CDMTCS Research Report Series

Reasoning about functional and full hierarchical dependencies over partial relations

Flavio Ferrarotti
School of Information Management,
The Victoria University of Wellington,
Wellington, New Zealand

Sven Hartmann
Department of Informatics,
Clausthal University of Technology,
Clausthal-Zellerfeld, Germany

Sebastian Link
Department of Computer Science,
University of Auckland,
Auckland, New Zealand

CDMTCS-449
December 2013
Centre for Discrete Mathematics and
Theoretical Computer Science

Reasoning about functional and full hierarchical dependencies over partial relations

Flavio Ferrarotti
School of Information Management
The Victoria University of Wellington
Wellington, New Zealand
flavio.ferrarotti@vuw.ac.nz

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

Sebastian Link
Department of Computer Science
The University of Auckland, Private Bag 92019
Auckland, New Zealand
s.link@auckland.ac.nz

December 17, 2013

Abstract

We study the implication problem for the combined class of functional and full hierarchical dependencies in the presence of SQL's NOT NULL constraints. Two different notions of implication are addressed: one where a dependency is implied by the given set of dependencies plus the underlying schema, and one where a dependency is implied by the given set of dependencies alone. We establish axiomatizations for both notions of implication, and reveal deep relationships between them.

Keywords: Axiomatization, Database, Full hierarchical dependency, Functional dependency, Partial relation

1 Introduction

Context. 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 Codd's relational model of data [22]. 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. According to [32] the class of functional dependencies (FDs) captures around two-thirds, and the class of full hierarchical dependencies (FHDs) around one-quarter of all uni-relational dependencies (those defined over a single relation schema) that arise in practice. In particular, FHDs are frequently exhibited in database applications [99], e.g. after de-normalization or in views [1]. Research on data dependencies has been extensive in the context of the relational model of data, see [14, 39, 89] for excellent surveys. Most of this research centers around the implication problem for classes of data dependencies. The problem is to decide for an arbitrarily given finite set $\Sigma \cup \{\varphi\}$ of data dependencies of a fixed class, if every relation that satisfies all elements of Σ also satisfies φ. Solutions to the implication problem are essential to database design, and fundamental to data processing tasks such as updates and queries. New application areas include data integration and exchange as well as database security. One of the most important extensions of Codd's relational model is incomplete information. This is mainly due to the high demand for the correct handling of such information in real-world applications. Approaches to deal with incomplete information comprise partial relations [24, 60, 69], or-relations [59, 92], fuzzy relations [85], rough sets [101] and probabilistic relations [26]. In this article we are interested in the implication problem of the combined class of functional and full hierarchical dependencies over partial relations that use Zaniolo's "no information" null value, denoted by `ni` [100]. The following example discusses an instance of this problem.

Example 1 *Consider a relation schema* DVD *with column headers* M(ovie), D(irector), A(ctor), F(eature) *and* L(anguage). *The schema collects information about DVDs, i.e., the title of the movie on the DVD together with the names of the movie's directors and actors, the features and languages available on the DVD. An example of a partial relation over* DVD *is*

Movie	*Director*	*Actor*	*Feature*	*Language*
The girl with the dragon tattoo	***ni***	*R. Mara*	*Commentary*	*English*
The girl with the dragon tattoo	*N.A. Oplev*	***ni***	*Subtitle*	*Swedish*

*where the null value **ni** indicates that no information is available about the director of the first movie, and no information is available about the actor of the second movie, respectively. Suppose the database management system enforces the following semantically meaningful constraints. The FD $MA \rightarrow D$ says that the director is uniquely determined by the title and an actor of the movie. That means, every pair of tuples that has the same non-null value on M and the same non-null value on A also has the same value on D (possibly the null value). The FHD $M : \{DF\}$ says that the sets of directors and features* *are determined by the title of the movie independently of the actors and languages. That means that for every pair of tuples that has the same non-null value on M there must be a tuple that has the same values on M, D and F as the first tuple and the same values on M, A and L as the second tuple. Finally, the FHD $MD : \{FL\}$ says that the sets of features and languages are determined by the title and director of the movie independently of the actors. It is a natural question to ask whether the following semantically meaningful constraints also need to be enforced explicitly, or whether they are already enforced implicitly: i) the FD $M \to D$ and ii) the FHD $M : \{A\}$?* ∎

Despite the demand to allow the storage of incomplete information, an effective and efficient management of databases requires that certain parts of the information are complete. For example, SQL table definitions permit column headers to be declared `NOT NULL`, i.e., no null value is allowed to occur in such columns [27]. Primary key attributes in SQL table definitions are `NOT NULL` by default [27]. With respect to Zaniolo's "no information" null value, the implication problem of the class of functional dependencies alone has been studied in the presence of a null-free subschema (NFS) [6]. An NFS over a given relation schema is simply the subset of attributes declared `NOT NULL`. The opportunity to specify an arbitrary NFS R_s provides the data administrator with a flexible mechanism to control the degree of certainty in partial relations. However, the following example illustrates that reasoning about FDs and FHDs in the presence of an arbitrary NFS is subtle, and automated tools for such reasoning tasks cannot be taken for granted.

Example 2 *Let $R = MDAFL$, $\Sigma = \{M : \{DF\}, MD : \{FL\}, MA \to D\}$ from Example 1. For the NFS $R_s = DA$ it turns out that Σ implies indeed the FD $M \to D$ and also the FHD $M : \{A\}$ in the presence of R_s. However, if $R_s = MDFL$, then the relation*

Movie	*Director*	*Actor*	*Feature*	*Language*
Psycho	*G. Van Sant*	*ni*	*Subtitle*	*English*
Psycho	*A. Hitchcock*	*ni*	*Deleted Scenes*	*English*

satisfies Σ and R_s, but violates $M \to D$. In particular, the FD $MA \to D$ and the FHD $M : \{A\}$ are both satisfied since the two tuples both have the null value on A. If we choose R_s to be $MAFL$ instead, then the relation

Movie	*Director*	*Actor*	*Feature*	*Language*
The girl with the dragon tattoo	*ni*	*R. Mara*	*Subtitle*	*English*
The girl with the dragon tattoo	*ni*	*N. Rapace*	*Subtitle*	*Swedish*

satisfies Σ and R_s, but violates $M : \{A\}$. In particular, the FHD $M : \{DF\}$ is satisfied as both tuples have the same values on D and F. ∎

Contributions. As the first contribution of this article we will establish an axiomatization $\mathfrak{W}$ for the implication of the combined class of functional and full hierarchical dependencies in the presence of a null-free subschema. The existence of an axiomatization for the implication of data dependencies can form the basis of an enumeration algorithm that lists all logical consequences. In practice, such an enumeration is often desirable to validate the correct specification of explicit knowledge. An axiomatization may also enable one to develop an algorithm which decides the implication of dependencies efficiently. This complements the enumeration algorithm by a further reasoning capability that can make efficient, but only partial decisions whether some dependency is implicitly specified or not. In contrast, the enumeration algorithm lists all of the implicitly specified dependencies. The axiomatization $\mathfrak{W}$ refers to the traditional notion of implication which takes into account the underlying relation schema R. The next example illustrates that there are dependencies that are implied by the given set of dependencies, the NFS and the underlying relation schema, but not implied by the given set of dependencies and the NFS alone. In order to distinguish between these two notions of implication we refer to R-implication in the first case, and to implication in the second case.

Example 3 *Consider again the relation schema $R = MDAFL$ and the set $\Sigma = \{M : \{DF\}, MD : \{FL\}, MA \to D\}$ of FDs and FHDs over R from Example 1. Let $R_s = AD$ denote a null-free subschema over R. Using the inference rules of Table 1 the following inference*

$$\dfrac{\mathcal{T}_H:\ \dfrac{M:\{DF\} \qquad \mathcal{C}_H^R:\ \dfrac{MD:\{FL\}}{MD:\{A\}}}{M:\{A\}} \qquad MA \to D}{\mathcal{T}_{FH}:\quad M \to D}$$

shows that the FHD $M : \{A\}$ is R-implied by Σ in the presence of R_s. Note that $\mathcal{T}_H$ is only applicable since $D \in R_s$ and $\mathcal{T}_{FH}$ is only applicable since $A \in R_s$. However, if we add the attribute W(riter) *to R to obtain the relation schema $R' = MDAFLW$, then the FHD $M : \{A\}$ is not R'-implied by Σ in the presence of R_s, as the relation*

Movie	*Director*	*Actor*	*Feature*	*Language*	*Writer*
Psycho	***ni***	*V. Vaughn*	*Subtitle*	*English*	*R. Bloch*
Psycho	***ni***	*A. Perkins*	*Subtitle*	*English*	*J. Stefano*

shows. ■

Example 3 demonstrates that there is a difference between those dependencies that are R-implied and those that are implied. As the second contribution of this article we show that the addition of further inference rules to $\mathfrak{W}$ results in an axiomatization $\mathfrak{W}_{\mathcal{C}}$ that is *complementary* and *adequate* for the R-implication of FDs and FHDs in the presence of an NFS. The property of complementarity means that for every R-implied FHD there is an inference by $\mathfrak{W}_{\mathcal{C}}$ in which the R-complementation rule $\mathcal{C}_{\mathrm{H}}^R$ is applied at most once, and if it is applied, then only in the last step of the inference. The property of adequacy means that for every R-implied FD there is an inference by $\mathfrak{W}_{\mathcal{C}}$ in which the R-complementation rule $\mathcal{C}_{\mathrm{H}}^R$ is not applied at all. As the third contribution of this article we show that the system $\mathfrak{U}$ resulting from the removal of the R-complementation rule $\mathcal{C}_{\mathrm{H}}^R$ from $\mathfrak{W}_{\mathcal{C}}$ is an axiomatization for the implication of FDs and FHDs in the presence of an NFS. In particular, the FHDs that can be inferred by $\mathfrak{W}$ but not by $\mathfrak{U}$ are exactly those FHDs that are R-implied but not implied by the given set of FDs and FHDs in the presence of an NFS. Finally, we show that our results also apply to Codd's null interpretation "value unknown at present" [24].

