Appropriate Reasoning about Data Dependencies in Fixed and Undetermined Universes

Joachim Biskup[1] and Sebastian Link[2,⋆]

[1] Fachbereich Informatik, Universität Dortmund, D-44221 Dortmund, Germany
biskup@ls6.cs.uni-dortmund.de
[2] Information Science Research Centre, Massey University, New Zealand
s.link@massey.ac.nz

Abstract. We study inference systems for the combined class of functional and full hierarchical dependencies in relational databases. Two notions of implication are considered: the original version in which the underlying set of attributes is fixed, and the alternative notion in which this set is left undetermined.

The first main result establishes a finite axiomatisation in fixed universes which clarifies the role of the complementation rule in the combined setting. In fact, we identify inference systems that are appropriate in the following sense: full hierarchical dependencies can be inferred without use of the complementation rule at all or with a single application of the complementation rule at the final step of the inference; and functional dependencies can be inferred without any application of the complementation rule. The second main result establishes a finite axiomatisation for functional and full hierarchical dependencies in undetermined universes.

1 Introduction

Modern database management systems provide commensurate tools to store, manage and process different kinds of data. The core of these systems still relies on the sound technology that is based on the relational model of data [15]. Relations permit the storage of inconsistent data, i.e., data that violate conditions which every legal database instance ought to satisfy. Consequently, additional assertions, called dependencies, are specified by the data administrator in order to restrict the databases to those which are considered meaningful to the application at hand. Most commercial database systems are only capable of enforcing consistency with respect to keys and foreign keys. The reason for this might be that good database design, e.g. Entity-Relationship modeling, will precisely lead to relational database schemata that are in Inclusion Dependency Normal Form [29]. During database normalisation join-related dependencies are explored

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

S. Hartmann and G. Kern-Isberner (Eds.): FoIKS 2008, LNCS 4932, pp. 58–77, 2008.

to minimise data redundancy for efficient means of updating. In practice, however, most normalised schemata are subject to denormalisation in order to facilitate the efficient processing of the most common types of queries. Therefore, any remaining join-related dependencies are examined for query optimisation. This describes the trade-off between efficient updates and efficient query processing. Hence, the quality of the target database with respect to these two criteria crucially depends on the ability to reason correctly and appropriately about such dependencies.

Full hierarchical dependencies (FHDs), called full first-order hierarchical decompositions in [16], constitute a large class of relational dependencies. A relation exhibits an FHD precisely when it is the natural join over at least two of its projections that all share the same join attributes. FHDs generalise multivalued dependencies (MVDs) in which the number of such projections is precisely two. The classical notion of an FHD [16] is dependent on the underlying universe R of attributes. For MVDs [18] their dependence on the relation schema R is syntactically reflected by the R-complementation rule which is part of the axiomatisation of MVDs [8]. The R-complementation rule is special in the sense that it is the only inference rule in this axiomatisation which is dependent on R. Further research on this fact has led to an alternative notion of semantic implication in which the underlying universe is left undetermined [12]. 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, every inference of an MVD by $\mathfrak{S}_C$ 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 ($\mathfrak{S}_C$ 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. This research has been extended recently [23, 24, 26, 30, 31].

Contributions. In this paper we analyse the completeness and appropriateness of inference systems for the *combined* class of functional dependencies (FDs) and full hierarchical dependencies in relational databases.

There are axiomatisations of multivalued and of full hierarchical dependencies for the case where the set of underlying attributes is undetermined [12, 23, 30, 32]. So far, however, no inference system has been proven complete for the combined class of FDs and MVDs, nevermind FDs and FHDs. Among other benefits, the ability to infer all implied FDs and FHDs effectively provides the data administrator with more choices on the final layout of a database schema during the design process. Without this ability the best approximation for efficiently processing the most common types of queries and the most common types of updates may remain unrevealed. Query optimisation in the presence of FDs and FHDs, such as the Chase [17], may benefit from the inference of additional dependencies, too.

Example 1. Suppose the three attributes *Movie*, *Director* and *Actor* represent information about movies titles, the name of their directors and the name of their actors. Moreover, we have specified the multivalued dependency $Movie \twoheadrightarrow Director$, stating that the set of *directors* is determined by the *movie* independently of any remaining attributes, and the functional dependency $Actor \rightarrow Director$, enforcing us to store only movie data in which no actor has acted in movies directed by different directors. For every relation schema R that contains (at least) these three attributes every R-relation r that satisfies both the MVD and FD above will also satisfy the FD $Movie \rightarrow Director$. For, if the tuples t_1, t_2 agree on *Movie*, then the MVD above guarantees that there is a tuple t that agrees with t_1 on *Movie* and *Director*, and agrees with t_2 on *Movie* and the rest of the attributes, i.e., at least on *Actor*. However, since t_2 and t agree on *Actor* the FD $Actor \rightarrow Director$ guarantees that t_2 and t also agree on *Director*. As t_1 and t also agree on *Director* it follows that t_1 and t_2 agree on *Director*. Hence, a complete set of inference rules for undetermined universes must enable us to infer the FD $Movie \rightarrow Director$ that is implied by $Movie \twoheadrightarrow Director$ and $Actor \rightarrow Director$. □

Moreover, we analyse the appropriateness of existing inference systems for the combined class of FDs and MVDs in fixed universes. In particular, we will clarify the role of the R-complementation rule in the combined setting. It turns out that some systems do not properly reflect the semantics of neither FDs nor MVDs, some properly reflect the semantics of MVDs but not of FDs and some systems properly reflect the semantics of FDs but not of MVDs. In particular, there are systems that require the application of the R-complementation rule in order to infer some implied FDs. Intuitively, an R-*adequate* inference system should be able to avoid such cases since the definition of an FD is independent of the underlying universe. Consequently, if R-*adequate* inference systems do not exist at all, then applications of the R-complementation rule may result in inferences of semantically meaningless functional dependencies.

Example 2. Suppose we fix the relation schema Film={*Movie*,*Director*,*Actor*} together with the MVD and FD from Example 1. We will show that there are axiomatisations in which the Film-complementation rule must be applied in order to infer the FD $Movie \rightarrow Director$. Therefore, one may argue that the inferred FD is actually meaningless since this line of reasoning is inadequate.

□

However, we will identify an inference system that does properly reflect the semantics of both FDs and MVDs in fixed universes, i.e., it is R-complementary and R-adequate for all relation schemata R. Strictly speaking, it is only this fact (i.e. that such an inference system does exist) that confirms our intuition that inferences by previously established axiomatisations always result in meaningful functional and multivalued dependencies. Therefore, there is no need to distrust the semantics of data dependencies that are inferred by complete yet inappropriate inference systems since, by our results, any inappropriate inference can be converted into an appropriate one.

Example 3. Consider the relation schema FILM as well as the set Σ of FDs and MVDs from Example 2 again. The *mixed subset rule* allows us to infer the FD $X \rightarrow Y \cap Z$ from the MVD $X \twoheadrightarrow Y$ and the FD $W \rightarrow Z$ in case that Y and W are disjoint. For example, the FD *Movie* $\rightarrow$ *Director* can be inferred from *Title* $\twoheadrightarrow$ *Director* and *Actor* $\rightarrow$ *Director* by means of the mixed subset rule, i.e., without using the FILM-complementation rule. □

Our results also confirm the intuition that the R-complementation rule is a mere means of database normalisation even in the combined setting of FDs and MVDs. Finally, we extend these results to the combined class of functional and full hierarchical dependencies.

