Fundamenta Informaticae 92 (2009) 83–102
DOI 10.3233/FI-2009-0067
IOS Press

On Inferences of Weak Multivalued Dependencies

Sven Hartmann
Department of Informatics, Clausthal University of Technology
Clausthal-Zellerfeld, Germany
sven.hartmann@tu-clausthal.de

Sebastian Link[*]
School of Information Management, Victoria University
Wellington, New Zealand
sebastian.link@vuw.ac.nz

Abstract. Nesting is a useful technique in many areas of database practice. For instance, nesting is a fundamental operation for the nested relational data model, it can be applied to reduce the level of data redundancy in a database instance, to improve query processing or to convert data from one model to another.

We further address the question when nesting operations commute with one another, i.e., when the final nested database relation is independent of the order in which the nesting operations are applied. In fact, it has been shown that the satisfaction of weak multivalued dependencies provides a sufficient and necessary condition for the commutativity of nesting operations.

We study inference systems for different notions of implication for weak multivalued dependencies. First, we establish an axiomatisation with the property that every weak multivalued dependency can be inferred either without any application of the complementation rule or by a single application of the complementation rule necessary only in the very last step of the inference. Consequently, the complementation rule is a mere means to achieve a decomposition of the database. Secondly, we drop the assumption of having a fixed underlying schema, and establish an axiomatisation of weak multivalued dependencies for the notion of implication in this context.

Keywords: Logic in Databases, Nested Relation, Weak Multivalued Dependency, Inference System

---

[*] Address for correspondence: School of Information Management, Victoria University, Wellington, New Zealand

1. Introduction

The relational model of data [7] still defines the basis for the core technologies of most commercial database management systems. This is mainly due to its sound, easy-to-apply framework that can address a large amount of data we are interested in, but also to the familiarity and proficiency that users and administrators have accrued in dealing with relational databases.

However, the first normal form condition of the relational model of data has proven too restrictive for several applications. Complex-value data models have been proposed to overcome severe limitations of the relational data model when designing many practical database applications. One of the most studied complex-value data models is the nested relational model of data [15, 20, 26, 29, 34] in which an attribute contains either an atomic value or a nested relation. In this context the operation of nesting is fundamental [20]: data tuples that have matching values on some fixed attribute set can be represented as a single nested tuple by collecting a set of the different tuple values on the remaining attributes. We illustrate this operation by the following example.

Example 1.1. We consider the following flat relation that shows a snapshot of different dancing courses in which ladies and gentleman partner up.

Flat relation		
Course	Lady	Gentleman
Swing	Dulcinea	Quixote
Swing	Dulcinea	Sancho
Swing	Theresa	Sancho
Swing	Theresa	Don Quixote
Swing	Beatrice	Dante
Street	Eve	Tupac
Street	Queen L.	DMX
Street	Eve	DMX

The flat relation may be subject to different nesting operations whose result is illustrated in the following tables.

Nest over Lady		
Course	Lady	Gentleman
Swing	{Dulcinea,Theresa}	Quixote
Swing	{Dulcinea,Theresa}	Sancho
Swing	{Beatrice}	Dante
Street	{Eve}	Tupac
Street	{Queen L., Eve}	DMX

Nest over Gentleman		
Course	Lady	Gentleman
Swing	Dulcinea	{Quixote, Sancho}
Swing	Theresa	{Quixote, Sancho}
Street	Eve	{Tupac, DMX}
Street	Queen L.	{DMX}

Nest over Lady, then Gentleman		
Course	Lady	Gentleman
Swing	{Dulcinea,Theresa}	{Quixote,Sancho}
Swing	{Beatrice}	{Dante}
Street	{Eve}	{Tupac}
Street	{Queen L., Eve}	{DMX}

Nest over Gentleman, then Lady		
Course	Lady	Gentleman
Swing	{Dulcinea,Theresa}	{Quixote, Sancho}
Street	{Eve}	{Tupac, DMX}
Street	{Queen L.}	{DMX}

As we can see nesting is not limited to a single application.

Previous Work. The operation of nesting has conceptual advantages [32, 35], is part of the deep nested relational algebra [22, 24], can benefit the efficiency of query processing [10, 20, 21], remove data redundancy [9], yield hybrid database decompositions [10], improve information integration [32] and the conversion of data into different formats [25], but has also considerable applications in Bayesian networks [6, 41].

As we can notice from Example 1.1 the operation of nesting is not commutative in general. However, exact conditions on the original relation are known that guarantee that the resulting nested relation is invariant under different orders of consecutive nesting operations. In fact, Jaeschke and Schek [20] introduced weak multivalued dependencies (wMVDs) for this purpose. In a nutshell, the nesting of the flat relation from Example 1.1 over Lady, then Gentleman is different from its nesting over Gentleman, then Lady because the flat relation does not exhibit the wMVD *Course* $\twoheadrightarrow_w$ *Lady*. An insertion of the tuple (Street,Queen L.,Tupac) into the flat relation would result in a new relation that does satisfy this dependency, and the two different orders of nesting would result in the same nested relations.

Due to their importance and wide applicability wMVDs have been well-studied. Fischer and Van Gucht [10] have given several characterisations for the satisfiability of such dependencies, and also established a finite ground axiomatisation. The notion of a wMVD is dependent on the underlying universe R of attributes [20]. Syntactically, this dependence is represented by the R-complementation rule which is the only inference rule in the axiomatisation of wMVDs that is dependent on R [10]. This is similar to the situation for multivalued dependencies (MVDs)[2, 3, 8]. Further research on MVDs has led to an alternative notion of semantic implication in which the underlying universe is left undetermined [3]. In the same paper Biskup shows that this notion can be captured syntactically by a sound and complete set of inference rules, denoted by $\mathfrak{S}$. If $\mathfrak{S}_C$ results from adding the R-complementation rule to $\mathfrak{S}$, then $\mathfrak{S}_C$ is R-sound and R-complete for the R-implication of MVDs for all relation schemata R. In fact, for every inference of an MVD by $\mathfrak{S}_C$ there is 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 ($\mathfrak{S}_C$ is said to be R-complementary). This indicates that the R-complementation rule simply reflects a part of the normalisation process, and does not necessarily infer semantically meaningful consequences, i.e., MVDs that are valid in all possible universes. Recently, this research has been extended into different directions: for MVDs themselves [14, 27, 28], to the combined class of functional dependencies and MVDs [4], and to (the combined class of functional and) full hierarchical dependencies [4, 12].

Contributions. In this paper we extend the line of research into yet another direction. In fact, we continue the investigation of weak multivalued dependencies and the properties of their corresponding inference systems. Similar to the case of MVDs one may argue that inferences of wMVDs by the R-complementation rule may generate dependencies that are only the result of the given universe R, but do not carry any actual semantic information. For instance, every relation over $\{$*Course,Lady,Gentleman*$\}$ that satisfies *Course* $\twoheadrightarrow_w$ *Lady* will also satisfy *Course* $\twoheadrightarrow_w$ *Gentleman* but this is just a consequence of $\{$*Gentleman*$\}$ being the complement of $\{$*Course,Lady*$\}$ relatively to $\{$*Course,Lady,Gentleman*$\}$. If we alter the universe to $\{$*Course,Lady,Gentleman,Dance*$\}$, then the wMVD *Course* $\twoheadrightarrow_w$ *Lady* does not imply *Course* $\twoheadrightarrow_w$ *Gentleman*. Strictly speaking, wMVDs that are dependent on the underlying universe should not be utilized to infer further wMVDs. Indeed, if the R-complementation rule is applied in a step different from the last, then one may wonder whether the result of the inference is universally valid (valid in all universes) or only valid in the underlying universe.