Organization. We summarize related work in Section 2. In Section 3 we give preliminary definitions regarding the extension of Codd's relational model of data with Zaniolo's "no information" null value. We also repeat the notion of R-implication, define functional and full hierarchical dependencies and null-free subschemata. In Section 4 we establish an axiomatization $\mathfrak{W}$ for the R-implication of functional and full hierarchical dependencies in the presence of a null-free subschema. In Section 5 we show that $\mathfrak{W}$ is neither complementary nor adequate, but that the addition of further inference rules to $\mathfrak{W}$ results in an axiomatization $\mathfrak{W}_{\mathcal{C}}$ that enjoys both properties. Section 6 discusses the alternative notion of implication where the set of underlying attributes is left undetermined. We establish an axiomatization for this notion of implication for the combined class of FDs and FHDs in the presence of a null-free subschema. In Section 7 we briefly discuss Codd's interpretation "value unknown at present". Finally, we conclude in Section 8 and briefly comment on possible future work.

2 Related Work

Codd's relational model of data [22] has been the context for a large body of research on data dependencies, and overviews include [1, 14, 39, 78, 89]. Traditional areas of applications for data dependencies include normalization [15, 34, 93], requirements engineering and schema validation [50, 76], data mining [77], database security [16], view maintenance [63] and query optimization [33]. New application areas are data cleaning [40], data transformations [28], consistent query answering [21], data exchange [38, 2] and data integration [19]. Data dependencies have received considerable attention in other data models [3, 6, 18, 42, 44, 49, 51, 52, 53, 61, 66, 68, 87, 88, 91, 97, 95]. FDs capture around two-thirds and FHDs around one-quarter of all uni-relational dependencies that arise in applications [32, 99]. Join, equality- and tuple-generating, and embedded dependencies are more expressive, but are beyond our scope here [11, 20, 36, 75]. Note that join dependencies are not Hilbert-style axiomatizable [80], and acyclic join dependencies are captured by sets of FHDs [10]. The use of equality- and tuple-generating dependencies [39] beyond FDs and FHDs have their *major* motivation in data exchange [37].

For total relations, Armstrong [5] established the first axiomatization for the class of FDs. Beeri, Fagin, and Howard extended this axiomatization to the combined class of FDs and multivalued dependencies (MVDs) [9]. Delobel introduced the class of hierarchical decompositions, including full hierarchical dependencies (FHDs) [31]. Biskup [13] studied the difference between the R-implication and implication of MVDs. Specifically, Biskup [13] established the first axiomatization $\mathfrak{S}_0$ for the R-implication of MVDs that is complementary. Biskup further showed that the removal of the R-complementation rule from $\mathfrak{S}_0$ results in an inference system that forms a finite axiomatization for the implication of MVDs. Hence, the R-complementation rule is a mere means of database normalization [13]. Link [73] established algebraic, proof-theoretical and logical characterizations for the implication of MVDs. Köhler, Hartmann and Link [64] extended these findings to the class of FHDs. Moreover, Biskup and Link [17] investigated the combined class of FDs and FHDs. In fact, an axiomatization for R-implication was established that is not only complementary for FHDs but also adequate for FDs.

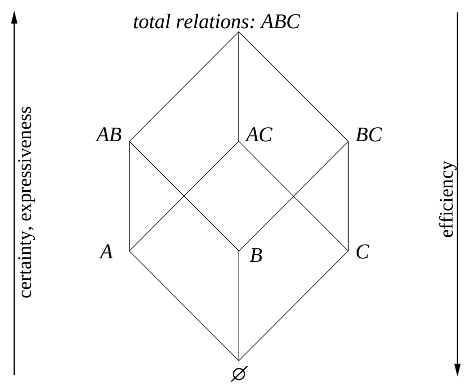

Figure 1: Control over trade-offs by specifying a null-free subschema

One of the most important extensions of Codd's basic relational model [22] is incomplete information. This is mainly due to the high demand for the correct handling of such information in real-world applications. Approaches to deal with incomplete information comprise partial relations [24, 60, 69], or-relations [59, 92], fuzzy relations [85] and rough sets [101]. In the literature many kinds of null values have been proposed; for example, "missing" or "value unknown at present" [47, 48], "non-existence" [74], "inapplicable" [48], "no information" [100] and "open" [45]. Results that are of most relevance to this article are based on Zaniolo's "no information" interpretation.

Lien [70] axiomatized the R-implication of FDs and MVDs over partial relations where the NFS is fixed to the empty attribute set. A complementary axiomatization for the R-implication of MVDs and an axiomatization for the implication of MVDs have been established [72]. Atzeni and Morfuni established axiomatizations for the R-implication of FDs in the presence of existence constraints, e.g., null-free subschemata [6]. Levene and Loizou introduced and axiomatized R-implication for the combined class of weak and strong FDs with respect to a possible world semantics [68]. The Armstrong axioms axiomatize strong FDs, while weak FDs have the same axiomatization as the FDs of Lien, Atzeni and Morfuni [6, 70].

For the "no information" interpretation, Hartmann and Link [54, 55] established recently an axiomatization for the R-implication of the combined class of FDs and MVDs in the presence of a null-free subschema R_s. Noticeably, the R-implication of FDs and MVDs in the presence of an NFS is equivalent to S-3 implication of a propositional fragment in Cadoli and Schaerf's para-consistent family of S-3 logics [54, 55]. Herein, S is the set of propositional variables that correspond to the attributes of the null-free subschema R_s. Essentially, R_s provides database engineers with full control to balance the expressiveness of database constraints with the efficiency to reason about them. This is illustrated in Figure 1 on the example of a relation schema with three attributes. In [41] a complementary and adequate axiomatization of the R-implication of FDs and MVDs in the presence of a null-free subschema was announced, as well as an axiomatization for the implication of this class. In the present article we prove all these results in detail for the more general class of FDs and FHDs in the presence of an NFS. This also complements the finding of [54, 55] where the difference between R-implication and implication was not studied.

3 Dependencies over partial relations

We summarize the basic notions required for our treatment of data dependencies over partial relations in the following sections.

3.1 Partial relations

Let $\mathfrak{A} = \{A_1, A_2, \ldots\}$ be a (countably) infinite set of distinct symbols, called attributes (column names of tables). A *relation schema* is a finite non-empty subset R of $\mathfrak{A}$. Each attribute A of a relation schema R is associated with an infinite domain $dom(A)$ which represents the possible values that can occur in column A. In order to encompass incomplete information every domain contains the same distinguished null value, i.e. $\mathtt{ni} \in dom(A)$ for all A. The intention of $\mathtt{ni}$ is to mean "no information". This is the most primitive interpretation, and it can model non-existing as well as unknown information [6, 100].

For attribute sets X and Y we may write XY for their set union $X \cup Y$. 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 (R-tuple or simply tuple, if R is understood) is a function $t : R \to \cup_{A \in R} dom(A)$ with $t(A) \in dom(A)$ for all $A \in R$. The null value occurrence $t(A) = \mathtt{ni}$ associated with the value $t(A)$ of a tuple t on attribute A means that "no information" is available about the attribute A for the tuple t. For $X \subseteq R$ let $t[X]$ denote the restriction of the tuple t over R to X. A (partial) *relation* r over R is a finite set of tuples over R. Let t_1 and t_2 be two tuples over R. It is said that t_1 *subsumes* t_2 if for every attribute $A \in R$, $t_1(A) = t_2(A)$ or $t_2(A) = \mathtt{ni}$ holds. In consistency with previous work [6, 70, 100], the following restriction will be imposed, unless stated otherwise: No relation shall contain two tuples t_1 and t_2 such that t_1 subsumes t_2. With no null values present this means that no duplicate tuples occur.

For a tuple t over R and a set $X \subseteq R$, t is said to be *X-total*, if for all $A \in X$, $t(A) \neq \mathtt{ni}$. Similar, a relation r over R is said to be *X-total*, if every tuple t of r is X-total. A relation r over R is said to be a *total relation*, if it is R-total.