2 Dependencies in Relational Databases

Let $\mathcal{A} = \{A_1, A_2, \ldots\}$ be a (countably) infinite set of symbols, called *attributes*. A *relation schema* is a finite set R of distinct *attributes* from $\mathcal{A}$, which represent column names of a relation. Each attribute A of a relation schema is associated an infinite domain $dom(A)$ which represents the set of possible values that can occur in the column named A. 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 the relation schema R (R-tuple or simply tuple, if R is understood) is a function $t : R \rightarrow \bigcup_{A \in R} dom(A)$ with $t(A) \in dom(A)$ for all $A \in R$. 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. Let $r[X] = \{t[X] \mid t \in r\}$ denote the *projection* of the relation r over R on $X \subseteq R$. For $X, Y \subseteq R$, finite $r_1 \subseteq dom(X)$ and $r_2 \subseteq dom(Y)$ let $r_1 \bowtie r_2 = \{t \in dom(XY) \mid \exists t_1 \in r_1, t_2 \in r_2$ with $t[X] = t_1[X]$ 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 any non-empty relation r on $\emptyset$ as well as left and right identity of the natural join operator.

Functional dependencies (FDs) between sets of attributes have played a central role in the study of relational databases [7, 9, 10, 14, 15], and seem to be central for the study of database design in other data models as well [1, 22, 25, 28, 37, 41, 42]. 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 [2, 3]. A *functional dependency* (FD) [15] on the relation schema R is an expression $X \rightarrow Y$ where $X, Y \subseteq R$. A relation r over R *satisfies* the FD $X \rightarrow Y$, denoted by $\models_r X \rightarrow Y$, if and only if every pair of tuples in r that agrees on each of the attributes in X also agrees on the attributes in Y. That is, $\models_r X \rightarrow Y$ if and only if $t_1[Y] = t_2[Y]$ whenever $t_1[X] = t_2[X]$ holds for any $t_1, t_2 \in r$.

FDs are incapable of modelling many important properties that database users have in mind. Multivalued dependencies (MVDs) provide a more general notion and offer a response to the shortcomings of FDs. A *multivalued dependency* (MVD) [18, 43] 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 $t[X(R - XY)] = t_2[X(R - XY)]$. 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 $R - XY$. This actually suggests that the relation schema R is overloaded in the sense that it carries two independent facts XY and $X(R-XY)$. More precisely, it is shown in [18] 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 $r = r[XY] \bowtie r[X(R - XY)]$. This characteristic of MVDs is fundamental to relational database design and 4NF [18]. A lot of research has therefore been devoted to studying the behaviour of these dependencies [4, 5, 6, 8, 11, 12, 19, 20, 21, 24, 26, 27, 32, 34, 35, 36, 39, 40]. Full hierarchical dependencies generalise multivalued dependencies [16, 23].

Definition 1. *A* full hierarchical dependency *(FHD) on a relation schema R is an expression $X : S$ where $X \subseteq R$ and S is a non-empty set of pairwise disjoint subsets of R that are also disjoint from X, i.e., $S \neq \emptyset$, for all $Y \in S$ we have $Y \subseteq R$ and for all $Y, Z \in S \cup \{X\}$ we have $Y \cap Z = \emptyset$. An R-relation $r \subseteq dom(R)$ is said to* satisfy *(or said to be a* model *of) the full hierarchical dependency $X : \{Y_1, \ldots, Y_k\}$ on R, denoted by $\models_r X : \{Y_1, \ldots, Y_k\}$, if and only if for all $t_1, \ldots, t_{k+1} \in r$ the following condition is satisfied: if $t_i[X] = t_j[X]$ for all $1 \leq i, j \leq k + 1$, then there is some $t \in r$ such that $t[XY_i] = t_i[XY_i]$ for $i = 1, \ldots, k$ and $t[X(R - XY_1 \cdots Y_k)] = t_{k+1}[X(R - XY_1 \cdots Y_k)]$.* □

Notice that Definition 1 reduces to the definition of MVDs in case that $k = 1$. Note that our definition of FHDs is slightly different from what Delobel originally introduced as full first-order hierarchical decompositions [16]. We prefer the form given above for the sake of simplifying the axiomatisation and emphasising the correspondence to the definition of MVDs. The following result is a straightforward generalisation from the MVD case [18].

Theorem 1. *Let $X, Y_1, \ldots, Y_k \subseteq R$ be pairwise disjoint and $k \geq 1$. An R-relation r satisfies the FHD $X : \{Y_1, \ldots, Y_k\}$ on R if and only if $r = r[XY_1] \bowtie \cdots \bowtie r[XY_k] \bowtie r[X(R - XY_1 \cdots Y_k)]$.* □

Recall that every FHD $X : \{Y_1, \ldots, Y_k\}$ is equivalent to the set $\{X \twoheadrightarrow Y_1, \ldots, X \twoheadrightarrow Y_k\}$ of MVDs [23]. With this in mind functional and full hierarchical dependencies with the same left-hand side permit a *stepwise* decomposition of the underlying relation schema by splitting one current component into two components. Hence, the left-hand side X represents exponentially many decompositions since no order is enforced in which the MVDs $X \twoheadrightarrow Y_i$ are to be applied. This feature distinguishes hierarchical dependencies from general join dependencies which do not have this property [38].

The tree obtained from any such stepwise decomposition provides the data administrator with two kinds of choices on the final layout of the relational database schema. Firstly, the set of relation schemata associated with a full decomposition tree does not permit any data redundancy with respect to the FHDs and, therefore, facilitates efficient updating for the broadest class of updates. However, in order to process queries efficiently several joining operations may be necessarily enforced. Secondly, a truncation of a selected tree at inner nodes of the decomposition tree represents some level of denormalisation that may meet the user's preferences for the level of i) efficient query processing and ii) efficient updating. Complete and appropriate reasoning about the class of functional and full hierarchical dependencies enables the administrator to infer all meaningful dependencies all of which can then be effectively used to determine the final database schema that best approximates the user's preferences for efficiently processing the most common type of queries and updates.

Example 4. Consider the four attributes *Article*, *Manufacturer*, *Location*, and *Costs* representing information about manufacturers that supply articles from their location at a certain cost. Suppose Σ consists of already identified functional dependencies $Article \rightarrow Manufacturer$ and $Article, Location \rightarrow Costs$ and the multivalued dependency $Manufacturer \twoheadrightarrow Location$. The target database is supposed to process efficiently updates on *Costs* based on the *Article*, e.g., values that result from queries such as

$$\pi_{\text{Costs}}(\sigma_{\text{Article=MP3-Player}}(\{Article, Costs\})).$$

Moreover, the target database is most commonly subject to queries about *Article,Location*-information of certain manufacturers such as

$$\pi_{\text{Article, Location}}(\sigma_{\text{Manufacturer=Sony}}(\{Article, Manufacturer, Location\})).$$

Notice that the FD $Article \rightarrow Costs$ is implied by Σ. If it is also a reasonable semantic constraint, then a good choice for a target schema would appear to have the two relation schemata $\{Article, Costs\}$ and $\{Article, Manufacturer, Location\}$. The first relation schema enables efficient updates, and the second one offers efficient query processing of our most common types of queries (no joining necessary). The question if the FD $Article \rightarrow Costs$ does indeed represent an appropriate semantic constraint will be further investigated in Examples 5 and 10. □

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. As just explained, 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 dependency σ, i.e., $lhs(\sigma) = X$ and $rhs(\sigma) = Y_1 \cdots Y_k$ if σ denotes the FHD $X : \{Y_1, \ldots, Y_k\}$, and $lhs(\sigma) = X$ and $rhs(\sigma) = Y$ if σ denotes the FD $X \rightarrow Y$. Let $Attr(\sigma)$ denote the set of attributes affected by σ, i.e., $Attr(\sigma) = lhs(\sigma) \cup rhs(\sigma)$.