Example 1.2. Suppose we fix the universe to $\{$*Course,Lady,Gentleman,Dance*$\}$, and suppose the wMVD *Course* $\twoheadrightarrow_w$ *Lady* has been specified. Consider the following inference in which rules from Table 1 are applied. We infer *Course* $\twoheadrightarrow_w$ *Course, Gentleman, Dance* from *Course* $\twoheadrightarrow_w$ *Lady* by means of the (natural) complementation rule. Subsequently, we infer *Course* $\twoheadrightarrow_w$ *Gentleman, Dance* from *Course* $\twoheadrightarrow_w$ *Course, Gentleman, Dance* by the *transport 1* rule. Finally, we infer *Course* $\twoheadrightarrow_w$ *Course, Lady* from *Course* $\twoheadrightarrow_w$ *Gentleman, Dance* by the natural complementation rule again. Since we used the (natural) complementation rule also at an inference step different from the last, this inference does not allow us to conclude whether the wMVD *Course* $\twoheadrightarrow_w$ *Course, Lady* is valid in all universes, or only valid in the universe $\{$*Course,Lady,Gentleman,Dance*$\}$.

An adequate inference system should be able to separate the wMVDs that are universally valid from those wMVDs that are only valid in the underlying universe. This line of reasoning suggests two different approaches:

1. we accept the original notion of implication that is dependent on the underlying universe. In this case, the R-complementation rule must be applied at most in the very last step of an inference. This is to ensure that all wMVDs that are universally valid can all be identified, and, subsequently, all those wMVDs that result from the fixed universe R are inferred (by a single application of the R-complementation rule, respectively) as well.

2. we reject the original notion of implication that is dependent on the underlying universe. In this case, we leave the set of underlying attributes undetermined. Intuitively, this notion should capture all those implied wMVDs that are universally valid.

In this article, we explore both approaches. In the case of fixed universes, we will show that the axiomatisation of wMVDs [10] is not R*-complementary* for all relation schemata R. That is, there are relation schemata R for which there are R-implied wMVDs that do require the application of the R-complementation rule at an inference step different from the last. Hence, the inference system in [10] is not adequate since it cannot always separate between those wMVDs that are universally valid and those which are only valid in a fixed universe. However, a slight modification of the R-complementation rule results in an inference system that is indeed R-complementary for all relation schemata R. This result shows that a clear separation is possible between wMVDs valid in all universes and wMVDs valid only in a fixed universe.

Example 1.3. Consider Example 1.2 again. The inference steps to infer *Course* $\twoheadrightarrow_w$ *Course, Lady* from *Course* $\twoheadrightarrow_w$ *Lady* can be replaced by a single application of the *transport 2* rule, cf. Table 1. It is evident that the wMVD *Course* $\twoheadrightarrow_w$ *Course, Lady* is universally valid.

Our second contribution establishes a finite axiomatisation of wMVDs for the alternative notion of implication in which the underlying set of attributes is left undetermined [3]. In that framework, inferences of possibly meaningless wMVDs are completely ruled out. For instance, the wMVD *Course* $\twoheadrightarrow_w$ *Gentleman* is not implied by *Course* $\twoheadrightarrow_w$ *Lady*.

These two contributions reveal the role of the R-complementation rule as a mere means to achieve database decompositions. This is in complete analogy to the R-complementation rule for multivalued dependencies, both in total and in partial databases (permitting null values) [3, 27, 28].

Outline. We summarise basic notions of the relational model of data and weak multivalued dependencies in Section 2. In Section 3 we illustrate the applications of wMVDs, in particular on nesting, database decompositions and Bayesian networks. We establish the first complementary axiomatisation of wMVDs for fixed universes in Section 4, and the first axiomatisation of wMVDs in undetermined universes in Section 5. We conclude in Section 6 and outline some related future research.

2. Weak Multivalued Dependencies

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 with an infinite domain $dom(A_i)$ which represents the set of possible values that can occur in the column named A_i. 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 $t : R \rightarrow \bigcup\limits_{i=1}^{n} dom(A_i)$ 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 set $Dom(r)$ of attributes over which the relation r over R is defined. Let $r[X] = \{t[X] \mid t \in r\}$ denote the *projection* of the relation r over R on $X \subseteq R$. For $X, Y \subseteq R$, $r_1 \subseteq dom(X)$ and $r_2 \subseteq dom(Y)$ let $r_1 \bowtie r_2 = \{t \in dom(XY) \mid \exists t_1 \in r_1, t_2 \in r_2 \text{ with } t[X] = t_1[X] \text{ and } t[Y] = t_2[Y]\}$ denote the *natural join* of r_1 and r_2. Note that the 0-ary relation $\{()\}$ is the projection $r[\emptyset]$ of a non-empty relation r on $\emptyset$ as well as left and right identity of the natural join operator.

A *multivalued dependency* (MVD) [8, 42] on R is an expression $X \twoheadrightarrow Y$ where $X, Y \subseteq R$. A 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 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 [8] 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 $r = r[XY] \bowtie r[X(R - Y)]$.

Weak multivalued dependencies have been introduced to characterise the commutativity of nesting operations [20]. They generalise MVDs in the sense that a relation may satisfy the wMVD $X \twoheadrightarrow_w Y$ but not the MVD $X \twoheadrightarrow Y$, but the MVD $X \twoheadrightarrow Y$ implies its associated wMVD $X \twoheadrightarrow_w Y$.

Definition 2.1. Let R be some relation schema. A *weak multivalued dependency* (wMVD) is an expression $X \twoheadrightarrow_w Y$ where $X, Y \subseteq R$. An R-relation r is said to satisfy the wMVD $X \twoheadrightarrow_w Y$, denoted by $\models_r X \twoheadrightarrow_w Y$, if and only if for all tuples $t_1, t_2, t_3 \in r$ such that $t_1[XY] = t_2[XY]$ and $t_1[X(R - Y)] = t_3[X(R - Y)]$ there is some tuple $t_4 \in r$ such that $t_4[XY] = t_3[XY]$ and $t_4[X(R - Y)] = t_2[X(R - Y)]$.

In the definition of wMVD satisfaction one may notice the dependence on the underlying set R of attributes.

Example 2.1. Suppose we have the relation schema DANCE with attributes *Course*, *Lady* and *Gentleman*. The intention is to record information about who partners up with whom in which course. Naturally, some pairs do not change their partners at all, but sometimes pairs switch. The following is an artificially small relation over DANCE.

Course	Lady	Gentleman
Latin	Racquel	Jose
Latin	Lorena	Flavio
Latin	Flor	Ruy
Latin	Flor	Flavio
Latin	Lorena	Ruy

In this relation, the pairs (Lorena,Flavio) and (Flor,Ruy) have switched at least once while the pair (Racquel,Jose) always dance together. It is evident that this relation does not satisfy the MVD *Course* $\twoheadrightarrow$ *Lady*, for instance because Racquel never dances with Flavio. However, the relation does satisfy the weak MVD *Course* $\twoheadrightarrow_w$ *Lady*.

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. Usually, the design process requires the data administrator to determine further dependencies which are logically implied by the given ones. In order to emphasise the dependence of implication on the underlying relation schema R we refer to R-implication. Let $lhs(\sigma)$ and $rhs(\sigma)$ denote the attribute sets on the left-hand side and right-hand side, respectively, of a wMVD σ, i.e., $lhs(\sigma) = X$ and $rhs(\sigma) = Y$ if σ denotes the wMVD $X \twoheadrightarrow_w Y$. Let $Attr(\sigma)$ denote the set of attributes affected by σ, i.e., $Attr(\sigma) = lhs(\sigma) \cup rhs(\sigma)$.

Definition 2.2. Let $\Sigma \cup \{\varphi\}$ be a set of wMVDs such that $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \subseteq R$. We say that Σ *R-implies* φ if and only if each relation r over R that satisfies all $\sigma \in \Sigma$ also satisfies φ.

Example 2.2. Consider again the relation schema DANCE from Example 2.1. The wMVD *Course* $\twoheadrightarrow_w$ *Lady* R-implies the wMVD *Course* $\twoheadrightarrow_w$ *Course,Lady* and the wMVD *Course* $\twoheadrightarrow_w$ *Gentleman*.

