of advanced database models. Moreover, the results clarify the power of several R-incomplete subsets.

Some interesting problems warrant future research. While the R-implication problem of MVDs has received a considerable amount of interest with the best current time bound proposed in (Galil 1982), no research has been devoted to the corresponding R-implication problem of NMVDs. In the spirit of our article it seems also interesting to investigate the implication problem of (N)MVDs in undetermined universes, and maybe derive further correspondences between implication and R-implication. An interesting open problem is to generalise the approach in (Levene & Loizou 1998) from functional to multivalued dependencies. The approach uses a possible world semantics exploring all extensions of an incomplete database to a complete database. Weak MVDs must be satisfied by some possible world while strong MVDs are satisfied by all possible worlds.

The probably most important open problem in the context of multivalued dependencies is the absence of a synthesis algorithm that extends the well-known synthesis approach to functional dependencies (Bernstein 1976, Biskup et al. 1979). Clearly, the notion of MVD implication in undetermined universes provides a much better basis for the development of such an algorithm than the original notion of MVD implication in fixed universes.

References

Arenas, M. & Libkin, L. (2004), 'A normal form for XML documents', *Trans. Database Syst.* **29**(1), 195–232.

Armstrong, W. W. (1974), 'Dependency structures of database relationships', *Information Processing* **74**, 580–583.

Armstrong, W. W., Nakamura, Y. & Rudnicki, P. (2002), 'Armstrong's axioms', *Journal of formalized Mathematics* **14**.

Atzeni, P. & Morfuni, N. (1986), 'Functional dependencies and constraints on null values in database relations', *Information and Control* **70**(1), 1–31.

Beeri, C. & Bernstein, P. A. (1979), 'Computational problems related to the design of normal form relational schemata', *Trans. Database Syst.* **4**(1), 30–59.

Beeri, C., Fagin, R. & Howard, J. H. (1977), A complete axiomatization for functional and multivalued dependencies in database relations, *in* 'SIGMOD Conference', ACM, pp. 47–61.

Bernstein, P. (1976), 'Synthesizing third normal form relations from functional dependencies', *Trans. Database Syst.* **1**(4), 277–298.

Bernstein, P. A. & Goodman, N. (1980), What does Boyce-Codd normal form do?, *in* 'VLDB Conference', IEEE Computer Society, pp. 245–259.

Biskup, J. (1978), 'On the complementation rule for multivalued dependencies in database relations', *Acta Inf.* **10**(3), 297–305.

Biskup, J. (1980), 'Inferences of multivalued dependencies in fixed and undetermined universes', *Theor. Comput. Sci.* **10**(1), 93–106.

Biskup, J., Dayal, U. & Bernstein, P. (1979), Synthesizing independent database schemas, *in* 'SIGMOD Conference', pp. 143–151.

Codd, E. F. (1970), 'A relational model of data for large shared data banks', *Commun. ACM* **13**(6), 377–387.

Codd, E. F. (1972), Further normalization of the database relational model, *in* 'Courant Computer Science Symposia 6: Data Base Systems', Prentice-Hall, pp. 33–64.

Codd, E. F. (1975), 'Understanding relations', *ACM SIGFIDET FDT Bulletin* **7**(3-4), 23–28.

Codd, E. F. (1979), 'Extending the database relational model to capture more meaning', *Trans. Database Syst.* **4**(4), 397–434.

Delobel, C. (1978), 'Normalisation and hierarchical dependencies in the relational data model', *Trans. Database Syst.* **3**(3), 201–222.

Fagin, R. (1977), 'Multivalued dependencies and a new normal form for relational databases', *Trans. Database Syst.* **2**(3), 262–278.

Fagin, R. & Vardi, M. Y. (1986), The theory of data dependencies: a survey, *in* 'Mathematics of Information Processing: Proceedings of Symposia in Applied Mathematics', American Mathematical Society, pp. 19–71.

Fischer, P. C., Saxton, L. V., Thomas, S. J. & Van Gucht, D. (1985), 'Interactions between dependencies and nested relational structures', *J. Comput. Syst. Sci.* **31**(3), 343–354.

Galil, Z. (1982), 'An almost linear-time algorithm for computing a dependency basis in a relational database', *J. ACM* **29**(1), 96–102.

Gottlob, G. & Zicari, R. (1988), Closed world databases opened through null values, *in* 'VLDB Conference', IEEE Computer Society, pp. 50–61.