Definition 2. *Let* $\Sigma \cup \{\varphi\}$ *be a set of FDs and FHDs 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. □

In order to determine the logical consequences of a set of FDs and MVDs with respect to R-implication one can use the inference rules [8, 11, 12] from Table 1. These *inference rules* have the form

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

and inference rules without a premise are called *axioms*. Let $\Sigma \cup \{\sigma\}$ be a set

Table 1. Inference Rules for Functional and Multivalued Dependencies

$\dfrac{}{X \to Y}\; Y \subseteq X$ (reflexivity, $\mathcal{R}_{\mathrm{F}}$)	$\dfrac{X \to Y}{X \to XY}$ (extension, $\mathcal{E}_{\mathrm{F}}$)
$\dfrac{X \to Y, Y \to Z}{X \to Z}$ (transitivity, $\mathcal{T}_{\mathrm{F}}$)	$\dfrac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\; V \subseteq U$ (augmentation, $\mathcal{A}_{\mathrm{M}}$)
$\dfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}$ (R-complementation, $\mathcal{C}^R_{\mathrm{M}}$)	$\dfrac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$ (pseudo-transitivity, $\mathcal{T}_{\mathrm{M}}$)
$\dfrac{X \twoheadrightarrow Y, W \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}\; Y \cap W = \emptyset$ (subset, $\mathcal{S}_{\mathrm{M}}$)	$\dfrac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow YZ}$ (additive transitivity, $\mathcal{T}^*_{\mathrm{M}}$)
$\dfrac{X \to Y}{X \twoheadrightarrow Y}$ (implication, $\mathcal{I}_{\mathrm{FM}}$)	$\dfrac{X \twoheadrightarrow Y, Y \to Z}{X \to Z - Y}$ (mixed pseudo-transitivity, $\mathcal{T}_{\mathrm{FM}}$)
$\dfrac{X \twoheadrightarrow Y, W \to Z}{X \to Y \cap Z}\; Y \cap W = \emptyset$ (mixed subset, $\mathcal{S}_{\mathrm{FM}}$)	

of FDs and FHDs (FDs and MVDs) on the relation schema R. Furthermore, we use $\mathfrak{S}$ to denote a set of inference rules. In this paper we consider only those sets of inference rules in which the R-complementation rule can be the only inference rule that is dependent on R. In particular, all sets $\mathfrak{S}$ we consider for FDs and MVDs will form a subset of the rule set in Table 1. Let $\Sigma \vdash_{\mathfrak{S}} \sigma$ denote the inference of σ from Σ with respect to $\mathfrak{S}$. Let $\Sigma^+_{\mathfrak{S}} = \{\sigma \mid \Sigma \vdash_{\mathfrak{S}} \sigma\}$ denote the *syntactic hull* of Σ under inference using only rules from $\mathfrak{S}$. An inference rule is called *R-sound* if the set of dependencies in the premise of the rule R-implies the dependency in the conclusion. The rules of Table 1 are R-sound for all R [8,13]. The set $\mathfrak{S}$ is called *R-sound* for the R-implication of FDs and FHDs if and only if for every set Σ of FDs and FHDs on the relation schema R we have $\Sigma^+_{\mathfrak{S}} \subseteq \Sigma^*_R = \{\sigma \mid \Sigma \ R\text{-implies } \sigma\}$. The set $\mathfrak{S}$ is called *R-complete* for the R-implication of FDs and FHDs if and only if for every set Σ of FDs and FHDs on R we have $\Sigma^*_R \subseteq \Sigma^+_{\mathfrak{S}}$. An inference rule $\mathfrak{R}$ is said to be *independent* of the set $\mathfrak{S}$ if and only if there is some relation schema R and some finite set $\Sigma \cup \{\varphi\}$ of FDs and FHDs on R such that $\varphi \notin \Sigma^+_{\mathfrak{S}}$, but $\varphi \in \Sigma^+_{\mathfrak{S}\cup\{\mathfrak{R}\}}$. An R-complete set $\mathfrak{S}$ is said to be *R-complementary* if and only if for every set $\Sigma \cup \{\varphi\}$ of FDs and FHDs on R the inference of an FHD φ from Σ using $\mathfrak{S}$ can be turned into an inference of φ from Σ using $\mathfrak{S}$ in which the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$ is applied at most once, and if it is applied, then it is applied in the last step of the inference. In what follows we use $\mathfrak{S}_{\mathcal{C}}$ to denote the inference system obtained from the system $\mathfrak{S}$ by adding the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$. The system

$$\mathfrak{F}_{\mathcal{C}} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{T}_{\mathrm{FM}}, \mathcal{C}^R_{\mathrm{M}}\}$$

is known to be both R-sound and R-complete for the R-implication of FDs and MVDs, for all relation schema R [33]. However, $\mathfrak{F}_{\mathcal{C}}$ is not R-complementary for all relation schemata R [12]. Biskup conjectured that the set

$$\mathfrak{A}\mathfrak{C}_{\mathcal{C}} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{S}_{\mathrm{M}}, \mathcal{T}^*_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{S}_{\mathrm{FM}}, \mathcal{C}^R_{\mathrm{M}}\}$$

is indeed R-complementary for the R-implication of FDs and MVDs for all relation schemata R [12]. We will formally prove that this conjecture is indeed true. Moreover, we will verify that $\mathfrak{A}\mathfrak{C}_{\mathcal{C}}$ enjoys a further property which makes it appropriate for reasoning about the combined class of FDs and MVDs. A system $\mathfrak{S}_{\mathcal{C}}$ of inference rules that is R-complete for the R-implication of FDs and MVDs is said to be *R-adequate* if for every set Σ of FDs and MVDs on R and every FD φ on R such that φ is R-implied by Σ there is an inference of φ from Σ by $\mathfrak{S}$, i.e., an inference in which the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$ is not utilised at all. When denoting inference systems we will use the letter $\mathfrak{A}$ to indicate the adequacy of the system, and the letter $\mathfrak{C}$ to indicate its complementarity.

Example 5. Consider the attributes *Article*, *Manufacturer*, *Location*, and *Costs* and dependency set $\Sigma = \{A \to M; A, L \to C; M \twoheadrightarrow L\}$ from Example 4 again. A possible inference of the FD $A \to C$ from Σ is the following:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \to M}{A \twoheadrightarrow M}\,\mathcal{I}_{\mathrm{FM}} \qquad M \twoheadrightarrow L}{A \twoheadrightarrow L}\,\mathcal{T}_{\mathrm{M}}}{A \twoheadrightarrow A, L}\,\mathcal{A}_{\mathrm{M}}}{A \twoheadrightarrow C, M}\,\mathcal{C}^R_{\mathrm{M}} \qquad A, L \to C}{A \to C}\,\mathcal{S}_{\mathrm{FM}}$$

One may argue that this line of reasoning is inappropriate since this inference applies the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$ in order to infer the FD $A \to C$.

Consequently, this inference may leave us in doubt whether the FD $A \rightarrow C$ is actually meaningful for our application. □

In order to avoid the inference of possibly meaningless MVDs, Biskup [12] introduced the alternative notion of implication in which the underlying set of attributes is left undetermined.

Definition 3. *Let $\Sigma \cup \{\varphi\}$ be a set of FDs and FHDs. 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. While there are axiomatisations for the class of MVDs [12, 30] and the class of FHDs [23, 38] in undetermined universes, no axiomatisation is known for the combined class of FDs and MVDs nor FDs and FHDs.

Let $\Sigma \cup \{\varphi\}$ be a set of FDs and FHDs, 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 and 3 it follows that Σ R-implies φ whenever Σ implies φ. Intuitively, the reverse direction also holds when φ is an FD, but the following Example 6 illustrates that the reverse direction does not hold when φ is an FHD.

Example 6. Let Film={Movie,Director,Actor} and

$$\Sigma = \{\text{Movie} \twoheadrightarrow \text{Director}, \text{Actor} \rightarrow \text{Director}\}.$$

While Σ Film-implies Movie $\twoheadrightarrow$ Actor it is relatively simple to give a counterexample for the implication of Movie $\twoheadrightarrow$ Actor by Σ. □

Before concluding this section we present a result on the interaction of FDs and MVDs in undetermined universes. In fixed universes the following fact is well-known. Let R denote a relation schema, and Σ a set of FDs on R. Then Σ R-implies the MVD $X \twoheadrightarrow Y$ if and only if Σ R-implies the FD $X \rightarrow Y$ or the FD $X \rightarrow R - Y$. This fact takes an even more convincing form in undetermined universes.