We recall the definition of projection and join operations on partial relations [6, 70]. Let r be some relation over R. Let X be some subset of R. The *projection* $r[X]$ of r on X is the set of tuples t for which (i) there is some $t_1 \in r$ such that $t = t_1[X]$ and (ii) there is no $t_2 \in r$ such that $t_2[X]$ subsumes t and $t_2[X] \neq t$. For $Y \subseteq X$, the *Y-total projection* $r_Y[X]$ of r on X is $r_Y[X] = \{t \in r[X] \mid t \text{ is } Y\text{-total}\}$. Given an X-total relation r over R and an X-total relation s over S such that $X = R \cap S$ the *natural join* $r \bowtie s$ of r and s is the relation over $R \cup S$ which contains those tuples t such that there are some $t_1 \in r$ and $t_2 \in s$ with $t_1 = t[R]$ and $t_2 = t[S]$ [6, 70].

Example 4 *Let* DVD *denote the relation schema that consists of the attributes* Movie*,* Director*,* Actor*,* Feature *and* Language*. Then*

Movie	*Director*	*Actor*	*Feature*	*Language*
The girl with the dragon tattoo	*D. Fincher*	***ni***	*Commentary*	*English*
The girl with the dragon tattoo	*N.A. Oplev*	***ni***	*Subtitle*	*Swedish*

denotes a relation over DVD *that is* $\{Movie, Director, Feature, Language\}$*-total.* $\blacksquare$

3.2 Constraints over partial relations

We define the classes of constraints that are of interest in this article.

Definition 1 *A* null-free subschema *(NFS)* R_s *over* R *is a subset* $R_s \subseteq R$*. The NFS* R_s *over* R *is satisfied by a relation* r *over* R*, if* r *is* R_s*-total.* $\blacksquare$

Note that an NFS R_s captures SQL's `NOT NULL` constraints: R_s is just the set of attributes declared `NOT NULL`.

Example 5 *The relation of Example 4 violates the NFS* $\mathrm{DVD}_s^1 = \{Actor\}$ *and satisfies the NFS* $\mathrm{DVD}_s^2 = \{Movie, Language\}$*.* $\blacksquare$

Functional dependencies (FDs) between sets of attributes have played a central role in the study of relational databases [5, 8, 12, 15, 14, 22, 23, 65], and seem to be central for the study of database design in other data models as well [4, 42, 49, 66, 68, 56, 88, 91, 96, 97, 95]. The notion of a functional dependency over total relations is well-understood and the semantic interaction between these dependencies has been syntactically captured by Armstrong's well-known axioms [5, 65].

Definition 2 *A* functional dependency *(FD) over a relation schema* R *is an expression* $X \to Y$ *where* $X, Y \subseteq R$*, and* $X \cap Y = \emptyset$*. A relation* r *over* R *satisfies the FD* $X \to Y$*, denoted by* $\models_r X \to Y$*, if and only if for all* $t_1, t_2 \in r$ *the following holds: if* $t_1[X] = t_2[X]$ *and* t_1, t_2 *are* X*-total, then* $t_1[Y] = t_2[Y]$*.* $\blacksquare$

Example 6 *The relation of Example 4 satisfies the FDs Movie → Actor and Actor → Feature but violates the FD Movie → Feature.* $\blacksquare$

FDs are incapable of modeling many important properties that database users have in mind. Full hierarchical dependencies (FHDs), including multivalued dependencies, provide a more general notion and offer a response to the shortcomings of FDs [7, 17, 31, 34, 43, 64, 54, 55, 71, 73, 93]. We will now introduce FHDs into the context of partial relations.

Definition 3 *A* full hierarchical dependency *(FHD) over a relation schema* R *is an expression* $X : S$ *where* $X \subseteq R$ *and* S *is a set of mutually disjoint non-empty subsets of* R *that are also disjoint from* X*, i.e., for all* $Y \in S$ *we have* $\emptyset \neq Y \subseteq R$ *and for all* $Y, Z \in S \cup \{X\}$ *we have* $Y \cap Z = \emptyset$*. A relation* r *over* R *is said to* satisfy *the full hierarchical dependency* $X : \{Y_1, \ldots, Y_k\}$ *over* 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* $1 \leq i, j \leq k+1$ *and* $t_1, \ldots, t_{k+1}$ *are* X*-total, then there is some* $t \in r$ *such that* $t[XY_i] = t_i[XY_i]$ *for all* $i = 1, \ldots, k$ *and* $t[X(R - XY_1 \cdots Y_k)] = t_{k+1}[X(R - XY_1 \cdots Y_k)]$*.* $\blacksquare$

Example 7 *The relation of Example 4 satisfies the FHD* $Movie : \{\{Director, Feature, Language\}\}$ *and violates the FHD* $Movie : \{\{Director, Actor\}, \{Feature\}\}$. ■

Informally, the relation r satisfies the FHD $X : \{Y_1, \ldots, Y_k\}$ if the X-total values determine the sets of values on the Y_i independently from the sets of values on $R - XY_i$. This actually suggests that the relation schema R is overloaded in the sense that it carries $k+1$ independent facts $XY_1, \ldots, XY_k, X(R - XY_1 \cdots Y_k)$. The following result follows from the characterization of satisfaction for multivalued dependencies [70], which are FHDs where $k = 1$. We omit the proof, but we remark that for the case of $k = 0$ we have $r = r[R] = r[X(R - X)]$.

Theorem 1 *Let* $X, Y_1, \ldots, Y_k \subseteq R$ *be mutually disjoint where* $Y_1, \ldots, Y_k$ *are non-empty, and let* k *denote a non-negative integer. A relation* r *over* R *satisfies the FHD* $X : \{Y_1, \ldots, Y_k\}$ *over* R *if and only if* $r_X[r] = r_X[XY_1] \bowtie \cdots \bowtie r_X[XY_k] \bowtie r_X[X(R - XY_1 \cdots Y_k)]$. ■

For total relations the characteristic of FHDs in Theorem 1 is fundamental to the theory of relational database design as it ensures the losslessness of decompositions [34]. For this reason a lot of research has been devoted to studying the behavior of these dependencies over total relations. Over partial relations this research direction has remained rather unexplored. We will study the class of functional and full hierarchical dependencies in the presence of null-free subschemata over partial relations.

Example 8 *The relation* r *of Example 4 has the following projections on {Movie, Director, Actor}, {Movie, Feature} and {Movie, Language}, respectively.*

Movie	*Director*	*Actor*
The girl with the dragon tattoo	*D. Fincher*	$\boldsymbol{ni}$
The girl with the dragon tattoo	*N.A. Oplev*	$\boldsymbol{ni}$

Movie	*Feature*
The girl with the dragon tattoo	*Commentary*
The girl with the dragon tattoo	*Subtitle*

Movie	*Language*
The girl with the dragon tattoo	*English*
The girl with the dragon tattoo	*Swedish*

.

Since the natural join of these two relations is different from (the {Movie}-total projection of) r *itself it follows that* r *does not satisfy* $Movie : \{\{Director, Actor\}, \{Feature\}\}$. ■

For the convenience of presentation we will introduce the following notation.

Definition 4 *For two tuples* t_1, t_2 *over relation schema* R *we define*

$$\begin{aligned} ag^s(t_1,t_2) &= \{A \in R \mid t_1(A) = t_2(A) \text{ and } t_1(A) \neq \boldsymbol{ni} \neq t_2(A)\}, \\ ag^w(t_1,t_2) &= \{A \in R \mid t_1(A) = \boldsymbol{ni} = t_2(A)\}, \\ ag(t_1,t_2) &= ag^s(t_1,t_2) \cup ag^w(t_1,t_2)\ . \end{aligned}$$ ■

Remark 1 *In Section 7, where we discuss Codd's null interpretation "value unknown at present", we will adjust the definition of weak agree sets to match the possible world semantics underlying this interpretation.* ∎

Remark 2 *Suppose we allow the members of the set S in an FHD $X : S$ to be empty. Then for all positive k we have the property that for all relations r the FHD $X : \{\emptyset, Y_2, \ldots, Y_k\}$ is satisfied by r if and only if r satisfies the FHD $X : \{Y_2, \ldots, Y_k\}$. In particular, if $k = 1$, then $X : \{\emptyset\}$ is equivalent to $X : \emptyset$; more specifically, they are satisfied by all relations.*

One may now define an equivalence relation over the set of FHDs defined over some fixed relation schema. Indeed, two such FHDs are equivalent whenever they are satisfied by the same relations over the schema. Strictly speaking, we will apply inference rules to these equivalence classes of FHDs.

For the sake of simplicity, however, we have limited Definition 3 to those FHDs where no empty sets are allowed to occur as elements of a right-hand side. As the property from the beginning of this remark shows, this is not a real limitation but just a suitable choice of a representative from the equivalence classes. ∎

3.3 Implication and inference

For the design of a relational database schema dependencies are normally specified as semantic constraints on the relations which are intended to be instances of the schema. During the design process one usually needs to determine further dependencies which are logically implied by the given ones. In order to emphasize the dependence of implication on the underlying relation schema R we refer to R-implication. Let $lhs(\sigma)$ denote the attribute set on the left-hand side and $rhs(\sigma)$ the set of attributes occurring on the right-hand side 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 $X \to Y$. Let $Attr(\sigma)$ denote the set of attributes affected by σ, i.e., $Attr(\sigma) = lhs(\sigma) \cup rhs(\sigma)$. For a relation r and a set Σ of FDs and FHDs over relation schema R we say that r satisfies Σ if r satisfies every $\sigma \in \Sigma$.

