On the logical Implication of Multivalued Dependencies with Null Values

Sebastian Link[†]

Department of Information Systems, Information Science Research Centre
Massey University, Palmerston North, New Zealand
E-mail: S.Link@massey.ac.nz

Abstract

The implication of multivalued dependencies (MVDs) in relational databases has originally been defined in the context of some fixed finite universe (Fagin 1977, Zaniolo 1976). While axiomatisability, implication problem and many design problems have been intensely studied with respect to this notion, almost no research has been devoted towards the alternative notion of implication in which the underlying universe of attributes is left undetermined (Biskup 1980).

A milestone in the advancement of database systems was the permission of null values in databases. In particular, many achievements on MVDs have been extended to encompass incomplete information. Multivalued dependencies with null values (NMVDs) were defined and axiomatised in (Lien 1982). The definition of NMVDs is again based on a fixed underlying universe of attributes, and any complete set of inference rules requires therefore some version of the complementation rule.

In this paper we show that the axiomatisation in (Lien 1982) does not reflect the fact that the complementation rule is merely a means to achieve database normalisation. Moreover, we provide an alternative axiomatisation for NMVDs that does reflect this property. We also suggest an alternative notion for the implication of NMVDs in which the underlying universe is left undetermined, and propose several sound and complete sets of inference rules for this notion. 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: Database Theory, Multivalued Dependency, Null Values, Implication, Axiomatisation

1 Introduction

Relational databases still form the core of most database management systems, even after more than three decades following their introduction in (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 database is compelled to obey. There are many different classes of dependencies which can be utilised for improving the representation of the target 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 2004) and XML (Vincent & Liu 2003, Vincent, Liu & Liu 2003).

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 in (Lien 1982). The notion of an NMVD from (Lien 1982), as well as the original notion of an MVD (Fagin 1977), is dependent on the underlying set R of attributes. This dependence is reflected syntactically by the R-complementation rule which is part of the axiomatisation of NMVDs, see (Lien 1982). The complementation rule is special in the sense that it is the only inference rule which is dependent on R. In the absence of null values, this observation has led to further research (Mendelzon 1979, Biskup 1978, Biskup 1980, Hartmann & Link 2006, Link 2006) on the complementation rule. In particular, Biskup introduced an alternative notion of semantic implication in which the underlying universe is left undetermined (Biskup 1980). In the same paper it was shown that this notion can be captured syntactically by a sound and complete set of inference rules, denoted by $\mathfrak{S}_0$. If $R\mathfrak{S}_0$ results from $\mathfrak{S}_0$ by adding the R-complementation rule, then $R\mathfrak{S}_0$ is R-sound and R-complete for the R-implication of MVDs. Moreover, every inference of an MVD by $R\mathfrak{S}_0$ can be turned into an inference of the same MVD in which the R-complementation rule is applied at most once, and if it is applied, then in the last step of the inference ($R\mathfrak{S}_0$ is said to be complementary). This indicates that the R-complementation rule simply re-

---

[†] Sebastian Link is supported by Marsden Funding, Royal Society of New Zealand

flects a part of the decomposition process, and is not necessarily essential for deriving valid consequences. Interestingly, this research has not been extended to encompass incomplete information, i.e., to NMVDs. Since research on (N)MVDs seems to experience a recent revival in the context of other data models (Fischer et al. 1985, Thalheim 2003, Hartmann & Link 2004, Vincent & Liu 2003, Vincent et al. 2003) it seems desirable to further extend the knowledge on (N)MVDs. An advancement of such knowledge may simplify the quest of finding suitable and comprehensible extensions of MVDs to currently popular data models.

In this paper we will extend the alternative notion of implication from MVDs (Biskup 1980) to the presence of null values. First, it is demonstrated that the sound and complete set $R\mathfrak{K}$ of inference rules for the R-implication of NMVDs from (Lien 1982) is not complementary. Moreover, we propose a sound and complete set $R\mathfrak{L}$ that is indeed complementary for the R-implication of NMVDs. Subsequently, we will identify a sound and complete set $\mathfrak{L}$ of inference rules for the implication of NMVDs in undetermined universes. Thus, $\mathfrak{L}$ does not permit the possibly semantically meaningless inference of complementation. Finally, the set $\mathfrak{L}$ is extended to obtain a sound and complete set of inference rules for the implication of functional and multivalued dependencies in the presnce of null values in undetermined universes. The problems studied in this paper are not just of theoretical interest. In practice one does not necessarily want to generate all consequences of a given set of NMVDs 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.