Theorem 2. *Let Σ be a set of functional dependencies. Then Σ implies the MVD $X \twoheadrightarrow Y$ if and only if Σ implies the FD $X \rightarrow Y$.*

Proof. If Σ implies the FD $X \rightarrow Y$, then, by soundness of the implication rule $\mathcal{I}_{\text{FM}}$, Σ also implies the MVD $X \twoheadrightarrow Y$.

Suppose Σ does not imply the FD $X \rightarrow Y$. This means that Y is not a subset of the attribute closure $X^*_\Sigma = \cup\{B \mid X \rightarrow B \in \Sigma^*\}$ of X under Σ, i.e., $Y - X^*_\Sigma$ is non-empty. Construct a relation r such that $Dom(r)$ is the disjoint union of X^*_Σ, $Y - X^*_\Sigma$ and a new attribute A, and r consists of exactly two tuples t, t' such that $t[C] = t'[C]$ if and only if $C \in X^*_\Sigma$. It is simple to observe that r satisfies Σ and violates $X \twoheadrightarrow Y$ as there is no tuple in r which agrees with t on $X^*_\Sigma Y$ and agrees with t' on $X^*_\Sigma A$. Consequently, Σ does not imply the MVD $X \twoheadrightarrow Y$. □

3 Inadequate Reasoning in Fixed Universes

Functional dependencies are satisfied by relations independently of the corresponding set of underlying attributes. Since this property should be properly reflected syntactically we may ask that inference systems for FDs and MVDs are R-adequate for all relation schemata R.

We will show in this section that adequacy of an inference system cannot be taken for granted. Let

$$\mathfrak{C} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{S}_{\mathrm{M}}, \mathcal{T}^*_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{T}_{\mathrm{FM}}\}$$

denote the system that is obtained from $\mathfrak{F}$ by adding $\mathcal{S}_{\mathrm{M}}$ and $\mathcal{T}^*_{\mathrm{M}}$.

Lemma 1. *The* mixed subset rule $\mathcal{S}_{FM}$ *is independent of* $\mathfrak{C}$.

Proof. Let $\Sigma = \{\emptyset \twoheadrightarrow A, B \rightarrow A\}$ and $\varphi = \emptyset \rightarrow A$. Neglecting all trivial FDs and MVDs with attributes not in AB we represent the closure $\Sigma^+_{\mathfrak{C}}$ of Σ with respect to $\mathfrak{C}$ as two tables. The MVD $X \twoheadrightarrow Y$ (FD $X \rightarrow Y$) belongs to $\Sigma^+_{\mathfrak{C}}$ if and only if in the $\twoheadrightarrow$-table ($\rightarrow$-table) the entry in row labelled X and column labelled Y is a cross $\times$. The $\rightarrow$-table can be obtained as follows. First, we apply $\mathcal{R}_{\mathrm{F}}$ to infer all trivial FDs with attributes in AB. Subsequently, we enter the premise $B \rightarrow A$ from Σ. Finally, we apply $\mathcal{E}_{\mathrm{F}}$ to infer $B \rightarrow AB$ from $B \rightarrow A$. The $\twoheadrightarrow$-table can be obtained as follows. First, we apply $\mathcal{I}_{\mathrm{FM}}$ to copy all $\times$ from the $\rightarrow$-table into the corresponding entries in the $\twoheadrightarrow$-table. Finally, we enter the premise $\emptyset \twoheadrightarrow A$ from Σ. This set is closed under inference using $\mathfrak{C}$. In particular, φ cannot be inferred from Σ by using $\mathfrak{C}$. In fact, one can observe that both premises in Σ are necessary to infer φ. The only inference rule capable of inferring φ from Σ is $\mathcal{T}_{\mathrm{FM}}$, but in order to apply this rule the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$ must first be applied to $\emptyset \twoheadrightarrow A$. However, $\mathcal{C}^R_{\mathrm{M}}$ is not available in $\mathfrak{C}$.

$\rightarrow$	$\emptyset$	A	B	AB
$\emptyset$	×			
A	×	×		
B	×	×	×	×
AB	×	×	×	×

$\twoheadrightarrow$	$\emptyset$	A	B	AB
$\emptyset$	×	×		
A	×	×		
B	×	×	×	×
AB	×	×	×	×

It follows that $\varphi \notin \Sigma^+_{\mathfrak{C}}$ but $\varphi \in \Sigma^+_{\mathfrak{C}\cup\{\mathcal{S}_{\mathrm{FM}}\}}$. □

The next lemma shows that the system $\mathfrak{C}_{\mathcal{C}}$ is inadequate, and thus, also the inadequacy of $\mathfrak{F}_{\mathcal{C}}$.

Lemma 2. *There is a relation schema R, a set Σ of FDs and MVDs on R and an FD φ on R such that $\varphi \in \Sigma^+_{\mathfrak{C}_{\mathcal{C}}}$ but $\varphi \notin \Sigma^+_{\mathfrak{C}}$.*

Proof. Let $R = AB$, and $\Sigma = \{\emptyset \twoheadrightarrow A, B \rightarrow A\}$ and $\varphi = \emptyset \rightarrow A$. The proof of Lemma 1 has shown that $\varphi \notin \Sigma^+_{\mathfrak{C}}$. It therefore remains to verify that $\varphi \in \Sigma^+_{\mathfrak{C}_{\mathcal{C}}}$. First, we apply $\mathcal{C}^R_{\mathrm{M}}$ to $\emptyset \twoheadrightarrow A$ to infer $\emptyset \twoheadrightarrow B$. Subsequently, we apply $\mathcal{T}_{\mathrm{FM}}$ to $\emptyset \twoheadrightarrow B$ and $B \rightarrow A$ and infer $\emptyset \rightarrow A$. □

Corollary 1. *The systems $\mathfrak{F}_\mathcal{C}$ and $\mathfrak{C}_\mathcal{C}$ are not R-adequate for the reasoning about FDs and MVDs for all relation schemata R.* □

Lemma 2 raises the question whether there is any adequate (or even appropriate) set of inference rules for the R-implication of FDs and MVDs. We will show in Section 4 that $\mathfrak{A}_\mathcal{C} = (\mathfrak{F}_\mathcal{C} - \{\mathcal{T}_{\mathrm{FM}}\}) \cup \{\mathcal{S}_{\mathrm{FM}}\}$ is adequate and $\mathfrak{AC}_\mathcal{C} = \mathfrak{A}_\mathcal{C} \cup \{\mathcal{S}_{\mathrm{M}}, \mathcal{T}^*_{\mathrm{M}}\}$ is indeed appropriate. Intuitively, the *mixed subset rule* $\mathcal{S}_{\mathrm{FM}}$ allows us to infer those FDs directly that otherwise had to be inferred by using the mixed pseudo-transitivity rule $\mathcal{T}_{\mathrm{FM}}$ and the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$. This is very much similar to the role of the subset rule $\mathcal{S}_{\mathrm{M}}$ which allows us to infer those MVDs directly that otherwise had to be inferred by using the pseudo-transitivity rule $\mathcal{T}_{\mathrm{M}}$ and the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$.

Example 7. Let FILM={Movie,Director,Actor} and

$$\Sigma = \{\text{Movie} \twoheadrightarrow \text{Director}, \text{Actor} \rightarrow \text{Director}\}.$$

The FD Movie → Actor can be inferred from Σ by first applying the FILM-complementation rule to Movie ↠ Director in order to infer Movie ↠ Actor and then applying the mixed pseudo-transitivity rule to Movie ↠ Actor and Actor → Director. However, this line of reasoning is inadequate since the FILM-complementation rule is utilised to infer a functional dependency. □

4 Appropriate Reasoning in Fixed Universes

Our first main result establishes

$$\mathfrak{AC}_\mathcal{C} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{S}_{\mathrm{M}}, \mathcal{T}^*_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{S}_{\mathrm{FM}}, \mathcal{C}^R_{\mathrm{M}}\}$$

as the first inference system for the R-implication of FDs and MVDs that is indeed appropriate: R-sound, R-complete, R-complementary and R-adequate for all relation schemata R. In particular, this clarifies the role of the R-complementation rule as a mere means of database normalisation in the combined setting of functional and multivalued dependencies.