Definition 5 *Let $\Sigma \cup \{\varphi\}$ be a set of FDs and FHDs, and R_s an NFS over the relation schema R, i.e., we have $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \cup R_s \subseteq R$. We say that Σ R-*implies φ in the presence of R_s, *denoted by $\Sigma \models^R_{R_s} \varphi$, if and only if every relation r over R that satisfies Σ and the NFS R_s also satisfies φ.* ∎

Let $\mathcal{C}$ denote a class of data dependencies. The *R-implication problem for $\mathcal{C}$ in the presence of a null-free subschema* is to decide, given any relation schema R, any NFS R_s over R, and any set $\Sigma \cup \{\varphi\}$ of data dependencies in $\mathcal{C}$ over R, whether $\Sigma \models^R_{R_s} \varphi$. For the classes $\mathcal{C}$ of dependencies we consider here, the sets $\Sigma \cup \{\varphi\}$ over a relation schema R are always finite, and it does not matter whether the relations are finite or not. For this reason, we will only speak of the R-implication problem. We will show later that it even suffices to consider two-tuple relations. We say that Σ *R-implies φ in the presence of an NFS in the world of two-tuple relations* R_s, denoted by $\Sigma \models^R_{2-R_s} \varphi$, if every two-tuple relation r over R that satisfies Σ and the NFS R_s also satisfies φ. The *two-tuple*

R-implication problem for $\mathcal{C}$ in the presence of a null-free subschema is to decide, given any relation schema R, any NFS R_s over R and any set $\Sigma \cup \{\varphi\}$ of dependencies in $\mathcal{C}$ over R, whether $\Sigma \models^{R}_{2-R_s} \varphi$ holds.

For a set Σ of data dependencies in $\mathcal{C}$ over a relation schema R and an NFS R_s over R, let $\Sigma^*_{R,R_s} = \{\varphi \in \mathcal{C} \mid \Sigma \models^{R}_{R_s} \varphi\}$ be its *semantic closure*. In order to determine the semantic closure one can utilize a syntactic approach by applying inference rules, e.g. those in Table 1. These inference rules have the form

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

and inference rules without any premise are called axioms. An inference rule is called *R-sound* for the R-implication of dependencies in the presence of an NFS, if whenever the set of dependencies in the premise of the rule and the NFS are satisfied by some relation over R and the dependencies and NFS satisfy the condition of the rule, then the relation also satisfies the dependency in the conclusion of the rule. For a finite set $\Sigma \cup \{\varphi\}$ of dependencies and a set $\mathfrak{R}$ of inference rules let $\Sigma \vdash_{\mathfrak{R}} \varphi$ denote the *inference* of φ from Σ by $\mathfrak{R}$. That is, there is some sequence $\gamma = [\sigma_1, \ldots, \sigma_n]$ of dependencies such that $\sigma_n = \varphi$ and every σ_i is an element of Σ or is the conclusion that results from an application of an inference rule in $\mathfrak{R}$ to some premises in $\{\sigma_1, \ldots, \sigma_{i-1}\}$. For a finite set Σ of dependencies in $\mathcal{C}$, let $\Sigma^+_{\mathfrak{R}} = \{\varphi \mid \Sigma \vdash_{\mathfrak{R}} \varphi\}$ be its *syntactic closure* under inferences by $\mathfrak{R}$. A set $\mathfrak{R}$ of inference rules is said to be *R-sound* (*R-complete*) for the R-implication of dependencies in $\mathcal{C}$ in the presence of an NFS if for every relation schema R, for every NFS R_s over R and for every set Σ of dependencies in $\mathcal{C}$ over R we have $\Sigma^+_{\mathfrak{R}} \subseteq \Sigma^*_{R,R_s}$ ($\Sigma^*_{R,R_s} \subseteq \Sigma^+_{\mathfrak{R}}$). The (finite) set $\mathfrak{R}$ is said to be a (finite) *axiomatization* for the R-implication of dependencies in $\mathcal{C}$ in the presence of an NFS if $\mathfrak{R}$ is both R-sound and R-complete for the R-implication of dependencies in $\mathcal{C}$ in the presence of an NFS.

Remark 3 *Note the following two global conditions that we enforce on all applications of inference rules that infer full hierarchical dependencies. Whenever we apply such an inference rule, we remove all empty sets that occur as elements of the right-hand side in the conclusion. Moreover, by applying an inference rule to $X : \emptyset$ we mean an application of the inference rule to $X : \{\emptyset\}$. These two conditions are justified due to Remark 2.* ∎

Example 9 *The* empty-set-axiom $\mathcal{R}_{\mathrm{H}}$ *is derivable from $\{\mathcal{R}_{\mathrm{F}}, \mathcal{I}_{\mathrm{FH}}\}$: we infer $\emptyset \to \emptyset$ by an application of the* empty-set-axiom $\mathcal{R}_{\mathrm{F}}$*, and $\emptyset : \emptyset$ by an application of the* implication rule $\mathcal{I}_{\mathrm{FH}}$ *to $\emptyset \to \emptyset$.*

The trivial FHDs $X : \{R - X\}$ are derivable from $\{\mathcal{R}_{\mathrm{F}}, \mathcal{I}_{\mathrm{FH}}, \mathcal{A}_{\mathrm{H}}, \mathcal{C}^R_{\mathrm{H}}\}$: first we infer $\emptyset : \emptyset$ as before, then we infer $X : \emptyset$ by an application of the FH augmentation rule $\mathcal{A}_{\mathrm{H}}$ *to $\emptyset : \emptyset$, and finally we infer $X : \{R - X\}$ by an application of the* R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ *to $X : \emptyset$.* ∎

$$\frac{}{\emptyset \to \emptyset}$$
(empty-set-axiom, $\mathcal{R}_{\mathrm{F}}$)

$$\frac{X \to YZ}{X \to Y}$$
(decomposition, $\mathcal{D}_{\mathrm{F}}$)

$$\frac{X \to Y \quad X \to Z}{X \to YZ}$$
(FD union, $\mathcal{U}_{\mathrm{F}}$)

$$\frac{X \to Y}{XZ \to Y - Z}$$
(FD augmentation, $\mathcal{A}_{\mathrm{F}}$)

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

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

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

$$\frac{X : \{W\} \quad XY : \{Y_1, \ldots, Y_k\}}{X : \{Y_1 - W, \ldots, Y_k - W, W\}} \, {\scriptstyle Y \subseteq W \cap R_s}$$
(transitivity, $\mathcal{T}_{\mathrm{H}}$)

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

$$\frac{X : \{W\} \quad XY \to Z}{X \to Z - W} \, {\scriptstyle Y \subseteq W \cap R_s}$$
(mixed transitivity, $\mathcal{T}_{\mathrm{FH}}$)

Table 1: Inference Rules for Functional and Full Hierarchical Dependencies in the presence of an NFS R_s

4 An Axiomatization of FDs and FHDs in the Presence of a Null-free Subschema

We will show now that the inference system

$$\mathfrak{W} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{A}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{A}_{\mathrm{H}}, \mathcal{O}_{\mathrm{H}}, \mathcal{T}_{\mathrm{H}}, \mathcal{C}_{\mathrm{H}}^{R}, \mathcal{I}_{\mathrm{FH}}, \mathcal{T}_{\mathrm{FH}}\},$$

as shown in Table 1, forms a finite axiomatization for the R-implication of FDs and FHDs in the presence of an NFS.

4.1 Sound Inference Rules

Lemma 1 *For all relation schemata R and all NFSs R_s over R, every inference rule of $\mathfrak{W}$ is R-sound.*

Proof We apply Definition 4 to show the R-soundness of each rule in $\mathfrak{W}$. Let R denote an arbitrary relation schema, and R_s an arbitrary NFS over R. Let r denote an arbitrary relation over R.

For the R-soundness of the *empty-set-axiom* $\mathcal{R}_{\mathrm{F}}$ note that for any two tuples $t_1, t_2 \in r$ we have $t_1[\emptyset] = t_2[\emptyset]$.

For the R-soundness of the *decomposition rule* $\mathcal{D}_{\mathrm{F}}$ assume that r violates the FD $X \to Y$. Then there are some $t_1, t_2 \in r$ such that $X \subseteq ag^s(t_1, t_2)$ and $Y \not\subseteq ag(t_1, t_2)$. We conclude that $YZ \not\subseteq ag(t_1, t_2)$. Consequently, r violates the FD $X \to YZ$.

For the R-soundness of the *FD union rule* $\mathcal{U}_{\mathrm{F}}$ assume that r violates the FD $X \to YZ$. Then there are some $t_1, t_2 \in r$ such that $X \subseteq ag^s(t_1, t_2)$ and $YZ \not\subseteq ag(t_1, t_2)$. We conclude that $Y \not\subseteq ag(t_1, t_2)$ or $Z \not\subseteq ag(t_1, t_2)$. Consequently, r violates the FD $X \to Y$ or the FD $X \to Z$.