The paper is structured as follows. Section 2 repeats fundamental notions from the relational model of data as well as incomplete information. In particular, the notions for implication of multivalued dependencies in the presence of null values are highlighted. After the axiomatisation of NMVDs from (Lien 1982) has been reviewed Section 3 identifies R-sound and R-complete sets of inference rules for the R-implication of NMVDs that are also R-complementary. Furthermore, the alternative notion of implication for NMVDs in which universes are left undetermined is introduced, and a sound and complete set $\mathfrak{L}$ of inference rules proposed. The result is extended to capture both functional and multivalued dependencies in the presence of nulls in undetermined universes. Some general results are proposed in Section 4 which show a correspondence between (minimal) complete sets of inference rules in undetermined universes and (minimal) complete and complementary sets of inference rules in fixed universes. Section 5 suggests some alternative axiomatisations for NMVDs in undetermined universes that only require weak versions of some of the inference rules in $\mathfrak{L}$. The paper concludes in Section 6.

2 MVDs in Relational Databases

We use this section to introduce some notation, and repeat notions and results for dependencies in the presence of null values.

2.1 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 particular, 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 "undefined", "inapplicable", or "nonexistent". For instance, the *maiden name* of a male *employee* may have a null value to mean inapplicable, or the *middle name* of an *employee* may have a null value to mean nonexistent.

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* 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* or simply *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 in the Presence of Nulls

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 2004, Levene & Loizou 1998, Tari, Stokes & Spaccapietra 1997, Weddell 1992, Wijsen 1999). The notion of a functional dependency is well-understood and the semantic interaction between these dependencies has been syntactically captured by Armstrong's well-known axioms (Armstrong 1974, Armstrong, Nakamura & Rudnicki 2002).

Let R be a relation schema. A *functional dependency with nulls* on R, abbreviated NFD, is a statement $X \to Y$ where $X, Y \subseteq R$. The NFD $X \to Y$ on R is satisfied by a partial relation r over R, denoted by $\models_r X \to Y$, 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 nonnull X-value, they agree on the Y-value, which may be partial.

FDs 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 in the presence of null values have been introduced in (Lien 1982).

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 from 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 (Lien 1982) 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). This means that $\models_r X \twoheadrightarrow Y$ if and only if $r_X[R] = r_X[XY] \bowtie r_X[X(R-Y)]$. This characteristic of NMVDs is fundamental to database design 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 et al. 1985), the Entity-Relationship model (Thalheim 2003), data models that support nested lists (Hartmann & Link 2004) and XML (Vincent & Liu 2003, Vincent et al. 2003).

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.

Definition 2.1. *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 NMVDs 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 NMVDs in Σ also satisfies $X \twoheadrightarrow Y$.* □

In order to determine all logical consequences of a finite set of NMVDs one can use the following set of inference rules which was proposed in (Lien 1982). Note that we use the natural complementation rule (Biskup 1978) instead of the complementation rule that was originally proposed (Lien 1982).

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

(reflexivity, $\mathcal{R}$) (augmentation, $\mathcal{A}$)

$$\frac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y} \qquad\qquad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ}$$

(R-complementation, $\mathcal{C}_R$) (union, $\mathcal{U}$)

$$\frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y} \qquad\qquad \frac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}$$

(difference, $\mathcal{D}$) (intersection, $\mathcal{I}$)