Lemma 3. *The* mixed pseudo-transitivity rule $\mathcal{T}_{FM}$ *can be inferred from the following set of inference rules* $\{\mathcal{T}_M, \mathcal{I}_{FM}, \mathcal{S}_{FM}\}$.

Proof

$$\mathcal{S}_{\mathrm{FM}}: \dfrac{\mathcal{T}_{\mathrm{M}}: \dfrac{X \twoheadrightarrow Y \qquad \mathcal{I}_{\mathrm{FM}}: \dfrac{Y \rightarrow Z}{Y \twoheadrightarrow Z}}{X \twoheadrightarrow Z - Y} \qquad Y \rightarrow Z}{X \rightarrow \underbrace{(Z - Y) \cap Z}_{= Z - Y}}$$

This completes the proof. □

Lemma 3 enables us to obtain another axiomatisation of FDs and MVDs in fixed universes: just replace $\mathcal{T}_{\mathrm{FM}}$ in $\mathfrak{F}_\mathcal{C}$ by $\mathcal{S}_{\mathrm{FM}}$.

Corollary 2. *For all relation schemata R, the system*

$$\mathfrak{A}_{\mathcal{C}} = \{\mathcal{R}_F, \mathcal{E}_F, \mathcal{T}_F, \mathcal{A}_M, \mathcal{T}_M, \mathcal{I}_{FM}, \mathcal{S}_{FM}, \mathcal{C}^R_M\}$$

is R-sound and R-complete for the R-implication of FDs and MVDs. □

Theorem 3. *Let R be some relation schema, and let Σ be a set of FDs and MVDs on R. For every inference γ from Σ by the system $\mathfrak{A}_{\mathcal{C}}$ there is an inference ξ from Σ by the system*

$$\mathfrak{AC}_{\mathcal{C}} = \mathfrak{A}_{\mathcal{C}} \cup \{\mathcal{S}_M, \mathcal{T}^*_M\}$$

with the following properties:

1. *if γ infers an MVD, then*
   - *γ and ξ infer the same MVD,*
   - *in ξ the R-complementation rule $\mathcal{C}^R_M$ is applied at most once, and*
   - *if $\mathcal{C}^R_M$ is applied in ξ, then $\mathcal{C}^R_M$ is applied as the last rule.*
2. *if γ infers an FD, then*
   - *γ and ξ infer the same FD, and*
   - *in ξ the R-complementation rule $\mathcal{C}^R_M$ is not applied at all.*

Proof (Sketch). The proof is done by induction on the length l of the inference γ. 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 $\mathfrak{A}_{\mathcal{C}}$ which has length l. All together, one needs to consider eight cases according to which inference rule in $\mathfrak{A}_{\mathcal{C}}$ was applied to infer σ_l from $[\sigma_1, \ldots, \sigma_{l-1}]$. However, we will only show the most interesting case in which the *mixed subset rule* $\mathcal{S}_{\mathrm{FM}}$ eliminates an application of the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$ during an inference of a functional dependency. Therefore, we assume that σ_l is inferred by applying the *mixed subset rule* $\mathcal{S}_{\mathrm{FM}}$ to the premises σ_i and σ_j with $i, j < l$. Let ξ_i (ξ_j) be obtained by using the induction hypothesis for $\gamma_i := [\sigma_1, \ldots, \sigma_i]$ ($\gamma_j := [\sigma_1, \ldots, \sigma_j]$). Consider the inference $\xi := [\xi_i, \xi_j, \sigma_l]$. Then we distinguish between two cases according to the occurrence of the R-complementation rule $\mathcal{C}^R_{\mathrm{M}}$ in ξ_i (assuming that ξ_j infers the FD in the premise). If $\mathcal{C}^R_{\mathrm{M}}$ is not applied in ξ_i, then ξ has the desired properties. It remains to consider the case where $\mathcal{C}^R_{\mathrm{M}}$ is applied in ξ_i (as the last rule), i.e., the last step of ξ_i and the last step of ξ are of the following form:

$$\cfrac{\mathcal{C}^R_{\mathrm{M}}:\ \cfrac{X \twoheadrightarrow Y}{X \twoheadrightarrow R-Y} \qquad\qquad W \to Z}{\mathcal{S}_{\mathrm{FM}}:\qquad X \to \underbrace{(R-Y)\cap Z}_{=Z-Y}}\ {\scriptstyle (R-Y)\cap W=\emptyset.}$$

Since $(R-Y)\cap W = \emptyset$ holds we have $W \subseteq Y$. Hence, these steps can be replaced as follows:

$$\cfrac{X \twoheadrightarrow Y \qquad \cfrac{\cfrac{}{\mathcal{R}_{\mathrm{F}}:\ Y \to W}\ {\scriptstyle W\subseteq Y} \qquad W \to Z}{\mathcal{T}_{\mathrm{F}}:\qquad Y \to Z}}{\mathcal{T}_{\mathrm{FM}}:\qquad X \to Z - Y}$$

The proof of Lemma 3 shows how the last application of this inference can be replaced by an inference that only uses rules in $\mathfrak{AC}$ (even in $\mathfrak{A}$ already). The result of this replacement is an inference with the desired properties. The cases in which the remaining inference rules are applied can be dealt with similarly.

□

Recall that the system $\mathfrak{C}_{\mathcal{C}} = \mathfrak{F}_{\mathcal{C}} \cup \{\mathcal{S}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}^*\}$ is R-complementary but not R-adequate for all relation schemata R [12]. On the other hand, the proof of Theorem 3 shows that the system $\mathfrak{A}_{\mathcal{C}}$ is already R-adequate for all relation schemata R, but similar to the case of $\mathfrak{F}_{\mathcal{C}}$ it can be shown that $\mathfrak{A}_{\mathcal{C}}$ is not R-complementary for all relation schemata R [12].

Corollary 3. *The inference system* $\mathfrak{C}_{\mathcal{C}}$ *satisfies* 1. *of Theorem 3, but not* 2. *The inference system* $\mathfrak{A}_{\mathcal{C}}$ *satisfies* 2. *of Theorem 3, but not* 1. □

Figure 1 illustrates the connection between the different inference systems and their semantic properties. In summary, one gains complementarity by including the subset rule $\mathcal{S}_{\mathrm{M}}$ and additive transitivity rule $\mathcal{T}_{\mathrm{M}}^*$, and adequacy by including the mixed subset rule $\mathcal{S}_{\mathrm{FM}}$.

$$\mathfrak{AC}_{\mathcal{C}} = \mathfrak{A}_{\mathcal{C}} \cup \{\mathcal{S}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}^*\} = (\mathfrak{C}_{\mathcal{C}} - \{\mathcal{T}_{\mathrm{FM}}\}) \cup \{\mathcal{S}_{\mathrm{FM}}\}$$

appropriate: adequate and complementary

$$\mathfrak{A}_{\mathcal{C}} = (\mathfrak{F}_{\mathcal{C}} - \{\mathcal{T}_{\mathrm{FM}}\}) \cup \{\mathcal{S}_{\mathrm{FM}}\} \qquad \mathfrak{C}_{\mathcal{C}} = \mathfrak{F}_{\mathcal{C}} \cup \{\mathcal{S}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}^*\}$$

adequate | complementary

$$\mathfrak{F}_{\mathcal{C}} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{T}_{\mathrm{FM}}, \mathcal{C}_{\mathrm{M}}^R\}$$

Fig. 1. Inference Systems and their Properties

Notice that Theorem 3 shows that no R-complete inference system can infer any semantically meaningless FD or MVD since any inappropriate inference of such a dependency can always be converted into an appropriate inference by $\mathfrak{AC}_{\mathcal{C}}$.