For the R-soundness of the *FD augmentation rule* $\mathcal{A}_{\mathrm{F}}$ assume that r violates the FD $XZ \to Y - Z$. Then there are some $t_1, t_2 \in r$ such that $XZ \subseteq ag^s(t_1, t_2)$ and $Y - Z \not\subseteq ag(t_1, t_2)$. We conclude that $Y \not\subseteq ag(t_1, t_2)$. Consequently, r violates the FD $X \to Y$.

For the R-soundness of the *omission rule* $\mathcal{O}_{\mathrm{H}}$ assume that r violates the FHD $X : \{Y_1, \ldots, Y_k\}$. Then there are some $t_1, \ldots, t_{k+1} \in r$ such that for all $1 \leq i < j \leq k+1$, $X \subseteq ag^s(t_i, t_j)$ and for all $t \in r$ there is some $i \in \{1, \ldots, k\}$ such that $Y_i \not\subseteq ag(t, t_i)$ or $R - XY_1 \cdots Y_k \not\subseteq ag(t, t_{k+1})$. Since $Y \subseteq R - XY_1 \cdots Y_k$ it follows that $Y \not\subseteq ag(t, t_{k+1})$ or $R - XY_1 \cdots Y_k Y \not\subseteq ag(t, t_{k+1})$ holds. Consequently, the $k+2$ tuples $t_1, \ldots, t_{k+1}, t_{k+1}$ show that r violates the FHD $X : \{Y_1, \ldots, Y_k, Y\}$.

The R-soundness of the *R-complementation rule* $\mathcal{C}_{\mathrm{H}}^R$ follows immediately from Theorem 1.

For the R-soundness of the *FH augmentation rule* $\mathcal{A}_{\mathrm{H}}$ assume that r violates the FHD $XZ : \{Y_1 - Z, \ldots, Y_k - Z\}$. Then there are some $t_1, \ldots, t_{k+1} \in r$ such that for all $1 \leq i < j \leq k+1$, $XZ \subseteq ag^s(t_i, t_j)$ and for all $t \in r$ there is some $i \in \{1, \ldots, k\}$ such that $Y_i - Z \not\subseteq ag(t, t_i)$ or $R - XZY_1 \cdots Y_k \not\subseteq ag(t, t_{k+1})$. Consequently, the tuples $t_1, \ldots, t_{k+1}$ show that r violates the FHD $X : \{Y_1, \ldots, Y_k\}$.

For the R-soundness of the *transitivity rule* $\mathcal{T}_{\mathrm{H}}$ assume that r satisfies the FHDs $X : \{W\}$ and $XY : \{Y_1, \ldots, Y_k\}$, and the NFS R_s. Furthermore, let $Y \subseteq W \cap R_s$. Let $t_1, \ldots, t_{k+2} \in r$ be such that for all $1 \leq i < j \leq k+2$, $X \subseteq ag^s(t_i, t_j)$. Since r satisfies $X : \{W\}$ we know that for all $i = 1, \ldots, k+1$ there is some $t'_i \in r$ such that $XW \subseteq ag(t'_i, t_{k+2})$ and $X(R - XW) \subseteq ag(t'_i, t_i)$. As $Y \subseteq W \cap R_s$ holds there are $t'_1, \ldots, t'_{k+1}$ and $t'_{k+2} = t_{k+2}$ such that $XY \subseteq ag^s(t'_i, t'_j)$ holds for all $1 \leq i < j \leq k+2$. From this and the fact that r satisfies $XY : \{Y_1, \ldots, Y_k\}$ we conclude that there is some $t \in r$ such that for all $i = 1, \ldots, k$, $XYY_i \subseteq ag(t, t'_i)$, $XW \subseteq ag(t, t_{k+1})$ and $R - XWY_1 \cdots Y_k \subseteq ag(t, t_{k+2})$. It follows that for all $i = 1, \ldots, k$, $X(Y_i - W) \subseteq ag(t, t_i)$, $XW \subseteq ag(t, t_{k+1})$ and $R - XWY_1 \cdots Y_k \subseteq ag(t, t_{k+2})$ hold. That is, r satisfies the FHD $X : \{Y_1 - W, \ldots, Y_k - W, W\}$, too.

For the R-soundness of the *implication rule* $\mathcal{I}_{\mathrm{FH}}$ assume that r satisfies the FD $X \to Y$. Let $t_1, t_2 \in r$ be such that $X \subseteq ag^s(t_1, t_2)$. Since r satisfies $X \to Y$ it follows that $Y \subseteq ag(t_1, t_2)$. Consequently, $t_1 \in r$ satisfies $XY \subseteq ag(t_1, t_2)$ and $X(R - XY) \subseteq ag(t_1, t_1)$. Hence, r satisfies the FHD $X : \{Y\}$.

For the R-soundness of the *mixed transitivity rule* $\mathcal{T}_{\mathrm{FH}}$ assume that r satisfies the FHD $X : \{W\}$ and the FD $XY \to Z$, and the NFS R_s. Furthermore, let $Y \subseteq W \cap R_s$. Let $t_1, t_2 \in r$ be such that $X \subseteq ag^s(t_1, t_2)$. Since r satisfies $X : \{W\}$ there is some $t \in r$ such that $XW \subseteq ag(t, t_1)$ and $X(R - XW) \subseteq ag(t, t_2)$. Since $Y \subseteq W \cap R_s$ it follows that $XY \subseteq ag^s(t, t_1)$. Since r satisfies $XY \to Z$ we conclude that $Z \subseteq ag(t, t_1)$. Let $A \in X(Z - W)$. Then $t_1(A) = t(A) = t_2(A)$. In particular, $Z - W \subseteq ag(t_1, t_2)$. That is, r satisfies the FD $X \to Z - W$.

In the next lemma we establish the R-soundness of further inference rules for the R-implication of FDs and FHDs in the presence of an NFS. These rules are important to settle our completeness argument.

Lemma 2 *The following inference rules are derivable from $\mathfrak{W}$:*

$$\frac{X \to Y \quad XY \to Z}{X \to Z}\,{\scriptstyle Y\subseteq R_s} \qquad \frac{X : \{Y_1,\dots,Y_k,Y_{k+1}\}}{X : \{Y_1,\dots,Y_kY_{k+1}\}} \qquad \frac{X : \{Y_1,\dots,Y_k\}}{X : \{Y_1,\dots,Y_k,\emptyset\}}$$

(FD transitivity, $\mathcal{T}_{\mathrm{F}}$) (merging, $\mathcal{M}_{\mathrm{H}}$) (empty-set-introduction, $\mathcal{I}_{\emptyset}$)

$$\frac{X : \{Y_1,\dots,Y_k\} \quad X : \{Z\}}{X : \{Y_1 - Z,\dots,Y_{k-1} - Z, Y_kZ\}} \qquad \frac{X : \{Y_1,\dots,Y_k\} \quad X : \{Z\}}{X : \{Y_1,\dots,Y_{k-1}, Y_k - Z\}} \qquad \frac{X : \{Y_1,\dots,Y_k\} \quad X : \{Z\}}{X : \{Y_1,\dots,Y_{k-1}, Y_k \cap Z\}}$$

(union, $\mathcal{U}_{\mathrm{H}}$) (difference, $\mathcal{D}_{\mathrm{H}}$) (intersection, $\mathcal{I}_{\mathrm{H}}$)

Proof We start with an inference of the *FD transitivity rule* $\mathcal{T}_{\mathrm{F}}$:

$$\dfrac{\dfrac{X \to Y}{\mathcal{D}_{\mathrm{F}}:\ X \to Y \cap Z} \quad \dfrac{\dfrac{X \to Y}{\mathcal{I}_{\mathrm{FH}}:\ X : \{Y\}} \quad XY \to Z}{\mathcal{T}_{\mathrm{FH}}:\ X \to Z - Y}\,{\scriptstyle Y\subseteq R_s}}{\mathcal{U}_{\mathrm{F}}:\ X \to Z} .$$

Next we present an inference of the *empty-set-introduction rule* $\mathcal{I}_{\emptyset}$:

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\mathcal{R}_{\mathrm{F}}:\ \emptyset \to \emptyset}}{\mathcal{I}_{\mathrm{FH}}:\ \emptyset : \{\emptyset\}}}{\mathcal{A}_{\mathrm{H}}:\ X : \{\emptyset\}} \quad \dfrac{X : \{Y_1,\dots,Y_k\}}{\mathcal{A}_{\mathrm{H}}:\ X \cup \emptyset : \{Y_1,\dots,Y_k\}}}{\mathcal{T}_{\mathrm{H}}:\ X : \{Y_1,\dots,Y_k,\emptyset\}} .$$