In (Lien 1982) the set $R\mathfrak{K} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{C}_R\}$ of inference rules is proven to be both R-sound and R-complete for the R-implication of MVDs, on all relation schemata R. Let $\Sigma \cup \{\sigma\}$ be a set of NMVDs on the relation schema R. Let $\Sigma \vdash_{\mathfrak{S}} \sigma$ denote the inference of σ from a set Σ of NMVDs with respect to the set $\mathfrak{S}$ of inference rules. Let $\Sigma^+_{\mathfrak{S}} = \{\sigma \mid \Sigma \vdash_{\mathfrak{S}} \sigma\}$ denote the *syntactic hull* of Σ under inference using only rules from $\mathfrak{S}$. The set $R\mathfrak{S}$ is called *R-sound* for the R-implication of NMVDs iff for every set Σ of NMVDs on R we have $\Sigma^+_{R\mathfrak{S}} \subseteq \Sigma^*_R = \{\sigma \mid \Sigma\ R\text{-implies}\ \sigma\}$. The set $R\mathfrak{S}$ is called *R-complete* for the implication of NMVDs if and only if for every set Σ of NMVDs on R we have $\Sigma^*_R \subseteq \Sigma^+_{R\mathfrak{S}}$. The set $R\mathfrak{S}$ is called sound (complete) for the R-implication of NMVDs iff it is R-sound (R-complete) for the R-implication of NMVDs for all relation schemata R.

An interesting question is now whether all the rules of a certain set are really necessary to capture the R-implication of NMVDs for every R. More precisely, an inference rule $\Re$ is said to be *R-independent* from the set $R\mathfrak{S}$ if and only if there is some set $\Sigma \cup \{\sigma\}$ of NMVDs on the relation schema R such that $\sigma \notin \Sigma^+_{R\mathfrak{S}}$, but $\sigma \in \Sigma^+_{R\mathfrak{S} \cup \{\Re\}}$. Moreover, the inference rule $\Re$ is said to be *independent* from $R\mathfrak{S}$ if and only if there is some relation schema R such that $\Re$ is *R-independent* from $R\mathfrak{S}$. Finally, a complete set $R\mathfrak{S}$ is said to be *minimal* for the R-implication of NMVDs if and only if every inference rule $\Re \in R\mathfrak{S}$ is independent from $R\mathfrak{S} - \{\Re\}$. This means that no proper subset of $R\mathfrak{S}$ is still complete.

Apart from $R\mathfrak{K}$ the following sets are also complete for the R-implication of NMVDs. This fact was not noticed in (Lien 1982), but is not difficult to see.

Theorem 2.1. *The sets $R\mathfrak{K}_1 = \{\mathcal{R}, \mathcal{A}, \mathcal{I}, \mathcal{C}_R\}$ and $R\mathfrak{K}_2 = \{\mathcal{R}, \mathcal{A}, \mathcal{D}, \mathcal{C}_R\}$ are sound and complete for the R-implication of NMVDs.*

Proof. The union rule $\mathcal{U}$ is derivable from $\{\mathcal{I}, \mathcal{C}_R\}$

$$\cfrac{\cfrac{\cfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y} \qquad \cfrac{X \twoheadrightarrow Z}{X \twoheadrightarrow R - Z}}{X \twoheadrightarrow (R-Y) \cap (R-Z)}}{X \twoheadrightarrow \underbrace{R - ((R-Y) \cap (R-Z))}_{=Y \cup Z}}$$

and the intersection rule $\mathcal{I}$ is derivable from $\{\mathcal{D}\}$

$$\cfrac{\cfrac{X \twoheadrightarrow Y \quad X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y} \qquad X \twoheadrightarrow Z}{X \twoheadrightarrow \underbrace{Z - (Z - Y)}_{=Y \cap Z}}$$

□

Note that $\dfrac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$, the *pseudo-transitivity* rule $\mathcal{T}$, which is essential for MVDs (Beeri, Fagin & Howard 1977), is not R-sound in the presence of null values (Lien 1982).

2.3 NMVDs in Undetermined Universes

Consider a slight modification of the classical example (Fagin 1977) in which the NMVD *Employee* $\twoheadrightarrow$ *Child* is specified, i.e., the set of children is completely determined by an employee, independently from 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 NMVD *Employee* $\twoheadrightarrow$ *Salary* by means of the R-complementation rule. However, if the underlying relation schema R consists of the four attributes *Employee*, *Child*, *Salary* and *Year*, then the NMVD *Employee* $\twoheadrightarrow$ *Salary* is no longer R-implied. Note the fundamental difference between the NMVDs

Employee $\twoheadrightarrow$ *Child* and *Employee* $\twoheadrightarrow$ *Salary*.