5 Nearly Complete Reasoning in Fixed Universes

Among others Theorem 3 shows that $\mathfrak{AC}$ is nearly R-complete for the R-implication of FDs and MVDs on any relation schema R. Indeed, $\mathfrak{AC}$ enables us to infer every R-implied FD. Moreover, for every R-implied MVD $X \twoheadrightarrow Y$ the system $\mathfrak{AC}$ enables us to infer $X \twoheadrightarrow Y$ itself or $X \twoheadrightarrow R - Y$.

Corollary 4. *Let $\Sigma \cup \{\varphi\}$ be a finite set of FDs and MVDs with $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \subseteq R$. Then*

– *If φ denotes an FD, then: $\varphi \in \Sigma^+_{\mathfrak{AC}_C}$ if and only if $\varphi \in \Sigma^+_{\mathfrak{AC}}$.*
– *If φ denotes the MVD $X \twoheadrightarrow Y$, then: $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{AC}_C}$ if and only if $X \twoheadrightarrow Y \in \Sigma^+_{\mathfrak{AC}}$ or $X \twoheadrightarrow (R-Y) \in \Sigma^+_{\mathfrak{AC}}$.* □

Another interpretation of Corollary 4 is the following: if $\mathfrak{AC}$ is utilised to infer FDs, then the underlying universe does not need to be fixed at all; and if $\mathfrak{AC}$ is utilised to infer MVDs, then the fixing of a universe can be deferred until the very last step of the inference.

Example 8. Let Film={Movie,Director,Actor} and

$$\Sigma = \{\text{Movie} \twoheadrightarrow \text{Director}, \text{Actor} \to \text{Director}\}.$$

It follows that Movie→Director $\in \Sigma^+_{\mathfrak{AC}}$, and Movie $\twoheadrightarrow$ Actor $\notin \Sigma^+_{\mathfrak{AC}}$ (cf. Lemma 4) but Movie $\twoheadrightarrow$ Actor $\in \Sigma^+_{\mathfrak{AC}_C}$. In the last inference we eventually commit ourselves to the relation schema Film by applying the Film-complementation rule in the final (only) step of the inference. □

6 Complete Reasoning in Undetermined Universes

In this section we establish the first axiomatisation for the combined class of FDs and MVDs in undetermined universes. While we have seen in the previous section that $\mathfrak{AC}$ is nearly R-complete for the R-implication on all relation schemata R it turns out that $\mathfrak{AC}$ is indeed complete for the implication of FDs and MVDs in undetermined universes.

Before sketching the proof we shall mention a lemma. Its correctness can easily be observed by inspecting the syntactic definitions of the inference rules in $\mathfrak{AC}$. 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 at least one of the premises.

Lemma 4. *Let $\Sigma \cup \{\varphi\}$ be a finite set of FDs and MVDs. If $\varphi \in \Sigma^+_{\mathfrak{AC}}$, then $rhs(\varphi) \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \cup lhs(\varphi)$.* □

We are now prepared to establish the first axiomatisation of FDs and MVDs in undetermined universes.

Theorem 4. *The set $\mathfrak{AC} = \{\mathcal{R}_F, \mathcal{E}_F, \mathcal{T}_F, \mathcal{A}_M, \mathcal{T}_M, \mathcal{S}_M, \mathcal{T}^*_M, \mathcal{I}_{FM}, \mathcal{S}_{FM}\}$ is sound and complete for the implication of FDs and MVDs in undetermined universes.*

Proof (Sketch). The soundness of the inference rules in $\mathfrak{AC}$ has been established in previous work [12, 13]. For the soundness of $\mathfrak{AC}$ one needs to show that every $\varphi \in \Sigma^+_{\mathfrak{AC}}$ is implied by Σ. That is, every relation r that satisfies $T := \cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \subseteq Dom(r)$ and $\models_r \sigma$ for all $\sigma \in \Sigma$ also satisfies

$\models_r \varphi$. One can show that there is an inference γ of φ from Σ by $\mathfrak{AC}$ such that $Attr(\psi) \subseteq T \subseteq Dom(r)$ holds for every ψ occurring in γ. Since each rule of $\mathfrak{AC}$ is sound we can therefore conclude by induction that each ψ occurring in γ is satisfied by r. In particular, r also satisfies φ.

For the completeness of $\mathfrak{AC}$ we assume that $\varphi \notin \Sigma^+_{\mathfrak{AC}}$. Let $R \subseteq \mathcal{A}$ 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 a functional dependency, then Corollary 4 shows that $\varphi \notin \Sigma^+_{\mathfrak{AC}_\mathcal{C}}$. However, $\mathfrak{AC}_\mathcal{C}$ is R-complete for the R-implication of FDs and MVDs. Hence, it follows that Σ does not R-imply φ. Consequently, Σ does not imply φ.

If φ denotes the multivalued dependency $X \twoheadrightarrow Y$, then Lemma 4 shows that $X \twoheadrightarrow R - Y \notin \Sigma^+_{\mathfrak{AC}}$ since $R - Y$ is not a subset of T. From $X \twoheadrightarrow Y \notin \Sigma^+_{\mathfrak{AC}}$ and $X \twoheadrightarrow R - Y \notin \Sigma^+_{\mathfrak{AC}}$ we conclude $X \twoheadrightarrow Y \notin \Sigma^+_{\mathfrak{AC}_\mathcal{C}}$ by Corollary 4. However, $\mathfrak{AC}_\mathcal{C}$ is R-complete for the R-implication of FDs and MVDs. Hence, it follows that Σ does not R-imply φ. Consequently, Σ does not imply φ. □

Theorem 4 proves the conjecture of Biskup in [12] for the combined class of FDs and MVDs.

Example 9. Let $\Sigma = \{\text{Movie} \twoheadrightarrow \text{Director}, \text{Actor} \rightarrow \text{Director}\}$. The multivalued dependency Movie $\twoheadrightarrow$ Actor is not implied by Σ and, thus, not derivable by using the inference rules in $\mathfrak{AC}$. Moreover, the FD Movie $\rightarrow$ Director is indeed implied by Σ and, consequently, also derivable from Σ by using $\mathfrak{AC}$. □

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

Example 10. Consider the attributes *<u>A</u>rticle*, *<u>M</u>anufacturer*, *<u>L</u>ocation*, and *<u>C</u>osts* and dependency set $\Sigma = \{A \rightarrow M; A, L \rightarrow C; M \twoheadrightarrow L\}$ from Example 4 again. Recall, that the inference from Example 5 has left us in doubt about the meaningfulness of the FD $A \rightarrow C$. The inference

$$\mathcal{S}_{\text{FM}}:\ \dfrac{\mathcal{T}_{\text{M}}:\ \dfrac{\mathcal{A}_{\text{M}}:\ \dfrac{\mathcal{T}_{\text{M}}:\ \dfrac{\mathcal{I}_{\text{FM}}:\ \dfrac{A \rightarrow M}{A \twoheadrightarrow M} \quad M \twoheadrightarrow L}{A \twoheadrightarrow L}}{A \twoheadrightarrow A, L} \quad \mathcal{I}_{\text{FM}}:\ \dfrac{A, L \rightarrow C}{A, L \twoheadrightarrow C}}{A \twoheadrightarrow C} \quad A, L \rightarrow C}{A \rightarrow C}$$

exhibits an adequate line of reasoning, the FD $A \rightarrow C$ is an appropriate consequence of Σ, and the target schema suggested in Example 4 represents therefore an excellent choice. □

7 Extension to Full Hierarchical Dependencies

Complete reasoning techniques for the class of full hierarchical dependencies have only been established independently of any other class of dependencies [23, 38].