Next we present an inference of the *FH union rule* $\mathcal{U}_{\mathrm{H}}$:

$$\dfrac{\dfrac{\dfrac{\dfrac{X : \{Z\} \qquad X : \{Y_1,\dots,Y_k\}}{\mathcal{T}_{\mathrm{H}}:\ X : \{Y_1 - Z,\dots,Y_k - Z, Z\}}}{\mathcal{C}^R_{\mathrm{H}}:\ X : \{Y_1 - Z,\dots,Y_{k-1} - Z, R - XZY_1\cdots Y_k, Z\}}}{\mathcal{O}_{\mathrm{H}}:\ X : \{Y_1 - Z,\dots,Y_{k-1} - Z, R - XZY_1\cdots Y_k\}}}{\mathcal{C}^R_{\mathrm{H}}:\ X : \{Y_1 - Z,\dots,Y_{k-1} - Z, \underbrace{R - (X(Y_1 - Z)\cdots(Y_{k-1} - Z)(R - XZY_1\cdots Y_k))}_{=Y_kZ}\}} .$$

Note that $X \cap Y_k = \emptyset$, $X \cap Z = \emptyset$ and $Y_i \cap Y_k = \emptyset$ for $i = 1,\dots,k-1$. Next we present an inference of the *merging rule* $\mathcal{M}_{\mathrm{H}}$:

$$\dfrac{\dfrac{X : \{Y_1,\dots,Y_k,Y_{k+1}\}}{\mathcal{O}_{\mathrm{H}}:\ X : \{Y_1,\dots,Y_k\}} \quad \dfrac{X : \{Y_1,\dots,Y_k,Y_{k+1}\}}{\mathcal{O}_{\mathrm{H}}:\ X : \{Y_{k+1}\}}}{\mathcal{U}_{\mathrm{H}}:\ X : \{Y_1,\dots,Y_kY_{k+1}\}} .$$

Next we present an inference of the *difference rule* $\mathcal{D}_{\mathrm{H}}$:

$$\dfrac{\dfrac{\dfrac{X : \{Y_1,\dots,Y_k\}}{\mathcal{O}_{\mathrm{H}}:\ X : \{Y_1,\dots,Y_{k-1}\}}}{\mathcal{I}_{\emptyset}:\ X : \{Y_1,\dots,Y_{k-1},\emptyset\}} \quad \dfrac{\dfrac{X : \{Z\} \quad \dfrac{X : \{Y_1,\dots,Y_k\}}{\mathcal{O}_{\mathrm{H}}:\ X : \{Y_k\}}}{\mathcal{T}_{\mathrm{H}}:\ X : \{Y_k - Z, \emptyset\}}}{\mathcal{O}_{\mathrm{H}}:\ X : \{Y_k - Z\}}}{\mathcal{U}_{\mathrm{H}}:\ X : \{Y_1,\dots,Y_{k-1}, Y_k - Z\}} .$$

Finally, we present an inference of the *intersection rule* $\mathcal{I}_{\mathrm{H}}$:

$$\dfrac{X:\{Y_1,\ldots,Y_k\} \quad \dfrac{\dfrac{X:\{Y_1,\ldots,Y_k\} \qquad X:\{Z\}}{\mathcal{D}_{\mathrm{H}}:\ X:\{Y_1,\ldots,Y_{k-1},Y_k-Z\}}}{\mathcal{O}_{\mathrm{H}}:\qquad X:\{Y_k-Z\}}}{\mathcal{D}_{\mathrm{H}}:\qquad X:\{Y_1,\ldots,Y_{k-1},Y_k\cap Z\}}\ .$$

Note that $Y_k \cap Z = Y_k - (Y_k - Z)$.

4.2 Completeness

Let R be some arbitrary relation schema, let Σ be a set of FDs and FHDs, and R_s an NFS over R. Let $Dep_{\Sigma,R_s}(X)$ be the set of all $W \subseteq R - X$ for which some FHD $X : S$ with $W \in S$ can be inferred from Σ and R_s by $\mathfrak{W}$, i.e., $Dep_{\Sigma,R_s}(X) = \{W \subseteq R - X \mid \text{there is some } X : S \in \Sigma^+_{\mathfrak{W}} \text{ such that } W \in S\} \cup \{\emptyset\}$. Note that $Dep_{\Sigma,R_s}(X)$ is finite, and $(Dep_{\Sigma,R_s}(X), \subseteq, \cup, \cap, (\cdot)^{\mathcal{C}}, \emptyset, R - X)$ constitutes a Boolean algebra due to the soundness of FH union rule $\mathcal{U}_{\mathrm{H}}$, difference rule $\mathcal{D}_{\mathrm{H}}$ and intersection rule $\mathcal{I}_{\mathrm{H}}$. Recall that an element $a \in P$ of a poset $(P, \sqsubseteq, 0)$ with least element 0 is called an *atom* of $(P, \sqsubseteq, 0)$ [46] if and only if $a \neq 0$ and every element $b \in P$ with $b \sqsubseteq a$ satisfies $b = 0$ or $b = a$. $(P, \sqsubseteq, 0)$ is called *atomic* if and only if for every element $b \in P$ with $b \neq 0$ there is an atom $a \in P$ with $a \sqsubseteq b$. In particular, every finite Boolean algebra is atomic. The set $DepB_{\Sigma,R_s}(X)$ of all atoms of $(Dep_{\Sigma,R_s}(X), \subseteq, \emptyset)$ is called the *dependency basis* [7] of X with respect to Σ and R_s. Moreover, let $X^+_{\Sigma,R_s} = \{A \in R - X \mid X \to A \in \Sigma^+_{\mathfrak{W}}\}$ be the attribute closure of X with respect to Σ and R_s. Furthermore, we define $\overline{X}_{\Sigma,R_s} = XX^+_{\Sigma,R_s}$, i.e., $\overline{X}_{\Sigma,R_s}$ is the disjoint union of X and X^+_{Σ,R_s}.

Theorem 2 *Let $\Sigma \cup \{X : S\}$ be a set of FDs and FHDs, and R_s an NFS over the relation schema R. Then the following hold:*

1. *$X : S \in \Sigma^+_{\mathfrak{W}}$ if and only if for every $Y \in S$ there is some $\mathcal{Y} \subseteq DepB_{\Sigma,R_s}(X)$ such that $Y = \bigcup \mathcal{Y}$;*
2. *$X \to Y \in \Sigma^+_{\mathfrak{W}}$ if and only if $Y \subseteq X^+_{\Sigma,R_s}$;*
3. *if $X \to A \in \Sigma^+_{\mathfrak{W}}$, then $\{A\} \in DepB_{\Sigma,R_s}(X)$.*

Proof 1. If $S = \emptyset$, then $X : \emptyset \in \Sigma^+_{\mathfrak{W}}$ by applications of the empty-set-axiom $\mathcal{R}_{\mathrm{F}}$, the implication rule $\mathcal{I}_{\mathrm{FH}}$ and the augmentation rule $\mathcal{A}_{\mathrm{H}}$. It remains to consider the case where $S \neq \emptyset$.

Let $Y \in S$ for $X : S \in \Sigma^+_{\mathfrak{W}}$. That is, $Y \in Dep_{\Sigma,R_s}(X)$, and since every element b of a Boolean algebra is the union over those atoms a with $a \sqsubseteq b$ we know that $Y = \bigcup \mathcal{Y}$ for $\mathcal{Y} = \{W \in DepB_{\Sigma,R_s}(X) \mid W \subseteq Y\}$.

Vice versa, let $Y \in S$ be arbitrary and suppose that $Y = \bigcup \mathcal{Y}$ for some $\mathcal{Y} \subseteq DepB_{\Sigma,R_s}(X)$. Since $DepB_{\Sigma,R_s}(X) \subseteq Dep_{\Sigma,R_s}(X)$ and $Dep_{\Sigma,R_s}(X)$ is closed under unions it follows that $Y \in Dep_{\Sigma,R_s}(X)$. Let $S = \{Y_1, \ldots, Y_k\}$. Then we know by

	$X(X^+_{\Sigma,R_s} \cap R_s)$	$X^+_{\Sigma,R_s} - R_s$	$W_1 \cap R_s$	$W_1 - R_s$	.	W_i	.	$W_k \cap R_s$	$W_k - R_s$
t_1	$0 \cdots 0$	$\mathtt{ni} \cdots \mathtt{ni}$	$0 \cdots 0$	$\mathtt{ni} \cdots \mathtt{ni}$		$0 \cdots 0$		$0 \cdots 0$	$\mathtt{ni} \cdots \mathtt{ni}$
t_2	$0 \cdots 0$	$\mathtt{ni} \cdots \mathtt{ni}$	$0 \cdots 0$	$\mathtt{ni} \cdots \mathtt{ni}$		$1 \cdots 1$		$0 \cdots 0$	$\mathtt{ni} \cdots \mathtt{ni}$

Table 2: The relation r_φ in the completeness proof

definition of $Dep_{\Sigma,R_s}(X)$ that $X : \{Y_i\} \in \Sigma^+_{\mathfrak{W}}$ holds for all $i = 1, \ldots, k$. Consecutive applications of the empty-set-introduction rule $\mathcal{I}_\emptyset$ and the union rule $\mathcal{U}_{\mathrm{H}}$ lead to $X : S \in \Sigma^+_{\mathfrak{W}}$.