The first NMVD has been specified to establish the relationship of employees and their children as a fact due to a set-valued correspondence. The second NMVD 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 NMVD disappears. We can therefore observe the following:

- If consequences of NMVDs are inferred by a set $R\mathfrak{S}$ of inference rules with respect to a fixed universe R, applications of the R-complementation rule during any inference by $R\mathfrak{S}$ should be either completely avoided or limited to the very last step of the inference.
- It may be argued that consequences which are dependent on the underlying universe are in fact no consequences at all. This implies, however, that the notion of R-implication is not suitable.

We follow the first observation first, and come back to the second observation later. One may ask whether the R-sound and R-complete set $R\mathfrak{K} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{C}_R\}$ of inference rules reflects the property of R-complementation. More precisely, a complete set $R\mathfrak{S}$ of inference rules for the R-implication of (N)MVDs is said to be *complementary* iff it is R-complementary for all relation schemata R, i.e., for each $X \twoheadrightarrow Y \in \Sigma^+_{R\mathfrak{S}}$ there is an inference of $X \twoheadrightarrow Y$ from Σ by $R\mathfrak{S}$ in which the R-complementation rule $\mathcal{C}_R$ is applied at most once and (if at all) as the last rule. The following example shows that the set $R\mathfrak{K}$ is not complementary.

Example 2.1. *Let* $\Sigma = \{A \twoheadrightarrow BC, A \twoheadrightarrow B\}$*. The following table represents the syntactic hull* $\Sigma^+_{\{\mathcal{R},\mathcal{A},\mathcal{U}\}}$ *of* Σ *under inferences using* $\{\mathcal{R}, \mathcal{A}, \mathcal{U}\}$*. The NMVD* $U \twoheadrightarrow V$ *belongs to* $\Sigma^+_{\{\mathcal{R},\mathcal{A},\mathcal{U}\}}$ *iff the entry at row labelled* U *and column labelled* V *is a cross* $\times$.

	$\emptyset$	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 3.1 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* $A \twoheadrightarrow C \in \Sigma^+_{R\mathfrak{K}}$*, say by*

$$\dfrac{\dfrac{\dfrac{A \twoheadrightarrow BC}{A \twoheadrightarrow AD} \qquad A \twoheadrightarrow B}{A \twoheadrightarrow ABD}}{A \twoheadrightarrow C} \quad .$$

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

Biskup introduced the following sound inference rules for the R-implication of MVDs

$$\underset{(\text{empty-set-axiom},\mathcal{R}_\emptyset)}{\dfrac{}{\emptyset \twoheadrightarrow \emptyset}} \qquad \underset{(\text{additive transitivity, }\mathcal{T}^*)}{\dfrac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow YZ}}$$

$$\underset{(\text{subset, }\mathcal{S})}{\dfrac{X \twoheadrightarrow Y, W \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}\, Y \cap W = \emptyset}$$

and showed that the complete set $R\mathfrak{B} = \{\mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{S}, \mathcal{C}_R\}$ is also complementary. That is,

$$\begin{gathered} X \twoheadrightarrow Y \in \Sigma^+_{R\mathfrak{B}} \\ \text{if and only if} \\ X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{B}} \text{ or } X \twoheadrightarrow (R - Y) \in \Sigma^+_{\mathfrak{B}} \end{gathered} \tag{2.1}$$

where $\Sigma = \{X_1 \twoheadrightarrow Y_1, \ldots, X_k \twoheadrightarrow Y_k\}$ and $X \cup Y \cup \bigcup_{i=1}^{k} (X_i \cup Y_i) \subseteq R$.

Moreover, Biskup introduced an alternative notion of implication for MVDs (Biskup 1980), which leaves the underlying relation schema undetermined. This brings us back to our second observation from above. We will now generalise the notion of MVDs in undetermined universes to the presence of null values. An NMVD is a syntactic expression $X \twoheadrightarrow Y$ with $X, Y \subseteq \mathfrak{A}$. 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 2.2. *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 2.2. *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 2.2 is false.