In this section we extend the results from previous sections to the combined class of FDs and FHDs. The main difference between MVDs and FHDs is that the latter permit several (disjoint) attribute sets on the right-hand side while MVDs only permit a single attribute set. One can observe that the inference rules for FDs and FHDs are similar to those for FDs and MVDs. What is required additionally, are inference rules that deal with the interactions of the several attribute sets on the right-hand side of an FHD. In fact, the merging of two attribute sets (by $\mathcal{M}_{\mathrm{H}}$) and the removal of an arbitrary attribute set (by $\mathcal{O}_{\mathrm{H}}$) capture these interactions completely. Further differences between the inference rules mainly result from the disjointness that we require (by definition) for attribute sets of an FHD while we do not require disjoint left- and right-hand sides in an MVD.

Table 2. Inference Rules for Functional and Full Hierarchical Dependencies

$$\frac{X:\{Y_1,\ldots,Y_k\}}{XZ:\{Y_1-Z,\ldots,Y_k-Z\}}$$
(augmentation, $\mathcal{A}_{\mathrm{H}}$)

$$\frac{XY:\{Y_1,\ldots,Y_k\},X:\{Y\}}{X:\{Y_1,\ldots,Y_k,Y\}}$$
(transitivity, $\mathcal{T}_{\mathrm{H}}$)

$$\frac{X:\{Y_1,\ldots,Y_k,Y\}}{X:\{Y_1,\ldots,Y_k\}}$$
(omission, $\mathcal{O}_{\mathrm{H}}$)

$$\frac{X:\{Y_1,\ldots,Y_k,Y_{k+1}\}}{X:\{Y_1,\ldots,Y_kY_{k+1}\}}$$
(merging, $\mathcal{M}_{\mathrm{H}}$)

$$\frac{X:\{Y_1,\ldots,Y_k\}}{X:\{Y_1,\ldots,Y_{k-1},R-XY_1\cdots Y_k\}}$$
(R-complementation, $\mathcal{C}_{\mathrm{H}}^R$)

$$\frac{X\rightarrow Y}{X:\{Y-X\}}$$
(implication, $\mathcal{I}_{\mathrm{FH}}$)

$$\frac{X:\{Y\},W\rightarrow Z}{X\rightarrow Y\cap Z}\;Y\cap W=\emptyset$$
(mixed subset, $\mathcal{S}_{\mathrm{FH}}$)

$$\frac{X:\{Y\},XY\rightarrow Z}{X\rightarrow Z-Y}$$
(mixed pseudo-transitivity, $\mathcal{T}_{\mathrm{FH}}$)

$$\frac{X:\{Y\},W:\{Y_1,\ldots,Y_k\}}{X:\{Y\cap Y_1,\ldots,Y\cap Y_k,Y-Y_1\cdots Y_k\}}\;Y\cap W=\emptyset$$
(subset, $\mathcal{S}_{\mathrm{H}}$)

To the authors' best knowledge Theorem 5 establishes the first axiomatisation for the combined class of FDs and FHDs extending a result for FHDs only [23].

Theorem 5. *For all relation schemata R, the inference systems*

- $\mathfrak{H}\mathfrak{F}_{\mathcal{C}}=\{\mathcal{R}_F,\mathcal{E}_F,\mathcal{T}_F,\mathcal{A}_H,\mathcal{T}_H,\mathcal{O}_H,\mathcal{I}_{FH},\mathcal{T}_{FH},\mathcal{C}_H^R\}$,
- $\mathfrak{H}\mathfrak{C}_{\mathcal{C}}=\mathfrak{H}\mathfrak{F}_{\mathcal{C}}\cup\{\mathcal{S}_H,\mathcal{M}_H\}$,
- $\mathfrak{H}\mathfrak{A}_{\mathcal{C}}=(\mathfrak{H}\mathfrak{F}_{\mathcal{C}}-\{\mathcal{T}_{FH}\})\cup\{\mathcal{S}_{FH}\}$, *and*
- $\mathfrak{H}\mathfrak{A}\mathfrak{C}_{\mathcal{C}}=\mathfrak{H}\mathfrak{A}_{\mathcal{C}}\cup\{\mathcal{S}_H,\mathcal{M}_H\}=(\mathfrak{H}\mathfrak{C}_{\mathcal{C}}-\{\mathcal{T}_{FH}\})\cup\{\mathcal{S}_{FH}\}$.

are R-sound and R-complete for the R-implication of FDs and FHDs. □

The next result shows that the systems for FDs and FHDs have the same properties as their counterparts for FDs and MVDs (cf. Figure 1). The definitions of R-adequacy and R-complementarity are easily extended to FDs and FHDs.

Theorem 6. *For all relation schemata R, the inference system*

- $\mathfrak{H}\mathfrak{C}_{\mathcal{C}}$ *is R-complementary,*
- $\mathfrak{H}\mathfrak{A}_{\mathcal{C}}$ *is R-adequate, and*
- $\mathfrak{H}\mathfrak{A}\mathfrak{C}_{\mathcal{C}}$ *is R-complementary and R-adequate*

for the R-implication of FDs and FHDs. □

Example 11. We illustrate how an application of the subset rule can shift applications of the R-complementation rule. Suppose we have the following inference

$$\cfrac{\mathcal{C}^R_{\mathrm{H}}: \cfrac{X:\{Y\}}{X:\{R-XY\}} \quad X(R-XY):\{Y_1,\ldots,Y_k\}}{\mathcal{T}_{\mathrm{H}}: \quad X:\{Y_1,\ldots,Y_k,R-XY\}}$$

Applying the subset rule $\mathcal{S}_{\mathrm{H}}$ instead one may infer the same FHD as follows:

$$\cfrac{\mathcal{S}_{\mathrm{H}}: \cfrac{X:\{Y\} \qquad X(R-XY):\{Y_1,\ldots,Y_k\}}{X:\{Y_1\cap Y,\ldots,Y_k\cap Y, Y-Y_1\cdots Y_k\}}}{\mathcal{C}^R_{\mathrm{H}}: \quad X:\{Y_1,\ldots,Y_k,\underbrace{R-XY_1\cdots Y_k(Y-Y_1\cdots Y_k)}_{=R-XYY_1\cdots Y_k=R-XY}\}}$$

Note that Y and $X(R-XY)$ are disjoint, and $Y_1\cdots Y_k$ and $X(R-XY)$ are disjoint, too. Hence, $Y_1\cdots Y_k\subseteq Y$, and thus $Y_i\cap Y=Y_i$ for all $i=1,\ldots,k$. □

Example 12. We illustrate how an application of the mixed subset rule can eliminate applications of the R-complementation rule. Consider the inference:

$$\cfrac{\mathcal{C}^R_{\mathrm{H}}: \cfrac{X:\{Y\}}{X:\{R-XY\}} \qquad X(R-XY)\to Z}{\mathcal{T}_{\mathrm{FH}}: \quad X\to\underbrace{Z-X(R-XY)}_{=Y\cap Z}}$$

For this inference notice that X and Y are disjoint. Hence, $X(R-XY)=R-Y$ and, consequently, $Z-(R-Y)=Y\cap Z$. Applying the mixed subset rule $\mathcal{S}_{FH}$ instead one may infer the same FHD by the following inference step:

$$\cfrac{X:\{Y\} \qquad X(R-XY)\to Z}{\mathcal{S}_{\mathrm{FH}}: \quad X\to Y\cap Z}$$

This eliminates the application of the R-complementation rule. □

In undetermined universes the following result extends Theorem 4.

Theorem 7. *The system $\mathfrak{H}\mathfrak{A}\mathfrak{C}$ of inference rules is sound and complete for the implication of FDs and FHDs in undetermined universes.* □

8 Conclusion

We have extended previous research on the appropriateness of inference systems for MVDs to the combined class of FDs and FHDs. In particular, we have established the first appropriate axiomatisation of FDs and FHDs in fixed universes, and the first ever axiomatisation of FDs and FHDs in undetermined universes. Our results demonstrate that the complementation rule is a mere means for achieving database normalisation: to infer an FHD at most one application of the complementation rule is necessary in the very last step of the inference; and to infer an FD the complementation rule does not need to be applied at all. Most importantly, we have formally demonstrated that previous axiomatisations for the class of FDs and FHDs (FDs and MVDs) cannot infer any semantically meaningless data dependencies since any inappropriate inference can be converted into an appropriate inference.