In order to determine the logical consequences of a set of wMVDs with respect to R-implication one can use the inference rules [10] from Table 1. These *inference rules* have the form

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

and inference rules without a premise are called *axioms*.

Let $\Sigma \cup \{\varphi\}$ be a set of wMVDs on the relation schema R. We will use $\mathfrak{S}$ to denote a set of inference rules. In this paper, we consider only those sets of inference rules in which either the R-complementation rule can be the only inference rule that is dependent on R or the natural R-complementation rule can be the only inference rule that is dependent on R. In particular, all sets $\mathfrak{S}$ we consider for wMVDs will

$\dfrac{}{X \twoheadrightarrow_w X}$ (reflexivity, $\mathcal{R}$)	$\dfrac{XV \twoheadrightarrow_w Y}{XV \twoheadrightarrow_w Y - V}$ (transport 1, $\mathcal{T}_1$)
$\dfrac{XV \twoheadrightarrow_w Y}{XV \twoheadrightarrow_w YV}$ (transport 2, $\mathcal{T}_2$)	$\dfrac{X \twoheadrightarrow_w Y}{XV \twoheadrightarrow_w Y}$ (augmentation, $\mathcal{A}$)
$\dfrac{X \twoheadrightarrow_w Y}{X \twoheadrightarrow_w R - XY}$ (R-complementation, $\mathcal{C}_R$)	$\dfrac{X \twoheadrightarrow_w Y}{X \twoheadrightarrow_w R - Y}$ (natural R-complementation, $\mathcal{C}_R^N$)

Table 1. Inference Rules for Weak Multivalued Dependencies

form a subset of the rule set in Table 1. The notion of syntactical inference ($\vdash_{\mathfrak{S}}$) with respect to a set $\mathfrak{S}$ of inference rules can be defined analogously to the notion in the relational data model [30]. That is, a finite sequence $\gamma = [\psi_1, \ldots, \psi_l]$ of wMVDs is called an *inference from* Σ *by* $\mathfrak{S}$ if every ψ_i is either an element of Σ or is obtained by applying one of the rules of $\mathfrak{S}$ to appropriate elements of $\{\psi_1, \ldots, \psi_{i-1}\}$. We say that the inference γ infers ψ_l, i.e., the last element of the sequence γ, and write $\Sigma \vdash_{\mathfrak{S}} \psi_l$. Let $\Sigma_{\mathfrak{S}}^+ = \{\varphi \mid \Sigma \vdash_{\mathfrak{S}} \varphi\}$ denote the *syntactic closure* of Σ under inferences by $\mathfrak{S}$. 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 rules of Table 1 are R-sound for all R [10]. The set $\mathfrak{S}$ is called *R-sound* for the R-implication of wMVDs if and only if for every set Σ of wMVDs on the relation schema R we have $\Sigma_{\mathfrak{S}}^+ \subseteq \Sigma_R^* = \{\varphi \mid \Sigma\ R\text{-implies } \varphi\}$. The set $\mathfrak{S}$ is called *R-complete* for the R-implication of wMVDs if and only if for every set Σ of wMVDs on R we have $\Sigma_R^* \subseteq \Sigma_{\mathfrak{S}}^+$. An R-complete set $\mathfrak{S}$ is said to be *R-complementary* if and only if for every set $\Sigma \cup \{\varphi\}$ of wMVDs on R the following holds: if there is an inference of a wMVD φ from Σ by $\mathfrak{S}$, then there is also an inference of φ from Σ by $\mathfrak{S}$ in which the R-complementation rule $\mathcal{C}_R$ (or the natural R-complementation rule $\mathcal{C}_R^N$, respectively) is applied at most once, and if it is applied, then it is applied in the last step of the inference. The system

$$\mathfrak{F} = \{\mathcal{R}, \mathcal{T}_1, \mathcal{T}_2, \mathcal{A}, \mathcal{C}_R\}$$

is known to be both R-sound and R-complete for the R-implication of wMVDs, for all relation schema R [10].

We will show in this paper that $\mathfrak{F}$ is not R-complementary for all relation schemata R. However, if $\mathfrak{W}_{\mathcal{C}}$ is obtained from $\mathfrak{F}$ by replacing the R-complementation rule $\mathcal{C}_R$ by the natural R-complementation rule $\mathcal{C}_R^N$, then $\mathfrak{W}_{\mathcal{C}}$ is indeed R-complementary for all relation schemata R.

In order to avoid the inference of possibly meaningless wMVDs, we can apply Biskup's alternative notion of implication [3] in which the underlying set of attributes is left undetermined. A wMVD $X \twoheadrightarrow_w Y$ with finite sets of attributes $X, Y \subseteq \mathfrak{A}$ is satisfied by some relation r if and only if $X, Y \subseteq Dom(r)$ and the following holds: for all tuples $t_1, t_2, t_3 \in r$ such that $t_1[XY] = t_2[XY]$ and $t_1[X(Dom(r) - Y)] = t_3[X(Dom(r) - Y)]$ there is some tuple $t_4 \in r$ such that $t_4[XY] = t_3[XY]$ and $t_4[X(Dom(r) - Y)] = t_2[X(Dom(r) - Y)]$.

Definition 2.3. Let $\Sigma \cup \{\varphi\}$ be a finite set of wMVDs. We say that Σ *implies* φ if and only if every relation r satisfies the following condition: if $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \subseteq Dom(r)$ and r satisfies all $\sigma \in \Sigma$, then r also satisfies φ.

The notions of *soundness* and *completeness* are simply adapted to the context of undetermined universes by dropping the reference to the underlying relation schema R from the corresponding notions in the context of fixed universes.

Let $\Sigma \cup \{\varphi\}$ be a set of wMVDs, and let R be some relation schema such that $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \subseteq R$ holds. Based on Definitions 2.2 and 2.3 it follows that Σ R-implies φ whenever Σ implies φ. However, the following Example 2.3 illustrates that the reverse direction does not hold.

Example 2.3. Consider the attributes *Course*, *Lady*, and *Gentleman* as well as the following set of wMVDs

$$\Sigma = \{\textit{Course} \twoheadrightarrow_w \textit{Lady}\}.$$

We notice that Σ $\{\textit{Course}, \textit{Lady}, \textit{Gentleman}\}$-implies *Course* $\twoheadrightarrow_w$ *Gentleman*. However, we give a counterexample relation for the implication of *Course* $\twoheadrightarrow_w$ *Gentelman* by Σ. In fact, the relation

Course	Lady	Gentleman	Dance
Latin	Lorena	Flavio	Salsa
Latin	Flor	Ruy	Rumba
Latin	Flor	Flavio	Rumba

satisfies the wMVD *Course* $\twoheadrightarrow_w$ *Lady* but violates the wMVD *Course* $\twoheadrightarrow_w$ *Gentleman* since Flavio dances with Lorena the Salsa, Ruy dances with Flor the Rumba and Flavio dances with Flor the Rumba, but Ruy does not dance with Lorena the Salsa.

Example 2.3 illustrates that the wMVD *Course* $\twoheadrightarrow_w$ *Gentleman* is not necessarily a valid consequence of *Course* $\twoheadrightarrow_w$ *Lady*.

3. Nesting and other Applications of Weak MVDs

Originally, the purpose for introducing weak multivalued dependencies was the study of nest operations in the nested relational data model [20]. Since then many other applications of wMVDs have been discussed in the literature, e.g. query processing and optimisation [20, 10, 21], deep nested relational algebra [22], database decompositions [10], information integration [32], conversion of data into different formats [25] and also for Bayesian networks [6, 41].

Nesting. Let r be a (possibly nested) relation over the relation schema R and X an attribute set with $X \subseteq R$. For each projected tuple $\pi \in r[R-X]$ we define a nested tuple t_π by:

$$t_\pi[R-X] = \pi \qquad \text{and} \qquad t_\pi[X] = \{t'[X] \mid t' \in r \text{ and } t'[R-X] = \pi\}.$$

The nested relation $\nu_X(r)$ obtained from nesting r on the attribute set X is defined as $\nu_X(r) = \{t_\pi \mid \pi \in r[R-X]\}$.