Example 2.2. *For* R *=*{*Employee,Child,Salary*} *and* Σ *=*{*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 $r_{Employee}[Employee,\ Child]$

Employee	*Child*
Don Juan	ν

and $r_{Employee}[Employee,\ Salary,\ Year]$

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

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

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

and $r_{Employee}[Employee,\ Child,\ Year]$

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.* ☐

A set $\mathfrak{S}$ of inference rules is called *sound* for the implication of (N)MVDs if and only if for every finite set Σ of (N)MVDs we have $\Sigma^+_{\mathfrak{S}} \subseteq \Sigma^* = \{\sigma \mid \Sigma \text{ implies } \sigma\}$. The set $\mathfrak{S}$ is called *complete* for the implication of (N)MVDs if and only if for every finite set Σ of (N)MVDs we have $\Sigma^* \subseteq \Sigma^+_{\mathfrak{S}}$. An inference rule $\Re$ is said to be *independent* from the set $\mathfrak{S}$ if and only if there is some finite set $\Sigma \cup \{\sigma\}$ of (N)MVDs such that $\sigma \notin \Sigma^+_{\mathfrak{S}}$, but $\sigma \in \Sigma^+_{\mathfrak{S}\cup\{\Re\}}$. A complete set $\mathfrak{S}$ of inference rules is said to be *minimal* for the implication of (N)MVDs if and only if every inference rule $\Re$ in $\mathfrak{S}$ is independent from $\mathfrak{S} - \{\Re\}$. This means that no proper subset of $\mathfrak{S}$ is still complete for the implication of (N)MVDs.

It should be noted that the singletons $\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}, \mathcal{I}$ are all sound, and the R-complementation rule is R-sound, but not sound as just demonstrated by the previous example.

The second major result in (Biskup 1980) shows that the set $\mathfrak{B} = \{\mathcal{R}_\emptyset, \mathcal{A}, \mathcal{T}, \mathcal{T}^*, \mathcal{S}\}$ is sound and complete for the implication of MVDs in undetermined universes.

3 NMVDs in fixed and undetermined Universes

The last section has identified two major objectives:

1. Find a set $R\mathfrak{L}$ of inference rules which is sound, complete and complementary for the R-implication of NMVDs.
2. Identify a set $\mathfrak{L}$ of inference rules which is sound and complete for the implication of NMVDs.

Biskup has successfully provided solutions to these two problems for MVDs, i.e., in the absence of null values. One may hope that the inclusion of the additive transitivity rule $\mathcal{T}^*$ and/or subset rule $\mathcal{S}$ into $R\mathfrak{K}$ result in a complete set of inference rules that is also complementary. Both rules, however, are not sound for the implication of NMVDs as the following example demonstrates.

Example 3.1. *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 the incorrectness of the subset rule for NMVDs.*

The incorrectness of the additive transitivity rule follows from the following example. 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*.* ☐

3.1 NMVDs in fixed Universes

Our first theorem shows that there are indeed complete sets which are complementary. 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 hull $\Sigma^+_{\mathfrak{S}}$ is the set of all NMVDs which can be inferred by some inference from Σ by $\mathfrak{S}$.

Theorem 3.1. *Let* Σ *be a set of NMVDs on the relation schema* R*. For each inference* γ *from* Σ *by the set* $R\mathfrak{K} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{C}_R\}$ *there is an inference* ξ *from* Σ *by the set* $R\mathfrak{L} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}, \mathcal{C}_R\}$ *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 $\mathfrak{L}$ 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:

$$\frac{\dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}}{XU \twoheadrightarrow (R - Y)V}\, V \subseteq U.$$

However, these two steps can be replaced as follows:

$$\dfrac{\dfrac{\dfrac{X \twoheadrightarrow Y}{XU \twoheadrightarrow Y} \qquad \dfrac{}{XU \twoheadrightarrow V}{}^{V\subseteq U\subseteq XU}}{XU \twoheadrightarrow Y-V}}{XU \twoheadrightarrow \underbrace{R-(Y-V)}_{=(R-Y)V}}$$

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 $\gamma_i = [\sigma_1, \dots, \sigma_i]$ and $\gamma_j = [\sigma_1, \dots, \sigma_j]$, 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:

$$\dfrac{\dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R-Y} \qquad \dfrac{}{X \twoheadrightarrow Z}}{X \twoheadrightarrow (R-Y)Z} \; .$$

However, these steps can be replaced as follows:

$$\dfrac{\dfrac{X \twoheadrightarrow Z \qquad X \twoheadrightarrow Y}{X \twoheadrightarrow Y-Z}}{X \twoheadrightarrow \underbrace{R-(Y-Z)}_{=(R-Y)Z}} \; .$$