A very interesting treatment of MVDs and FHDs in the context of Entity-Relationship modeling can be found in [39]. 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.

References

1. Arenas, M., Libkin, L.: A normal form for XML documents. ACM Trans. Database Syst. 29(1), 195–232 (2004)
2. Armstrong, W.W.: Dependency structures of database relationships. Information Processing 74, 580–583 (1974)
3. Armstrong, W.W., Nakamura, Y., Rudnicki, P.: Armstrong's axioms. Journal of formalized Mathematics 14 (2002)
4. Balcázar, J., Baixeries, J.: Characterizations of multivalued dependencies and related expressions. In: Suzuki, E., Arikawa, S. (eds.) DS 2004. LNCS (LNAI), vol. 3245, pp. 306–313. Springer, Heidelberg (2004)
5. Balcázar, J., Baixeries, J.: Characterization and armstrong relations for degenerate multivalued dependencies using formal concept analysis. In: Ganter, B., Godin, R. (eds.) ICFCA 2005. LNCS (LNAI), vol. 3403, pp. 162–175. Springer, Heidelberg (2005)
6. Beeri, C.: On the membership problem for functional and multivalued dependencies in relational databases. ACM Trans. Database Syst. 5(3), 241–259 (1980)
7. Beeri, C., Bernstein, P.A.: Computational problems related to the design of normal form relational schemata. ACM Trans. Database Syst. 4(1), 30–59 (1979)
8. Beeri, C., Fagin, R., Howard, J.H.: A complete axiomatization for functional and multivalued dependencies in database relations. In: SIGMOD Conference, pp. 47–61. ACM, New York (1977)
9. Bernstein, P.: Synthesizing third normal form relations from functional dependencies. ACM Trans. Database Syst. 1(4), 277–298 (1976)
10. Bernstein, P.A., Goodman, N.: What does Boyce-Codd normal form do? In: VLDB Conference, pp. 245–259. IEEE Computer Society, Los Alamitos (1980)

11. Biskup, J.: On the complementation rule for multivalued dependencies in database relations. Acta Inf. 10(3), 297–305 (1978)
12. Biskup, J.: Inferences of multivalued dependencies in fixed and undetermined universes. Theor. Comput. Sci. 10(1), 93–106 (1980)
13. Biskup, J.: Grundlagen von Informationssystemen. Vieweg (1995)
14. Biskup, J., Dayal, U., Bernstein, P.: Synthesizing independent database schemas. In: SIGMOD Conference, pp. 143–151 (1979)
15. Codd, E.F.: A relational model of data for large shared data banks. Commun. ACM 13(6), 377–387 (1970)
16. Delobel, C.: Normalisation and hierarchical dependencies in the relational data model. ACM Trans. Database Syst. 3(3), 201–222 (1978)
17. Deutsch, A., Popa, L., Tannen, V.: Query reformulation with constraints. SIGMOD Record 35(1), 65–73 (2006)
18. Fagin, R.: Multivalued dependencies and a new normal form for relational databases. ACM Trans. Database Syst. 2(3), 262–278 (1977)
19. Fischer, P.C., Saxton, L.V., Thomas, S.J., Van Gucht, D.: Interactions between dependencies and nested relational structures. J. Comput. Syst. Sci. 31(3), 343–354 (1985)
20. Galil, Z.: An almost linear-time algorithm for computing a dependency basis in a relational database. J. ACM 29(1), 96–102 (1982)
21. Hagihara, K., Ito, M., Taniguchi, K., Kasami, T.: Decision problems for multivalued dependencies in relational databases. SIAM J. Comput. 8(2), 247–264 (1979)
22. Hara, C., Davidson, S.: Reasoning about nested functional dependencies. In: PODS Conference, pp. 91–100. ACM, New York (1999)
23. Hartmann, S., Köhler, H., Link, S.: Full hierarchical dependencies in fixed and undetermined universes. Ann. Math. Artif. Intell. 50(1–2), 195–226 (2007)
24. Hartmann, S., Link, S.: On a problem of Fagin concerning multivalued dependencies in relational databases. Theor. Comput. Sci. 353(1–3), 53–62 (2006)
25. Hartmann, S., Link, S., Schewe, K.-D.: Axiomatisations of functional dependencies in the presence of records, lists, sets and multisets. Theor. Comput. Sci. 355(2), 167–196 (2006)
26. Hartmann, S., Link, S., Schewe, K.-D.: Functional and multivalued dependencies in nested databases generated by record and list constructor. Ann. Math. Artif. Intell. 46(1–2), 114–164 (2006)
27. Lakshmanan, V., VeniMadhavan, C.: An algebraic theory of functional and multivalued dependencies in relational databases. Theor. Comput. Sci. 54, 103–128 (1987)
28. Levene, M., Loizou, G.: Axiomatisation of functional dependencies in incomplete relations. Theor. Comput. Sci. 206(1–2), 283–300 (1998)
29. Levene, M., Vincent, M.: Justification for inclusion dependency normal form. IEEE Trans. Knowl. Data Eng. 12(2), 281–291 (2000)
30. Link, S.: On multivalued dependencies in fixed and undetermined universes. In: Dix, J., Hegner, S.J. (eds.) FoIKS 2006. LNCS, vol. 3861, pp. 257–276. Springer, Heidelberg (2006)
31. Link, S.: On the logical implication of multivalued dependencies with null values. In: Twelfth Computing: The Australasian Theory Symposium. Conferences in Research and Practice in Information Technology, vol. 51, pp. 113–122 (2006)
32. Mendelzon, A.: On axiomatising multivalued dependencies in relational databases. J. ACM 26(1), 37–44 (1979)
33. Paredaens, J., De Bra, P., Gyssens, M., Van Gucht, D.: The Structure of the Relational Database Model. Springer, Heidelberg (1989)

34. Sagiv, Y.: An algorithm for inferring multivalued dependencies with an application to propositional logic. J. ACM 27(2), 250–262 (1980)
35. Sagiv, Y., Delobel, C., Parker Jr., D.S., Fagin, R.: An equivalence between relational database dependencies and a fragment of propositional logic. J. ACM 28(3), 435–453 (1981)
36. Sözat, M., Yazici, A.: A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations. ACM Fuzzy Sets and Systems 117(2), 161–181 (2001)
37. Tari, Z., Stokes, J., Spaccapietra, S.: Object normal forms and dependency constraints for object-oriented schemata. ACM Trans. Database Syst. 22, 513–569 (1997)
38. Thalheim, B.: Dependencies in Relational Databases. Teubner-Verlag (1991)
39. Thalheim, B.: Conceptual treatment of multivalued dependencies. In: Song, I.-Y., Liddle, S.W., Ling, T.-W., Scheuermann, P. (eds.) ER 2003. LNCS, vol. 2813, pp. 363–375. Springer, Heidelberg (2003)
40. Vincent, M., Liu, J., Liu, C.: A redundancy free 4NF for XML. In: Bellahsène, Z., Chaudhri, A.B., Rahm, E., Rys, M., Unland, R. (eds.) Database and XML Technologies. LNCS, vol. 2824, pp. 254–266. Springer, Heidelberg (2003)
41. Weddell, G.: Reasoning about functional dependencies generalized for semantic data models. ACM Trans. Database Syst. 17(1), 32–64 (1992)
42. Wijsen, J.: Temporal FDs on complex objects. ACM Trans. Database Syst. 24(1), 127–176 (1999)
43. Zaniolo, C., Melkanoff, M.: On the design of relational database schemata. ACM Trans. Database Syst. 6(1), 1–47 (1981)