2. If $X \to Y \in \Sigma^+_{\mathfrak{W}}$, then it follows that for all $A \in Y$ we have $X \to A \in \Sigma^+_{\mathfrak{W}}$ by the soundness of the decomposition rule $\mathcal{D}_F$. That is, $A \in X^+_{\Sigma,R_s}$ for all $A \in Y$. Vice versa, if $Y \subseteq X^+_{\Sigma,R_s}$, then $X \to A \in \Sigma^+_{\mathfrak{W}}$ for all $A \in Y$ due to the definition of X^+_{Σ,R_s}. Consequently, if Y is non-empty, then $X \to Y \in \Sigma^+_{\mathfrak{W}}$ by applications of the union rule $\mathcal{U}_{\mathrm{F}}$. If Y is empty, then $X \to \emptyset \in \Sigma^+_{\mathfrak{W}}$ by an application of the empty-set-axiom $\mathcal{R}_{\mathrm{F}}$ and an application of the augmentation rule $\mathcal{A}_{\mathrm{F}}$.

3. If $X \to A \in \Sigma^+_{\mathfrak{W}}$, then $X : \{A\} \in \Sigma^+_{\mathfrak{W}}$ by an application of the implication rule $\mathcal{I}_{\mathrm{FH}}$. Since $\{A\}$ is an atom the definition of $Dep_{\Sigma,R_s}(X)$ implies that $\{A\} \in DepB_{\Sigma,R_s}(X)$.

Theorem 3 *The set $\mathfrak{W}$ of inference rules forms a finite axiomatization for the R-implication of FDs and FHDs in the presence of an NFS.*

Proof The R-soundness of $\mathfrak{W}$ follows by a simple induction on the length of an inference and the R-soundness of the individual rules proven in Lemma 1. It remains to show the R-completeness of $\mathfrak{W}$. Let R be an arbitrary relation schema, let Σ be an arbitrary set of FDs and FHDs, and let R_s be an arbitrary NFS over R.

Suppose first there is some FHD φ, say $X : S$, such that $\varphi \notin \Sigma^+_{\mathfrak{W}}$. We will now construct a two-tuple relation r_φ that violates $X : S$ but satisfies Σ and the NFS R_s.

Let $DepB_{\Sigma,R_s}(X)$ be the disjoint union of $\{\{A\} \mid A \in X^+_{\Sigma,R_s}\}$ and $\{W_1, \ldots, W_k\}$. In particular, it follows that $\{X, X^+_{\Sigma,R_s}, W_1, \ldots, W_k\}$ forms a partition of R. Since $\varphi \notin \Sigma^+_{\mathfrak{W}}$ we conclude by Theorem 2 that there is some attribute set $Y \in S$ such that Y is not the union of some elements of $DepB_{\Sigma,R_s}(X)$. Consequently, there is some $i \in \{1, \ldots, k\}$ such that $Y \cap W_i \neq \emptyset$ and $Y - W_i \neq \emptyset$ hold. Let $r_\varphi := \{t_1, t_2\}$ be the relation in Table 2. That is, for all $A \in R$, i) $t_1(A) = t_2(A)$ if and only if $A \notin W_i$, and ii) $t_1(A)$ and $t_2(A)$ are A-total if and only if $A \in XR_sW_i$. Note that r_φ satisfies the following property: if $Z = \bigcup_{B \in \mathcal{B}} B$ for some $\mathcal{B} \subseteq DepB_{\Sigma,R_s}(X)$, then $t_1[Z] = t_2[Z]$, if $W_i \notin \mathcal{B}$, or $t_1[R - Z] = t_2[R - Z]$, if $W_i \in \mathcal{B}$. Also note that $t_1[\overline{X}_{\Sigma,R_s}] = t_2[\overline{X}_{\Sigma,R_s}]$.

It follows from the construction that r_φ violates φ and r_φ satisfies the NFS R_s. In order to show that $\varphi \notin \Sigma^*_{R,R_s}$ it remains to prove that r_φ satisfies Σ.

Let $U : \{V_1, \ldots, V_l\} \in \Sigma$. Suppose that $U \subseteq ag^s(t_1, t_2)$. Let

$$W := \bigcup \{W_j \in DepB_{\Sigma,R_s}(X) \mid W_j \cap U \neq \emptyset\}.$$

From $U \subseteq ag(t_1, t_2)$ and the construction of r_φ we conclude that $W \subseteq ag(t_1, t_2)$. Since W is the union of elements from $DepB_{\Sigma,R_s}(X)$ we conclude by Theorem 2 that $X : \{W\} \in \Sigma^+_{\mathfrak{W}}$. Note that X^+_{Σ,R_s} is also the union of elements from $DepB_{\Sigma,R_s}(X)$, i.e., $X : \{X^+_{\Sigma,R_s}\} \in \Sigma^+_{\mathfrak{W}}$, and by an application of the *FHD union rule* $\mathcal{U}_{\mathrm{H}}$, $X : \{X^+_{\Sigma,R_s}W\} \in \Sigma^+_{\mathfrak{W}}$, too. An application of the *FHD augmentation rule* $\mathcal{A}_{\mathrm{H}}$ to $U : \{V_1, \ldots, V_l\} \in \Sigma$ results in $UX : \{V_1 - X, \ldots, V_l - X\} \in \Sigma^+_{\mathfrak{W}}$.

Since $U \subseteq ag^s(t_1, t_2)$, the construction of r_φ implies that

$$U \subseteq X((X^+_\Sigma W) \cap R_s).$$

We now apply the *transitivity rule* $\mathcal{T}_{\mathrm{H}}$ to $X : \{X^+_{\Sigma,R_s}W\} \in \Sigma^+_{\mathfrak{W}}$, $XU : \{V_1 - X, \ldots, V_l - X\} \in \Sigma^+_{\mathfrak{W}}$ and $U - X \subseteq (X^+_{\Sigma,R_s}W) \cap R_s$ to infer $X : \{V_1 - XX^+_{\Sigma,R_s}W, \ldots, V_l - XX^+_{\Sigma,R_s}W, X^+_{\Sigma,R_s}W\} \in \Sigma^+_{\mathfrak{W}}$. Consequently, for all $j = 1, \ldots, l$, $X : \{V_j - XX^+_{\Sigma,R_s}W\} \in \Sigma^+_{\mathfrak{W}}$ by means of the *omission rule* $\mathcal{O}_{\mathrm{H}}$. From the definition of X^+_{Σ,R_s} it follows that $X \to X^+_{\Sigma,R_s} \in \Sigma^+_{\mathfrak{W}}$ by applications of the *FD union rule* $\mathcal{U}_{\mathrm{F}}$. From $X \to X^+_{\Sigma,R_s} \in \Sigma^+_{\mathfrak{W}}$ we conclude $X \to (V_j - W) \cap X^+_{\Sigma,R_s} \in \Sigma^+_{\mathfrak{W}}$ by means of the *decomposition rule* $\mathcal{D}_{\mathrm{F}}$, and $X : \{(V_j - W) \cap X^+_{\Sigma,R_s}\} \in \Sigma^+_{\mathfrak{W}}$ by an application of the *implication rule* $\mathcal{I}_{\mathrm{FM}}$, for all $j = 1, \ldots, l$. Moreover, an application of the *FHD union rule* $\mathcal{U}_{\mathrm{H}}$ to $X \twoheadrightarrow V_j - XX^+_{\Sigma,R_s}W \in \Sigma^+_{\mathfrak{W}}$ and $X : \{((V_j - W) \cap X^+_{\Sigma,R_s})\} \in \Sigma^+_{\mathfrak{W}}$ results in $X : \{V_j - XW\} \in \Sigma^+_{\mathfrak{W}}$ for all $j = 1, \ldots, l$. Therefore, $V_j - XW$ is the union of elements from $DepB_{\Sigma,R_s}(X)$ for all $j = 1, \ldots, l$. Consequently, $V_j - XW \subseteq ag(t_1, t_2)$ or $XW(R - V_j) \subseteq ag(t_1, t_2)$ for all $j = 1, \ldots, l$.

In summary, we have $XX^+_{\Sigma,R_s}WV_j \subseteq ag(t_1, t_2)$ or $XX^+_{\Sigma,R_s}W(R - V_j) \subseteq ag(t_1, t_2)$ for all $j = 1, \ldots, l$. The first case implies $UV_j \subseteq ag(t_1, t_2)$ and the second case implies $U(R - V_j) \subseteq ag(t_1, t_2)$ for every $j \in \{1, \ldots, l\}$. This shows that for all $j = 1, \ldots, l$, r_φ satisfies $U : \{V_j\}$. Due to the soundness of the *empty-set-introduction rule* $\mathcal{I}_\emptyset$ and the *FHD union rule* $\mathcal{U}_{\mathrm{H}}$ we conclude that r_φ satisfies $U : \{V_1, \ldots, V_l\}$.