Grahne, G. (1984), Dependency satisfaction in databases with incomplete information, *in* 'VLDB Conference', IEEE Computer Society, pp. 37–45.

Grant, J. (1977), 'Null values in a relational data base', *Information Processing Letters* **6**(5), 156–157.

Hara, C. & Davidson, S. (1999), Reasoning about nested functional dependencies, *in* 'Proceedings of the 18th SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems', ACM, pp. 91–100.

Hartmann, S. & Link, S. (2006*a*), 'Deciding implication for functional dependencies in complex-value databases', *Theor. Comput. Sci.* **364**(2), 212–240.

Hartmann, S. & Link, S. (2006*b*), 'On a problem of Fagin concerning multivalued dependencies in relational databases', *Theor. Comput. Sci.* **353**(1-3), 53–62.

Hartmann, S., Link, S. & Köhler, H. (2007), 'Full hierarchical dependencies in fixed and undetermined universes', *Ann. Math. Artif. Intell.* **50**(1-2), 195–226.

Hartmann, S., Link, S. & Schewe, K.-D. (2006*a*), 'Axiomatisations of functional dependencies in the presence of records, lists, sets and multisets', *Theor. Comput. Sci.* **355**(2), 167–196.

Hartmann, S., Link, S. & Schewe, K.-D. (2006*b*), 'Functional and multivalued dependencies in nested databases generated by record and list constructor', *Ann. Math. Artif. Intell.* **46**(1-2), 114–164.

Johnson, M. & Rosebrugh, R. (2003), 'Three approaches to partiality in the sketch data model', *Electronic Notes in Theoretical Computer Science* **78**, 1–18.

Levene, M. & Loizou, G. (1993), 'Semantics for null extended nested relations', *Trans. Database Syst.* **18**(3), 414–459.

Levene, M. & Loizou, G. (1998), 'Axiomatisation of functional dependencies in incomplete relations', *Theor. Comput. Sci.* **206**(1-2), 283–300.

Lien, Y. E. (1982), 'On the equivalence of data models', *J. ACM* **29**(2), 333–363.

Link, S. (2006*a*), On multivalued dependencies in fixed and undetermined universes, *in* 'Foundations of Information and Knowledge Systems, 4th International Symposium', number 3861 *in* 'Lecture Notes in Computer Science', Springer, pp. 258–277.

Link, S. (2006*b*), 'On the logical implication of multivalued dependencies with null values', *Conferences in Research and Practice in Information Technology* **51**, 113–122.

Mendelzon, A. (1979), 'On axiomatising multivalued dependencies in relational databases', *J. ACM* **26**(1), 37–44.

Mikinouchi, A. (1977), A consideration on normal form of not-necessarily-normalised relation in the relational data model, *in* 'VLDB Conference', IEEE Computer Society, pp. 447–453.

Tari, Z., Stokes, J. & Spaccapietra, S. (1997), 'Object normal forms and dependency constraints for object-oriented schemata', *Trans. Database Syst.* **22**, 513–569.

Thalheim, B. (1991), *Dependencies in Relational Databases*, Teubner-Verlag.

Thalheim, B. (2003), Conceptual treatment of multivalued dependencies, *in* 'Conceptual Modeling - ER 2003, 22nd International Conference on Conceptual Modeling, Chicago, IL, USA, October 13-16, 2003, Proceedings', number 2813 *in* 'Lecture Notes in Computer Science', Springer, pp. 363–375.

Vincent, M. & Liu, J. (2003), Multivalued dependencies in XML, *in* 'Proceedings of the 20th British National Conference on Databases', number 2712 *in* 'Lecture Notes in Computer Science', Springer, pp. 4–18.

Vincent, M., Liu, J. & Liu, C. (2003), A redundancy free 4NF for XML, *in* 'Proceedings of the 1st International XML Database Symposium', number 2824 *in* 'Lecture Notes in Computer Science', Springer, pp. 254–266.

Weddell, G. (1992), 'Reasoning about functional dependencies generalized for semantic data models', *Trans. Database Syst.* **17**(1), 32–64.

Wijsen, J. (1999), 'Temporal FDs on complex objects', *Trans. Database Syst.* **24**(1), 127–176.

Zaniolo, C. (1976), Analysis and Design of Relational Schemata for Database Systems, PhD thesis, UCLA, Tech. Rep. UCLA-ENG-7769.

Zaniolo, C. (1984), 'Database relations with null values', *J. Comput. Syst. Sci.* **28**(1), 142–166.