Example 3.1. Consider the *Dance*-relation r from Example 2.1. Below we have given examples of nesting several relations on various attribute sets.

r		
Course	Lady	Gentleman
Latin	Racquel	Jose
Latin	Lorena	Flavio
Latin	Flor	Ruy
Latin	Flor	Flavio
Latin	Lorena	Ruy

$\nu_{\text{Lady}}(r)$		
Course	Lady	Gentleman
Latin	{Racquel}	Jose
Latin	{Lorena, Flor}	Flavio
Latin	{Flor, Lorena}	Ruy

$\nu_{\text{Gentleman}}(r)$		
Course	Lady	Gentleman
Latin	Racquel	{Jose}
Latin	Lorena	{Flavio,Ruy}
Latin	Flor	{Ruy,Flavio}

$\nu_{\text{Gentleman}}(\nu_{\text{Lady}}(r)) = \nu_{\text{Lady}}(\nu_{\text{Gentleman}}(r))$		
Course	Lady	Gentleman
Latin	{Racquel}	{Jose}
Latin	{Lorena,Flor}	{Flavio,Ruy}

Of course, nesting can be applied consecutively.

One may have noticed that in the specific case of the previous example it did not matter whether we first nest on *Lady* and then on *Gentleman*, or vice versa. We say that the two nesting operations commute. However, this is not always the case. In fact, weak multivalued dependencies characterise the situation when two nest operations on disjoint attribute sets commute.

Theorem 3.1. [20] Let R be some relation schema and r some R-relation. Let $X, Y \subseteq R$. Then $\models_r X \twoheadrightarrow_w Y$ if and only if $\nu_Y(\nu_{R-XY}(r)) = \nu_{R-XY}(\nu_Y(r))$.

The result has been generalised to deal with an arbitrary finite number of consecutive nesting operations [37]. In undetermined universes, Theorem 3.1 reads as follows.

Theorem 3.2. Let r be some finite relation. Then $\models_r X \twoheadrightarrow_w Y$ if and only if $\nu_Y(\nu_{Dom(r)-XY}(r)) = \nu_{Dom(r)-XY}(\nu_Y(r))$.

Example 3.2. Consider again the following relation r:

Course	Lady	Gentleman	Dance
Latin	Lorena	Flavio	Salsa
Latin	Flor	Ruy	Rumba
Latin	Flor	Flavio	Rumba

Applying Theorem 3.2 we will illustrate that r satisfies the wMVD *Course* $\twoheadrightarrow_w$ *Lady* but violates the wMVD *Course* $\twoheadrightarrow_w$ *Gentleman*.

$\nu_{\text{Lady}}(r)$			
Course	Lady	Gentleman	Dance
Latin	{Lorena}	Flavio	Salsa
Latin	{Flor}	Ruy	Rumba
Latin	{Flor}	Flavio	Rumba

$\nu_{\text{Gentleman,Dance}}(\nu_{\text{Lady}}(r))$		
Course	Lady	(Gentleman,Dance)
Latin	{Lorena}	{(Flavio,Salsa)}
Latin	{Flor}	{(Ruy,Rumba), (Flavio,Rumba)}

$\nu_{\text{Gentleman,Dance}}(r)$		
Course	Lady	(Gentleman,Dance)
Latin	Lorena	{(Flavio,Salsa)}
Latin	Flor	{(Ruy,Rumba), (Flavio,Rumba)}

$\nu_{\text{Lady}}(\nu_{\text{Gentleman,Dance}}(r))$		
Course	Lady	(Gentleman,Dance)
Latin	{Lorena}	{(Flavio,Salsa)}
Latin	{Flor}	{(Ruy,Rumba), (Flavio,Rumba)}

We recognise that $\nu_{\text{Lady}}(\nu_{\text{Gentleman,Dance}}(r)) = \nu_{\text{Gentleman,Dance}}(\nu_{\text{Lady}}(r))$. That is, r satisfies the wMVD *Course* $\twoheadrightarrow_w$ *Lady*.

$\nu_{\text{Gentleman}}(r)$			
Course	Lady	Gentleman	Dance
Latin	Lorena	{Flavio}	Salsa
Latin	Flor	{Ruy,Flavio}	Rumba

$\nu_{\text{Lady,Dance}}(\nu_{\text{Gentleman}}(r))$		
Course	Gentleman	(Lady,Dance)
Latin	{Flavio}	{(Lorena,Salsa)}
Latin	{Ruy,Flavio}	{(Flor,Rumba)}

$\nu_{\text{Lady,Dance}}(r)$		
Course	Gentleman	(Lady,Dance)
Latin	Flavio	{(Lorena,Salsa), (Flor,Rumba)}
Latin	Ruy	{(Flor,Rumba}

$\nu_{\text{Gentleman}}(\nu_{\text{Lady,Dance}}(r))$		
Course	Gentleman	(Lady,Dance)
Latin	{Flavio}	{(Lorena,Salsa), (Flor,Rumba)}
Latin	{Ruy}	{(Flor,Rumba}

We recognise that $\nu_{\text{Lady,Dance}}(\nu_{\text{Gentleman}}(r)) \neq \nu_{\text{Gentleman}}(\nu_{\text{Lady,Dance}}(r))$. That is, r violates the wMVD *Course* $\twoheadrightarrow_w$ *Gentleman*.

Database Decompositions. Fischer and Van Gucht [10] give several characterisations of the satisfaction of wMVDs by a database relation. Among others, the following definition is equivalent to the satisfaction of the wMVD $X \twoheadrightarrow_w Y$ by the R-relation r: r can be partitioned horizontally into subrelations $r_1, \ldots, r_n$ such that i) r_i satisfies the multivalued dependency $X \twoheadrightarrow Y$ for $i = 1, \ldots, n$, ii) $r_i[XY] \cap r_j[XY] = \emptyset$ for $1 \leq i < j \leq n$, and iii) $r_i[X(R-Y)] \cap r_j[X(R-Y)] = \emptyset$ for $1 \leq i < j \leq n$. This suggests that a wMVD naturally captures a combination of horizontal and vertical decompositions in order to reduce data redundancy in a relation. We consider an example to illustrate this point.

Example 3.3. Consider again the DANCE-relation from Example 2.1. Take a look at its horizontal partition into r_1 and r_2:

r_1

Course	Lady	Gentleman
Latin	Racquel	Jose

r_2

Course	Lady	Gentleman
Latin	Lorena	Flavio
Latin	Flor	Ruy
Latin	Flor	Flavio
Latin	Lorena	Ruy

Both r_1 and r_2 satisfy the MVD *Course* $\twoheadrightarrow$ *Lady*, and are disjoint in their projections on $\{$Course,Lady$\}$ and $\{$Course,Gentleman$\}$. Consequently we would apply a lossless 4NF-decomposition [8] of both r_1 and r_2 obtaining

$r_1[$*Course,Lady*$]$

Course	Lady
Latin	Racquel

$r_2[$*Course,Lady*$]$

Course	Lady
Latin	Lorena
Latin	Flor

$r_1[$*Course,Gentleman*$]$

Course	Gentleman
Latin	Jose

$r_2[$*Course,Gentleman*$]$

Course	Gentleman
Latin	Flavio
Latin	Ruy

Hence, we see the immediate benefits since a lossless 4NF-decomposition is not applicable to the original relation r because it does not satisfy the MVD *Course* $\twoheadrightarrow$ *Lady*.

Weak MVDs provide a more general framework for database decompositions than MVDs: a relation r may violate an MVD but satisfy the respective wMVD. In this case r can be decomposed horizontally into subrelations each of which exhibits the MVD and can be decomposed vertically. In this sense, the role of the R-complementation rule for inferring wMVDs is similar to that for inferring MVDs: it is merely a means for database normalisation. This role should be reflected syntactically in the axiomatisation of wMVDs.