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:

$$\dfrac{X \twoheadrightarrow Y \qquad \dfrac{X \twoheadrightarrow Z}{X \twoheadrightarrow R-Z}}{X \twoheadrightarrow Y(R-Z)} \; .$$

However, these steps can be replaced as follows:

$$\dfrac{\dfrac{X \twoheadrightarrow Y \qquad X \twoheadrightarrow Z}{X \twoheadrightarrow Z-Y}}{X \twoheadrightarrow \underbrace{R-(Z-Y)}_{=Y(R-Z)}} \; .$$

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:

$$\dfrac{\dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R-Y} \qquad \dfrac{X \twoheadrightarrow Z}{X \twoheadrightarrow R-Z}}{X \twoheadrightarrow (R-Y) \cup (R-Z)} \; .$$

However, these steps can be replaced as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{X \twoheadrightarrow Y \qquad X \twoheadrightarrow Z}{X \twoheadrightarrow Z-Y} \qquad X \twoheadrightarrow Z}{X \twoheadrightarrow \underbrace{Z-(Z-Y)}_{=Y\cap Z}}}{X \twoheadrightarrow \underbrace{R-(Y\cap Z)}_{=(R-Y)\cup(R-Z)}}} \; .$$

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 $\gamma_i := [\sigma_1, \dots, \sigma_i]$. 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:

$$\dfrac{\dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R-Y}}{X \twoheadrightarrow \underbrace{R-(R-Y)}_{=Y}} \; .$$

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

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

3.2 NMVDs in Undetermined Universes

We now explore the power of the common part of the sets $R\mathfrak{L}$, namely $\mathfrak{L} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{D}\}$, which can be obtained from any of the sets $R\mathfrak{L}$ by removing the R-complementation rule $\mathcal{C}_R$. Hence, $\mathfrak{L}$ does not permit the possibly semantically meaningless inference of complementation.

Theorem 3.1 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 3.1. *Let $R \subseteq \mathfrak{A}$ be a finite set of attributes. Then for all finite sets $\Sigma = \{X_1 \twoheadrightarrow Y_1, \dots, X_k \twoheadrightarrow Y_k\}$ 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*

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

□

Corollary 3.1 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 2.2, that is by $\mathfrak{L}$ we can generate exactly all implications in an undetermined universe.

Lemma 3.1. *Let $\Sigma = \{X_1 \twoheadrightarrow Y_1, \dots, X_k \twoheadrightarrow Y_k\}$ 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 $\gamma = [\sigma_1, \dots, \sigma_l]$ 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. ☐

Lemma 3.2. *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 3.1 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 3.2. *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

$$\Sigma \text{ implies } X \twoheadrightarrow Y \quad \text{iff} \quad X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{L}} \quad . \tag{3.2}$$

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 (3.2)) 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 3.2 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 (3.2)) 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 3.1, $X \twoheadrightarrow (R - Y) \notin \Sigma^+_{\mathfrak{L}}$. Now from $X \twoheadrightarrow Y \notin \Sigma^+_{\mathfrak{L}}$ and $X \twoheadrightarrow (R - Y) \notin \Sigma^+_{\mathfrak{L}}$ we conclude that $X \twoheadrightarrow Y \notin R\mathfrak{L}$ by Corollary 3.1. Since $R\mathfrak{L}$ 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 2.2. ☐

3.3 All minimal Sets of Inference Rules

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. The proof of the following result consists of independence proofs. These independence proofs have been computationally verified using GNU pascal (providing set arithmetic) programs.

Theorem 3.3. $\mathfrak{L}$ *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. We show that all four inference rules of $\mathfrak{L}$ are essential for gaining completeness, i.e., each inference rule is independent from the rest of the inference rules.

- The reflexivity axiom $\mathcal{R}$ is independent from $\mathfrak{S} = \{\mathcal{A}, \mathcal{I}, \mathcal{D}, \mathcal{U}\}$. 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 from $\mathfrak{S} = \{\mathcal{R}, \mathcal{I}, \mathcal{D}, \mathcal{U}\}$. 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 $X, Y \subseteq \mathfrak{A}$.