Conditional Independencies. Bayesian networks provide a semantic modeling tool which greatly facilitates the acquisition of probabilistic knowledge [31]. While multivalued dependencies allow us to decompose a database relation into two of its projections without the loss of information, Bayesian multivalued dependencies allow us to decompose a joint probability distribution into two of its marginalisations without the loss of information [41]. Consequently, the probability of an event can be obtained, in principle, by appropriate marginalisations of the joint probability distribution.

Object-oriented Bayesian networks were introduced to facilitate the construction of large Bayesian networks [23]. Similar to how the nested relational model of data allows the value of an attribute to be itself a nested relation, object-oriented Bayesian networks allow the value of a variable to be itself a distribution. In particular, *weak conditional independencies* provide a necessary and sufficient condition to obtain a unique coarsended distribution [6]. This is equivalent to weak multivalued dependencies providing a characterisation of nesting operations that commute [20].

4. Axiomatisations in Fixed Universes

We show in this section that the system $\mathfrak{F}$ [10] cannot always separate those wMVDs that are universally valid from those that are only valid in the underlying universe. In fact, the R-complementation rule $\mathcal{C}_R$ cannot fully justify its purpose for inferring only those wMVDs that result from the context of the fixed relation schema R. However, by replacing the R-complementation rule $\mathcal{C}_R$ in $\mathfrak{F}$ by the natural R-complementation rule $\mathcal{C}_R^N$ we obtain an axiomatisation for wMVDs that is indeed R-complementary for all relation schemata R.

Theorem 4.1. There is some relation schema R which $\mathfrak{F}$ is not R-complementary for.

Proof:
Fix $R = ABCD$, let $\Sigma = \{A \twoheadrightarrow_w B\}$ and $\varphi = AD \twoheadrightarrow_w CD$. We show first that φ is R-implied by Σ. In fact, consider the following inference of φ from Σ by $\mathfrak{F}$.

$$\cfrac{\cfrac{A \twoheadrightarrow_w B}{\mathcal{C}_R:\ A \twoheadrightarrow_w CD}}{\mathcal{A}:\ AD \twoheadrightarrow_w CD}$$

Since $\mathfrak{F}$ is R-sound it follows that φ is R-implied by Σ. Since φ cannot be inferred from Σ by $\mathfrak{W} = \mathfrak{F} - \{\mathcal{C}_R\}$, according to Lemma 5.1, one must apply the R-complementation rule $\mathcal{C}_R$ at least once to infer φ from Σ by $\mathfrak{F}$. However, in any inference of φ from Σ by $\mathfrak{F}$ it is not possible to apply the R-complementation rule $\mathcal{C}_R$ in the very last step of the inference since the attribute D occurs in both $lhs(\varphi)$ and $rhs(\varphi)$ but any wMVD resulting from an application of the R-complementation rule $\mathcal{C}_R$ has disjoint left- and right-hand sides. □

We will now show that the system $\mathfrak{W}_{\mathcal{C}} = \{\mathcal{R}, \mathcal{T}_1, \mathcal{T}_2, \mathcal{A}, \mathcal{C}_R^N\}$ is indeed R-complementary for all relation schemata R. Consequently, the natural R-complementation rule $\mathcal{C}_R^N$ adequately justifies its purpose in capturing all those wMVDs that result from the fixation of a relation schema.

Theorem 4.2. Let R be some relation schema, and let Σ be a set of wMVDs on R. For every inference γ from Σ by the system $\mathfrak{F}$ there is an inference ξ from Σ by the system $\mathfrak{W}_{\mathcal{C}}$ with the following properties:

- γ and ξ infer the same wMVD,
- in ξ the natural R-complementation rule $\mathcal{C}_R^N$ is applied at most once, and
- if $\mathcal{C}_R^N$ is applied in ξ, then $\mathcal{C}_R^N$ is applied as the last rule.

Proof:
The proof is done by strong induction on the length l of the inference γ. If $l = 1$, then $\xi := \gamma$ has the desired properties. Let $l > 1$, and $\gamma = [\psi_1, \ldots, \psi_l]$ be an inference from Σ by $\mathfrak{F}$ which has length l. All together, one needs to consider five cases according to which inference rule in $\mathfrak{F}$ was applied to infer ψ_l from $[\psi_1, \ldots, \psi_{l-1}]$.
Case 1. In this case ψ_l has been inferred by an application of the reflexivity axiom $\mathcal{R}$ or is an element of Σ. It follows immediately that $\xi := [\psi_l]$ has the desired properties.

Case 2. In this case ψ_l has been inferred by an application of the transport rule $\mathcal{T}_1$ to the premise ψ_i where $i < l$. Let ξ_i be obtained by applying the induction hypothesis to $\gamma_i = [\psi_1, \ldots, \psi_i]$. Consider the inference $\bar{\xi} = [\xi_i, \psi_l]$. If in ξ_i the natural R-complementation rule $\mathcal{C}_R^N$ is not applied, then $\xi := \bar{\xi}$ has the desired properties. If in ξ_i the natural R-complementation rule $\mathcal{C}_R^N$ is applied as the last rule, then $\bar{\xi}$ has the form $[\underbrace{\xi_i', XV \twoheadrightarrow_w R-Y}_{=\xi_i}, XV \twoheadrightarrow_w (R-Y)-V]$ where the last two steps are of the following form:

$$\cfrac{\cfrac{XV \twoheadrightarrow_w Y}{\mathcal{C}_R^N : \; XV \twoheadrightarrow_w R-Y}}{\mathcal{T}_1 : \; XV \twoheadrightarrow_w \underbrace{(R-Y)-V}_{=R-YV}}$$

These steps, however, can be replaced by the following steps:

$$\cfrac{\cfrac{XV \twoheadrightarrow_w Y}{\mathcal{T}_2 : \; XV \twoheadrightarrow_w YV}}{\mathcal{C}_R^N : \; XV \twoheadrightarrow_w R-YV}$$

The inference $\xi := [\xi_i', XV \twoheadrightarrow_w YV, XV \twoheadrightarrow_w R-YV]$, resulting from this replacement, meets the desired properties.

Case 3. In this case ψ_l has been inferred by an application of the transport rule $\mathcal{T}_2$ to the premise ψ_i where $i < l$. Let ξ_i be obtained by applying the induction hypothesis to $\gamma_i = [\psi_1, \ldots, \psi_i]$. Consider the inference $\bar{\xi} = [\xi_i, \psi_l]$. If in ξ_i the natural R-complementation rule $\mathcal{C}_R^N$ is not applied, then $\xi := \bar{\xi}$ has the desired properties. If in ξ_i the natural R-complementation rule $\mathcal{C}_R^N$ is applied as the last rule, then $\bar{\xi}$ has the form $[\underbrace{\xi_i', XV \twoheadrightarrow_w R-Y}_{=\xi_i}, XV \twoheadrightarrow_w (R-Y)V]$ where the last two steps are of the following form:

$$\cfrac{\cfrac{XV \twoheadrightarrow_w Y}{\mathcal{C}_R^N : \; XV \twoheadrightarrow_w R-Y}}{\mathcal{T}_2 : \; XV \twoheadrightarrow_w \underbrace{(R-Y)V}_{=R-(Y-V)}}$$

These steps, however, can be replaced by the following steps:

$$\cfrac{\cfrac{XV \twoheadrightarrow_w Y}{\mathcal{T}_1 : \; XV \twoheadrightarrow_w Y-V}}{\mathcal{C}_R^N : \; XV \twoheadrightarrow_w R-(Y-V)}$$

The inference $\xi := [\xi_i', XV \twoheadrightarrow_w Y-V, XV \twoheadrightarrow_w R-(Y-V)]$, resulting from this replacement, meets the desired properties.

Case 4. In this case ψ_l has been inferred by an application of the augmentation rule $\mathcal{A}$ to the premise ψ_i where $i < l$. Let ξ_i be obtained by applying the induction hypothesis to $\gamma_i = [\psi_1, \ldots, \psi_i]$. Consider the inference $\bar{\xi} = [\xi_i, \psi_l]$. If in ξ_i the natural R-complementation rule $\mathcal{C}_R^N$ is not applied, then ξ has the desired properties. If in ξ_i the natural R-complementation rule $\mathcal{C}_R^N$ is applied as the last rule, then $\bar{\xi}$ has the form $[\underbrace{\xi_i', X \twoheadrightarrow_w R - Y}_{=\xi_i}, XV \twoheadrightarrow_w R - Y]$ where the last two steps are of the following form:

$$\cfrac{\cfrac{X \twoheadrightarrow_w Y}{\mathcal{C}_R^N : X \twoheadrightarrow_w R - Y}}{\mathcal{A} : XV \twoheadrightarrow_w R - Y}$$

These steps, however, can be replaced by the following steps:

$$\cfrac{\cfrac{X \twoheadrightarrow_w Y}{\mathcal{A} : XV \twoheadrightarrow_w Y}}{\mathcal{C}_R^N : XV \twoheadrightarrow_w R - Y}$$

The inference $\xi := [\xi_i', XV \twoheadrightarrow_w Y, XV \twoheadrightarrow_w R - Y]$, resulting from this replacement, meets the desired properties.

Case 5. In this case ψ_l has been inferred by an application of the R-complementation rule $\mathcal{C}_R$ to the premise ψ_i where $i < l$. Let ξ_i be obtained by applying the induction hypothesis to $\gamma_i = [\psi_1, \ldots, \psi_i]$. Consider the inference $\bar{\xi} = [\xi_i, \psi_l]$. If in ξ_i the natural R-complementation rule $\mathcal{C}_R^N$ is not applied, then $\bar{\xi}$ has the form $[\xi_i, X \twoheadrightarrow_w R - XY]$ where the last step is of the following form:

$$\frac{X \twoheadrightarrow_w Y}{\mathcal{C}_R : X \twoheadrightarrow R - XY}$$

This step, however, can be replaced by the following steps:

$$\cfrac{\cfrac{X \twoheadrightarrow_w Y}{\mathcal{T}_2 : X \twoheadrightarrow_w XY}}{\mathcal{C}_R^N : X \twoheadrightarrow_w R - XY}$$

The inference $\xi := [\xi_i, X \twoheadrightarrow_w XY, X \twoheadrightarrow_w R - XY]$, resulting from this replacement, meets the desired properties.

In the remaining case, the natural R-complementation rule $\mathcal{C}_R^N$ is applied as the last rule in ξ_i. Consequently, $\bar{\xi}$ has the form $[\underbrace{\xi_i', X \twoheadrightarrow_w R - Y}_{=\xi_i}, XV \twoheadrightarrow_w R - ((R - Y)X)]$ where the last two steps are of the following form:

$$\cfrac{\cfrac{X \twoheadrightarrow_w Y}{\mathcal{C}_R^N : X \twoheadrightarrow_w R - Y}}{\mathcal{C}_R : X \twoheadrightarrow \underbrace{R - ((R - Y)X)}_{=Y-X}}$$

These steps, however, can be replaced by the following step:

$$\frac{X \twoheadrightarrow_w Y}{\mathcal{T}_1 : X \twoheadrightarrow_w Y - X}$$

The inference $\xi := [\xi_i', X \twoheadrightarrow_w Y - X]$, resulting from this replacement, meets the desired properties. This concludes the proof. $\square$

Among others Theorem 4.2 shows that $\mathfrak{W} = \mathfrak{W}_{\mathcal{C}} - \{\mathcal{C}_R^N\}$ is nearly R-complete for the R-implication of wMVDs on any relation schema R. Indeed, for every R-implied wMVD $X \twoheadrightarrow_w Y$ the system $\mathfrak{W}$ enables us to infer $X \twoheadrightarrow_w Y$ itself or $X \twoheadrightarrow_w R - Y$.

Corollary 4.1. Let $\Sigma \cup \{X \twoheadrightarrow_w Y\}$ be a finite set of wMVDs with $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup XY \subseteq R$. Then $X \twoheadrightarrow_w Y \in \Sigma^+_{\mathfrak{W}_{\mathcal{C}}}$ if and only if $X \twoheadrightarrow_w Y \in \Sigma^+_{\mathfrak{W}}$ or $X \twoheadrightarrow_w (R - Y) \in \Sigma^+_{\mathfrak{W}}$.

Another interpretation of Corollary 4.1 is the following: the fixing of a relation schema can be deferred until the very last step of an inference.

Example 4.1. Consider the relation schema DANCE from Example 2.1 and $\Sigma = \{$*Course* $\twoheadrightarrow_w$ *Lady*$\}$. It follows that *Course* $\twoheadrightarrow_w$ *Course, Lady* $\in \Sigma^+_{\mathfrak{W}}$, and *Course* $\twoheadrightarrow_w$ *Gentleman* $\notin \Sigma^+_{\mathfrak{W}}$ but *Course* $\twoheadrightarrow_w$ *Gentleman* $\in \Sigma^+_{\mathfrak{W}_{\mathcal{C}}}$. In the last inference we eventually commit ourselves to the relation schema DANCE by applying the natural DANCE-complementation rule in the final step of the inference.

5. An Axiomatisation for Undetermined Universes

In this section, we will establish the first axiomatisation for wMVDs in undetermined universes. While we have seen in Corollary 4.1 that $\mathfrak{W} = \{\mathcal{R}, \mathcal{T}_1, \mathcal{T}_2, \mathcal{A}\}$ is nearly R-complete for the R-implication on all relation schemata R, it turns out that $\mathfrak{W}$ is indeed complete for the implication of wMVDs in undetermined universes.

Before proving this result we show two technical lemmata. The correctness of the first lemma can easily be observed by inspecting the syntactic definitions of the inference rules in $\mathfrak{W}$. For each of the rules, the right-hand side of the conclusion does not contain any attribute that did not already occur in the left-hand side of the conclusion or in the right-hand side of its premise.

Lemma 5.1. Let $\Sigma \cup \{\varphi\}$ be a finite set of wMVDs. If $\varphi \in \Sigma^+_{\mathfrak{W}}$, then $rhs(\varphi) \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup lhs(\varphi)$.

Proof:
We will show the following: if $\gamma = [\psi_1, \dots, \psi_l]$ denotes an inference of $\psi_l = \varphi$ by $\mathfrak{W}$, then $rhs(\varphi) \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup lhs(\varphi)$. The proof is done by strong induction on the length l of γ.

Let $l = 1$. In this case φ has been inferred either by a single application of the reflexivity axiom $\mathcal{R}$ or $\varphi \in \Sigma$. In either case we have $rhs(\varphi) \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup lhs(\varphi)$.

Let $l > 1$. We distinguish between four different cases according to which φ is inferred from $[\psi_1, \dots, \psi_{l-1}]$ by $\mathfrak{W}$.

In the first case we assume that $\varphi \in \Sigma$ or φ has been inferred by an application of the reflexivity axiom $\mathcal{R}$. This is just the same situation as in the case where $l = 1$.

In the second case we assume that φ has been inferred by an application of the transport rule $\mathcal{T}_1$ to ψ_i with $i < l$. Then the last step of γ has the form:

$$\frac{XV \twoheadrightarrow_w Y}{XV \twoheadrightarrow_w Y - V}$$

where $\psi_i = XV \twoheadrightarrow_w Y$ and $Y \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup XV$ by induction hypothesis, and $\psi_l = XV \twoheadrightarrow Y - V$. It follows that $Y - V \subseteq Y \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup XV$.

In the third case we assume that φ has been inferred by an application of the transport rule $\mathcal{T}_2$ to ψ_i with $i < l$. Then the last step of γ has the form:

$$\frac{XV \twoheadrightarrow_w Y}{XV \twoheadrightarrow_w YV}$$

where $\psi_i = XV \twoheadrightarrow_w Y$ and $Y \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup XV$ by induction hypothesis, and $\psi_l = XV \twoheadrightarrow YV$. It follows that $YV \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup XV$.