	∅	A	B	C	AB	AC	BC	ABC
∅	×							
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 from $\mathfrak{S} = \{\mathcal{R}, \mathcal{A}, \mathcal{I}, \mathcal{D}\}$. 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 $X, Y \subseteq \mathfrak{A}$.

	∅	A	B	C	AB	AC	BC	ABC
∅	×							
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 from $\mathfrak{S} = \{\mathcal{R}, \mathcal{A}, \mathcal{I}, \mathcal{U}\}$. 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 $X, Y \subseteq \mathfrak{A}$.

	∅	A	B	C	AB	AC	BC	ABC
∅	×							
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 $\mathfrak{S}$.

□

3.4 NFDs and NMVDs in undetermined Universes

Finally, we use Theorem 3.2 and the results from (Lien 1982) to obtain an axiomatisation for NFDs and NMVDs in undetermined universes.

Theorem 3.4. *The following set of inference rules*

$$\begin{array}{ccc}
\dfrac{}{X \to Y}\; Y \subseteq X & \dfrac{X \to Y}{XU \to YV}\; V \subseteq U & \dfrac{X \to Y, X \to Z}{X \to YZ} \\[2ex]
\dfrac{X \to Y}{X \to Z}\; Z \subseteq Y & \dfrac{X \to Y}{X \twoheadrightarrow Y} & \dfrac{}{X \twoheadrightarrow Y}\; Y \subseteq X \\[2ex]
\dfrac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\; V \subseteq U & \dfrac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow YZ} & \dfrac{X \twoheadrightarrow Y, X \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}
\end{array}$$

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 emphasize 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).

4 General Results

We will use this section to show an equivalence between complete sets of inference rules in undetermined universes and complete and complementary sets in fixed universes. Due to space limitations we omit the proofs in this section. For a set $\mathfrak{S}$ of inference rules that is sound for the implication of NMVDs let $R\mathfrak{S}$ denote the set $\mathfrak{S} \cup \{\mathcal{C}_R\}$.

Theorem 4.1. *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 the set $R\mathfrak{S}$ is complete and complementary for the R-implication of NMVDs.* □

This equivalence from Theorem 4.1 can even be extended to minimal complete sets of inference rules.

Corollary 4.1. *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 the set $R\mathfrak{S}$ is complete and complementary for the R-implication of NMVDs, and there is no inference rule $\mathfrak{R} \in \mathfrak{S}$ such that the set $R(\mathfrak{S} - \{\mathfrak{R}\})$ is still both complete and complementary for the R-implication of NMVDs.* □

The last corollary helps us finding all subsets of $\{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I}, \mathcal{D}\}$ that are complete and complementary for the R-implication of NMVDs.

Lemma 4.1. *The R-complementation rule $\mathcal{C}_R$ is independent from $\mathfrak{S} = \{\mathcal{R}, \mathcal{A}, \mathcal{U}, \mathcal{I}, \mathcal{D}\}$.*

Proof. Let $R = A$, $\Sigma = \emptyset$ and $\sigma = \emptyset \twoheadrightarrow A$. Since $\sigma \notin \Sigma^+_{\mathfrak{S}}$, but $\sigma \in \Sigma^+_{\mathfrak{S} \cup \{\mathcal{C}_R\}}$ we have found witnesses R, Σ and σ for the independence of $\mathcal{C}_R$ from $\mathfrak{S}$. □

The next corollary is a consequence of Theorem 3.3, Corollary 4.1 and Lemma 4.1.

Corollary 4.2. *There are no proper subsets of $R\mathfrak{L}$ which are both complete and complementary for the R-implication of NMVDs.* □

5 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 $\dfrac{}{X \twoheadrightarrow Y}\,{\scriptstyle Y\subseteq X}$ and the augmentation rule $\dfrac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\,{\scriptstyle V\subseteq U}$ are present in $\mathfrak{L}$ one may replace the reflexivity axiom $\mathcal{R}$ by the empty-set-axiom $\mathcal{R}_\emptyset$: $\dfrac{}{\emptyset \twoheadrightarrow \emptyset}$ and still maintain completeness. In fact, the empty-set-axiom $\mathcal{R}_\emptyset$ 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 $\{\mathcal{R}_\emptyset, \mathcal{A}\}$:

$$\dfrac{\dfrac{}{\emptyset \twoheadrightarrow \emptyset}}{X \twoheadrightarrow Y}\,{\scriptstyle Y\subseteq X} \quad .$$