In the fourth case we assume that φ has been inferred by an application of the augmentation rule $\mathcal{A}$ to ψ_i with $i < l$. Then the last step of γ has the form:

$$\frac{X \twoheadrightarrow_w Y}{XV \twoheadrightarrow_w Y}$$

where $\psi_i = X \twoheadrightarrow_w Y$ and $Y \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup X$ by induction hypothesis, and $\psi_l = XV \twoheadrightarrow Y$. It follows that $Y \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup XV$. This concludes the proof. □

For the next lemma notice that attributes outside of $\cup_{\sigma\in\Sigma} Attr(\sigma)$ can always be introduced only in the two final steps of the inference utilising the *augmentation rule* $\mathcal{A}$ and the *transport rule* $\mathcal{T}_2$.

Lemma 5.2. Let $\Sigma \cup \{\varphi\}$ be a finite set of wMVDs. If $\varphi \in \Sigma^+_{\mathfrak{W}}$, then there is some inference $\xi = [\psi_1, \ldots, \psi_{l+2}]$ of φ from Σ by $\mathfrak{W}$ such that every attribute occurring in $\psi_1, \ldots, \psi_l$ is an element of $\cup_{\sigma\in\Sigma} Attr(\sigma)$, and every attribute occurring in ψ_{l+1} is an element of $Attr(\varphi)$.

Proof:
For convenience let us define $T := \cup_{\sigma\in\Sigma} Attr(\sigma)$. Moreover, let $\psi \cap T$ denote the wMVD $lhs(\psi) \cap T \twoheadrightarrow_w rhs(\psi) \cap T$.

Let $\xi' = [\psi_1, \ldots, \psi_l]$ be any inference of φ from Σ by $\mathfrak{W}$. Consider the sequence

$$\bar{\xi} = [\psi_1 \cap T, \ldots, \psi_l \cap T].$$

We claim that $\bar{\xi}$ is an inference of $\varphi \cap T$ from Σ by $\mathfrak{W}$. For if ψ_i is an element of Σ, then $\psi_i \cap T = \psi_i$. Furthermore, one can easily verify that if ψ_i is the result of applying one of the rules $\mathcal{T}_1, \mathcal{T}_2, \mathcal{A}$ in ξ', then $\psi_i \cap T$ is the result of applying the same rule to the corresponding premises in $\bar{\xi}$.

According to Lemma 5.1 we have $rhs(\varphi) \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup lhs(\varphi) \subseteq T \cup lhs(\varphi)$. Hence, $rhs(\varphi) - T \subseteq lhs(\varphi)$. If φ denotes the wMVD $X \twoheadrightarrow_w Y$, then this means that $Y - T \subseteq X$ holds. This implies that we can infer φ from $X \cap T \twoheadrightarrow Y \cap T$ as follows:

$$\cfrac{\cfrac{X \cap T \twoheadrightarrow_w Y \cap T}{\mathcal{A}:\ (X \cap T) \cup X \twoheadrightarrow_w Y \cap T}}{\mathcal{T}_2:\ X \twoheadrightarrow_w \underbrace{(Y \cap T) \cup (Y - T)}_{=Y}}$$

Hence, the inference $\xi := [\bar{\xi}, X \twoheadrightarrow_w Y \cap T, X \twoheadrightarrow Y]$ has the desired properties. □

We can now establish the first axiomatisation of wMVDs in undetermined universes.

Theorem 5.1. The system $\mathfrak{W} = \{\mathcal{R}, \mathcal{T}_1, \mathcal{T}_2, \mathcal{A}\}$ is sound and complete for the implication of wMVDs.

Proof:
For all relation schemata R, the R-soundness of the inference rules in $\mathfrak{W}$ has been established in previous work [10]. According to Definitions 2.2 and 2.3 an inference rule that is R-sound for all R is also sound. Hence, the inference rules in $\mathfrak{W}$ are all sound.

For the soundness of $\mathfrak{W}$ one needs to show that every $\varphi \in \Sigma^+_{\mathfrak{W}}$ is implied by Σ. Let $T := \cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi)$. Then we need to show that every relation r that satisfies $T \subseteq Dom(r)$ and $\models_r \sigma$ for all $\sigma \in \Sigma$ also satisfies $\models_r \varphi$. We have established in Lemma 5.2 that there is an inference ξ of φ from Σ by $\mathfrak{W}$ such that $Attr(\psi) \subseteq T \subseteq Dom(r)$ holds for every ψ occurring in ξ. Since each rule of $\mathfrak{W}$ is sound we can therefore conclude by induction that each ψ occurring in ξ is satisfied by r. In particular, r also satisfies φ.

For the completeness of $\mathfrak{W}$ we assume that $\varphi \notin \Sigma^+_{\mathfrak{W}}$. Let $R \subseteq \mathfrak{A}$ be a finite set of attributes such that T is a proper subset of R, i.e., $T \subset R$. In particular, it follows that $R - Y$ is not a subset of T.

If φ denotes the wMVD $X \twoheadrightarrow_w Y$, then Lemma 5.1 shows that $X \twoheadrightarrow_w R - Y \notin \Sigma^+_{\mathfrak{W}}$ since $R - Y$ is not a subset of T. From $X \twoheadrightarrow_w Y \notin \Sigma^+_{\mathfrak{W}}$ and $X \twoheadrightarrow_w R - Y \notin \Sigma^+_{\mathfrak{W}}$ we conclude $X \twoheadrightarrow_w Y \notin \Sigma^+_{\mathfrak{W}_C}$ by Corollary 4.1. However, $\mathfrak{W}_C$ is R-complete for the R-implication of wMVDs for all relation schemata R. Hence, it follows that Σ does not R-imply φ. Consequently, Σ does not imply φ. □

Example 5.1. Let $\Sigma = \{$*Course* $\twoheadrightarrow_w$ *Lady*$\}$. The wMVD *Course* $\twoheadrightarrow_w$ *Gentleman* is not implied by Σ and, thus, not derivable by using the inference rules in $\mathfrak{W}$.

The inference system $\mathfrak{W}$ does not permit the application of the natural R-complementation rule, and does therefore not result in the inference of wMVDs that are possibly semantically meaningless.

6. Conclusion and Future Work

We have investigated inference systems for weak multivalued dependencies for two different notions of logical implication. On the one hand, we have established the first axiomatisation in fixed universes which is R-complementary for all relation schemata R. On the other hand, we have established the first axiomatisation of wMVDs in undetermined universes. This extends previous research on the class of multivalued dependencies [3, 4, 12, 27, 28].

Finally, we would like to suggest further lines of research into this direction. There exists an axiomatisation for the combined class of MVDs and wMVDs in fixed universes [11]. The system includes an R-complementation rule for MVDs, an R-complementation rule for wMVDs and the so-called *weak diminution rule*

$$\frac{XZ_1 \twoheadrightarrow_w Y, XY \twoheadrightarrow Z_1, XZ_2 \twoheadrightarrow_w Y}{X \twoheadrightarrow_w Y}$$

where $\{Y, Z_1, Z_2\}$ forms a partition of $R - X$ and R denotes the underlying universe. It represents a challenge to identify an inference system in which both R-complementation rules are the only inference rules that depend on the underlying universe, in which for every inference of an MVD there is always an inference of the same MVD in which the R-complementation rule for MVDs is at most applied at the very last step of the inference but an application of the R-complementation rule for wMVDs is not necessary at all, and in which for every inference of a wMVD there is always an inference of the same wMVD in which the R-complementation rule for wMVDs is at most applied at the very last step of the inference but an application of the R-complementation rule for MVDs is not necessary at all. It seems intuitive that this system (if it exists) without the two R-complementation rules forms an axiomatisation of MVDs and wMVDs in undetermined universes. It is also still an open problem to find an axiomatisation for the combined class of FDs, MVDs and wMVDs [11].

There are equivalences between the logical R-implication of classes of relational dependencies and classes of conditional independencies in Bayesian networks [6, 41]. It would be interesting to investigate whether these equivalences are also valid for the notion of implication in undetermined universes. Perhaps more interestingly, this notion of implication has not been studied previously for conditional independencies.

While embedded multivalued dependencies do not have a finite ground axiomatisation [18, 19] the class of so-called *conflict-free* embedded multivalued dependencies is finitely axiomatisable [31]. It would be interesting to study this class of relational dependencies from the perspective of undetermined universes.

A very interesting treatment of MVDs in the context of Entity-Relationship modeling can be found in [36]. There, the R-complete inference rules do not directly apply an R-complementation rule but make use of R's partitions into components and attributes where R denotes some relationship type. This is another way of indicating the dependence of implication on the underlying universe R. In this context it would therefore be very interesting to investigate the notion of implication in undetermined universes.

Several classes of relational dependencies such as keys, functional and multivalued dependencies have also been studied in the context of the eXtensible Markup Language XML [1, 5, 13, 16, 17, 33, 38, 39, 40]. To the best of our knowledge weak multivalued dependencies and their associated decompositions have not been investigated so far for XML.

7. Acknowledgement

This research is supported by the Marsden fund council from Government funding, administered by the Royal Society of New Zealand.

S. Hartmann is supported by a research grant of the Alfried Krupp von Bohlen and Halbach Foundation, administered by the German Scholars Organization.

References

[1] Arenas, M., Libkin, L.: A Normal Form for XML Documents, *ACM Trans. Database Syst.*, **29**(1), 2004, 195–232.

[2] Beeri, C., Fagin, R., Howard, J. H.: A Complete Axiomatization for Functional and Multivalued Dependencies in Database Relations, *SIGMOD Conference*, ACM, 1977.

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

[4] Biskup, J., Link, S.: Appropriate Reasoning about Data Dependencies in Fixed and Undetermined Universes, *FoIKS Conference*, 4932, Springer, 2008.

[5] Buneman, P., Davidson, S., Fan, W., Hara, C., Tan, W.: Reasoning about keys for XML, *Inf. Syst.*, **28**(8), 2003, 1037–1063.

[6] Butz, C., Sanscartier, M.: Properties of Weak Conditional Independence, *Rough Sets and Current Trends in Computing Conference*, 2475, Springer, 2002.

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

[8] Fagin, R.: Multivalued Dependencies and a New Normal Form for Relational Databases, *ACM Trans. Database Syst.*, **2**(3), 1977, 262–278.

[9] Fischer, P., Saxton, L. V., Thomas, S. J., Van Gucht, D.: Interactions between Dependencies and Nested Relational Structures, *J. Comput. Syst. Sci.*, **31**(3), 1985, 343–354.

[10] Fischer, P., Van Gucht, D.: Weak Multivalued Dependencies, *PoDS Conference*, ACM, 1984.

[11] van Gucht, D.: *Theory of unnormalized relational structures*, Technical Report CS-85-07, Vanderbilt University, 1985.

[12] Hartmann, S., Köhler, H., Link, S.: Full Hierarchical Dependencies in Fixed and Undetermined Universes, *Ann. Math. Artif. Intell.*, **50**(1-2), 2007, 195–226.

[13] Hartmann, S., Link, S.: More Functional Dependencies for XML, *ADBIS Conference*, number 2798 in Lecture Notes in Computer Science, Springer, 2003.

[14] Hartmann, S., Link, S.: On a Problem of Fagin concerning Multivalued Dependencies in Relational Databases, *Theor. Comput. Sci.*, **353**(1-3), 2006, 53–62.

[15] Hartmann, S., Link, S.: Characterising nested database dependencies by fragments of propositional logic, *Ann. Pure Appl. Logic*, **152**(1-3), 2007, 84–106.

[16] Hartmann, S., Link, S.: Efficient Reasoning about a Robust XML Key Fragment, *ACM Trans. Database Syst.*, **34**(1:3), 2009.

[17] Hartmann, S., Trinh, T.: Axiomatising functional dependencies for XML with frequencies, *FoIKS Conference*, number 3861 in Lecture Notes in Computer Science, Springer, 2006.

[18] Herrmann, C.: On the Undecidability of Implications Between Embedded Multivalued Database Dependencies, *Inf. Comput.*, **122**(2), 1995, 221–235.

[19] Herrmann, C.: Corrigendum to ”On the Undecidability of Implications Between Embedded Multivalued Database Dependencies”, *Inf. Comput.*, **204**(12), 2006, 1847–1851.

[20] Jaeschke, G., Schek, H.-J.: Remarks on the Algebra of Non First Normal Form Relations, *PoDS Conference*, 1982.

[21] Kambayashi, Y., Yoshikawa, M.: Query Processing utilizing dependencies and horizontal decomposition, *SIGMOD Conference*, 1983.

[22] Kitagawa, H., Kunii, T., Ohbo, N.: Classification of nested tables under deeply nested algebra, *System Sciences Conference*, 1991.

[23] Koller, D., Pfeffer, A.: Object-oriented Bayesian Networks, *UAI Conference*, Morgan Kaufmann, 1997.

[24] Korth, H., Roth, M.: Query languages for Nested Relational Databases, *Nested Relations and Complex Objects in Databases*, 361, Springer, 1991.

[25] Lee, D., Mani, M., Chiu, F., Chu, W.: NeT & CoT: translating relational schemas to XML schemas using semantic constraints, *CIKM Conference*, 2002.

[26] Levene, M., Loizou, G.: The Nested Universal Relation Data Model, *J. Comput. Syst. Sci.*, **49**(3), 1994, 683–717.

[27] Link, S.: Charting the completeness frontier of inference systems for multivalued dependencies, *Acta Inf.*, **45**(7-8), 2008, 565–591.

[28] Link, S.: On the Implication of Multivalued Dependencies in Partial Database Relations, *Int. J. Found. Comput. Sci.*, **19**(3), 2008, 691–715.

[29] Makinouchi, A.: A Consideration on Normal Form of Not-Necessarily-Normalised Relation in the Relational Data Model, *VLDB Conference*, IEEE Computer Society, 1977.

[30] Paredaens, J., De Bra, P., Gyssens, M., Van Gucht, D.: *The Structure of the Relational Database Model*, Springer, 1989.

[31] Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan-Kaufmann, 1988.

[32] Saxton, L., Raghavan, V.: Design of an Integrated Information Retrieval/Database Management System, *IEEE Trans. Knowl. Data Eng.*, **2**(2), 1990, 210–219.

[33] Saxton, L., Tang, X.: Tree Multivalued Dependencies for XML Datasets, *WAIM Conference*, number 3129 in Lecture Notes in Computer Science, Springer, 2004.

[34] Schek, H.-J., Scholl, M.: The relational model with relation-valued attributes, *Inf. Syst.*, **11**(2), 1986, 137–147.

[35] Thalheim, B.: *Entity-Relationship Modeling: Foundations of Database Technology*, Springer, 2000.

[36] Thalheim, B.: Conceptual Treatment of Multivalued Dependencies, *Conceptual Modeling Conference*, number 2813 in Lecture Notes in Computer Science, Springer, 2003.

[37] Thomas, S.: *A Non-First-Normal-Form Relational Database Model*, Ph.D. Thesis, Vanderbilt University, 1983.

[38] Vincent, M., Liu, J., Liu, C.: A redundancy free 4NF for XML, *XSym Conference*, number 2824 in Lecture Notes in Computer Science, Springer, 2003.

[39] Vincent, M., Liu, J., Liu, C.: Strong Functional Dependencies and their Application to Normal Forms in XML, *ACM Trans. Database Syst.*, **29**(3), 2004, 445–462.

[40] Vincent, M., Liu, J., Mohania, M.: On the equivalence between FDs in XML and FDs in relations, *Acta Inf.*, **44**(3-4), 2007, 207–247.

[41] Wong, S., Butz, C., Wu, D.: On the Implication Problem for Probabilistic Conditional Independency, *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, **30**(6), 2000, 785–805.

[42] Zaniolo, C., Melkanoff, M.: On the Design of Relational Database Schemata, *ACM Trans. Database Syst.*, **6**(1), 1981, 1–47.