Theorem 5.1. *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 $U = \{A_1, \ldots, A_k\}$):

$$\dfrac{\dfrac{\begin{array}{c}\dfrac{X \twoheadrightarrow Y}{XA_1 \twoheadrightarrow Y}\\ \vdots\end{array}}{XA_1 \cdots A_k \twoheadrightarrow Y} \qquad \dfrac{}{XU \twoheadrightarrow V}\,{\scriptstyle V\subseteq U\subseteq XU}}{XU \twoheadrightarrow YV} \quad .$$

The reflexivity axiom $\mathcal{R}$ may also be replaced by the empty-set-axiom $\mathcal{R}_\emptyset$ and the attribute axiom $\mathcal{A}t$: $\dfrac{}{A \twoheadrightarrow A}$. 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 5.2. *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.* ☐

6 Conclusion

We have explored multivalued dependencies in the presence of null values (NMVDs) with meaning “undefined”, “inapplicable”, or “non-existent”. It was shown that Lien’s original axiomatisation of NMVDs (Lien 1982) is not complementary. That is, there are inferences of NMVDs in which the application of the complementation rule can neither be avoided nor deferred until the last step of the inference. The fact that the complementation rule simply reflects a part of the normalisation process is therefore not reflected by Lien’s axiomatisation. In this paper sound and complete sets of inference rules for the R-implication of NMVDs have been proposed that are indeed complementary. Moreover, Biskup’s alternative notion of implication for MVDs, 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 the implication of NMVDs have been proposed, which can be 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.

References

Arenas, M. & Libkin, L. (2004), ‘A normal form for XML documents’, *Tansactions on Database Systems (ToDS)* **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’, *ACM Tans. 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* ‘Proceedings of the SIGMOD International Conference on Management of Data’, ACM, pp. 47–61.

Bernstein, P. (1976), ‘Synthesizing third normal form relations from functional dependencies’, *ACM Tans. Database Syst.* **1**(4), 277–298.

Bernstein, P. A. & Goodman, N. (1980), What does Boyce-Codd normal form do?, *in* ‘Proceedings of the 6th International Conference on Very Large Data Bases’, IEEE Computer Society, pp. 245–259.

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

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

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', *Transactions on Database Systems (ToDS)* **4**(4), 397–434.

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

Fagin, R. (1977), 'Multivalued dependencies and a new normal form for relational databases', *ACM Tans. 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* 'Proceedings of the 14th International Conference on Very Large Data Bases (VLDB)', pp. 50–61.

Grahne, G. (1984), Dependency satisfaction in databases with incomplete information, *in* 'Proceedings of the 10th International Conference on Very Large Data Bases (VLDB)', 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. (2004), Multi-valued dependencies in the presence of lists, *in* 'Proceedings of the 23rd SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems', ACM, pp. 330–341.

Hartmann, S. & Link, S. (2006), 'On a problem of Fagin concerning multivalued dependencies in relational databases', accepted for Theoretical Computer Science (TCS).

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', *Transactions on Database Systems (ToDS)* **18**(3), 414–459.

Levene, M. & Loizou, G. (1998), 'Axiomatisation of functional dependencies in incomplete relations', *Theoretical Computer Science (TCS)* **206**(1-2), 283–300.

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

Link, S. (2006), 'On multivalued dependencies in fixed and undetermined universes', accepted for the 4th International Symposium on Foundations of Information and Knowledge Systems (FoIKS), Springer, LNCS.

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* 'Proceedings of the 3rd International Conference on Very Large Data Bases (VLDB)', IEEE Computer Society, pp. 447–453.

Tari, Z., Stokes, J. & Spaccapietra, S. (1997), 'Object normal forms and dependency constraints for object-oriented schemata', *ACM Tans. 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', *ACM Tans. Database Syst.* **17**(1), 32–64.

Wijsen, J. (1999), 'Temporal FDs on complex objects', *ACM Tans. 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', *Journal of Computer and System Sciences* **28**(1), 142–166.