Let $U \to V \in \Sigma$. Suppose that $U \subseteq ag^s(t_1, t_2)$. As before let

$$W := \bigcup\{W_j \in DepB_{\Sigma,R_s}(X) \mid W_j \cap U \neq \emptyset\}.$$

From $U \subseteq ag(t_1, t_2)$ and the construction of r_φ we conclude that $W \subseteq ag(t_1, t_2)$. An application of the *FD augmentation rule* $\mathcal{A}_{\mathrm{F}}$ to $U \to V \in \Sigma$ results in $XU \to V - X \in \Sigma^+_{\mathfrak{W}}$. As before we conclude that $X : \{X^+_{\Sigma,R_s}W\} \in \Sigma^+_{\mathfrak{W}}$ and that it follows from the construction of r_φ that

$$U \subseteq X((X^+_{\Sigma,R_s}W) \cap R_s).$$

We now apply the *mixed transitivity rule* $\mathcal{T}_{\mathrm{FH}}$ to $X : \{X^+_{\Sigma,R_s}W\} \in \Sigma^+_{\mathfrak{W}}$, $XU \to V - X \in \Sigma^+_{\mathfrak{W}}$ and $U - X \subseteq (X^+_{\Sigma,R_s}W) \cap R_s$ to infer $X \to V - XX^+_{\Sigma,R_s}W \in \Sigma^+_{\mathfrak{W}}$. As before, we conclude that $X \to (V - W) \cap X^+_{\Sigma,R_s} \in \Sigma^+_{\mathfrak{W}}$, and therefore $X \to V - XW \in \Sigma^+_{\mathfrak{W}}$ by means of the *FD union rule* $\mathcal{U}_{\mathrm{F}}$. Therefore, $V - XW \subseteq X^+_{\Sigma,R_s}$. Consequently, $V - XW \subseteq ag(t_1, t_2)$ and since $XW \subseteq ag(t_1, t_2)$ holds as well, we conclude $V \subseteq ag(t_1, t_2)$. Therefore, r_φ satisfies $U \to V$.

Finally, suppose there is some FD φ, say $X \to Y$, such that $\varphi \notin \Sigma^+_{\mathfrak{W}}$. Due to the *FD union rule* $\mathcal{U}_{\mathrm{F}}$ there is some $A \in Y$ such that $X \to A \notin \Sigma^+_{\mathfrak{W}}$. It follows that $A \notin X^+_{\Sigma,R_s}$. Without loss of generality let $A \in W_i$. Let r_φ be the two-tuple relation from before. It follows that r_φ violates $X \to Y$ since $X \subseteq ag^s(t_1, t_2)$, and $A \notin ag(t_1, t_2)$. We know that r_φ satisfies Σ and the NFS R_s. Consequently, $\varphi \notin \Sigma^*_{R,R_s}$.

We have shown the completeness of $\mathfrak{W}$ for the implication of FDs and FHDs in the presence of an NFS.

The two-tuple counterexample relation that we utilize in the proof of Theorem 3 allows us to derive the following corollary.

Corollary 1 *Let $\Sigma \cup \{\varphi\}$ denote a set of FDs and FHDs, and let R_s denote an NFS over the relation schema R. Then Σ R-implies φ in the presence of R_s if and only if Σ R-implies φ in the presence of R_s in the world of all two-tuple relations.*

Proof If $\Sigma \models^R_{2-R_s} \varphi$ does not hold, then $\Sigma \models^R_{R_s} \varphi$ does not hold. If $\Sigma \models^R_{R_s} \varphi$ does not hold, then $\Sigma \vdash_{\mathfrak{W}} \varphi$ does not hold by the R-soundness of $\mathfrak{W}$. Consequently, we can utilize the same two-tuple relation r_φ as in the proof of Theorem 3 to derive that $\Sigma \models^R_{2-R_s} \varphi$ does not hold.

4.3 A Weaker Version of R-complementation

Let $\dfrac{}{\emptyset : \{R\}}$ be the R-axiom for FHDs. This inference rule is R-sound for the R-implication of FHDs for all relation schemata R. As it turns out, we can simply replace the *R-complementation rule* $\mathcal{C}^R_{\mathrm{H}}$ in $\mathfrak{W}$ by the *R-axiom* and still maintain R-completeness for all R.

Theorem 4 *The set*

$$\mathfrak{W}_0 = \{\mathcal{R}_{\mathrm{F}}, \mathcal{A}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{A}_{\mathrm{H}}, \mathcal{O}_{\mathrm{H}}, \mathcal{T}_{\mathrm{H}}, R - \mathit{axiom}, \mathcal{I}_{\mathrm{FH}}, \mathcal{T}_{\mathrm{FH}}\},$$

of inference rules forms a finite axiomatization for the R-implication of FDs and FHDs in the presence of an NFS.

Proof The proof follows immediately from Theorem 3 and the following inference

$$\dfrac{\dfrac{\dfrac{X : \{Y_1, \ldots, Y_k\}}{\mathcal{O}_{\mathrm{H}} :\ X : \{Y_1, \ldots, Y_{k-1}\}}}{\mathcal{I}_{\emptyset} :\ X : \{Y_1, \ldots, Y_{k-1}, \emptyset\}} \quad \dfrac{\dfrac{\dfrac{X : \{Y_1, \ldots, Y_k\}}{\mathcal{M}_{\mathrm{H}} :\ X : \{Y_1 \cdots Y_k\}} \quad \dfrac{\overline{R - \text{axiom} :\ \emptyset : \{R\}}}{\mathcal{A}_{\mathrm{H}} :\ X : \{R - X\}}}{\mathcal{T}_{\mathrm{H}} :\ X : \{R - XY_1 \cdots Y_k, Y_1 \cdots Y_k\}}}{\mathcal{O}_{\mathrm{H}} :\ X : \{R - XY_1 \cdots Y_k\}}}{\mathcal{U}_{\mathrm{H}} :\ X : \{Y_1, \ldots, Y_{k-1}, R - XY_1 \cdots Y_k\}}$$

of the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ from the R-axiom and $\mathfrak{W} - \{\mathcal{C}^R_{\mathrm{H}}\}$.

5 From Inappropriate to Appropriate Inference Systems

We have seen before that the inference system

$$\mathfrak{W} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{A}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{A}_{\mathrm{H}}, \mathcal{O}_{\mathrm{H}}, \mathcal{T}_{\mathrm{H}}, \mathcal{C}^R_{\mathrm{H}}, \mathcal{I}_{\mathrm{FH}}, \mathcal{T}_{\mathrm{FH}}\},$$

forms an axiomatization for the R-implication of the combined class of FDs and FHDs in the presence of an NFS. In this section, we will analyze the role of the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ in the combined setting of FDs, FHDs and NFSs.

5.1 The notion of an appropriate inference system

A quick inspection shows that the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ is the only rule in $\mathfrak{W}$ that depends on the underlying relation schema R. This raises the question to which degree applications of the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ are necessary to infer FDs and FHDs. In particular, if there are inferences of FDs and FHDs in which the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ does not need to be applied, then the inferred dependencies are already implied by Σ in the presence of the NFS without having to fix the underlying relation schema R.

Example 10 *Consider again the relation schema $R = MDAFL$ and the set $\Sigma = \{M : \{DF\}, MD : \{FL\}, MA \to D\}$ of FDs and FHDs over R from Example 1. Let $R_s = AD$ denote a null-free subschema over R. The inference of Example 1 shows that the FHD $M : \{A\}$ and FD $M \to D$ can be inferred from Σ in the presence of R_s by $\mathfrak{W}$. Since the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ is applied in both inferences, it is unclear whether either of these dependencies is already implied by Σ in the presence of R_s without fixing the relation schema R.* ▮

The goal of this section is to establish an axiomatization for the R-implication of FDs and FHDs in the presence of an NFS that appropriately reflects the role of the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$. For this purpose, we assume now that sets $\mathfrak{R}$ of inference rules we consider do not contain rules that are dependent on the underlying relation schema R with the exception of the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$. For example, $\mathfrak{W}$ is such an axiomatization. First we extend the notion of an *appropriate inference system* [17] to the presence of an arbitrary NFS.

Definition 6 *Let $\mathfrak{R}$ denote a set of inference rules that is R-sound for the R-implication of FDs and FHDs in the presence of an NFS. $\mathfrak{R}$ is said to be* complementary *for the R-implication of FDs and FHDs in the presence of an NFS if for every relation schema R, for every NFS R_s over R, for every set Σ of FDs and FHDs over R, and for every FHD φ over R such that φ is R-implied by Σ in the presence of R_s there is an inference of φ from Σ by $\mathfrak{R}$ in which the R-complementation rule $\mathcal{C}^R_H$ is applied at most once and if it is applied, then it is applied only in the very last step of the inference. $\mathfrak{R}$ is said to be* adequate *for the R-implication of FDs and FHDs in the presence of an NFS if for every relation schema R, for every NFS R_s over R, for every set Σ of FDs and FHDs over R,* *and for every FD φ over R such that φ is R-implied by Σ in the presence of R_s there is an inference of φ from Σ by $\mathfrak{R}$ in which the R-complementation rule $\mathcal{C}_H^R$ is not applied at all. $\mathfrak{R}$ is said to be* appropriate *for the R-implication of FDs and FHDs in the presence of an NFS if $\mathfrak{R}$ is both complementary and adequate.* ∎

5.2 $\mathfrak{W}$ is an inappropriate axiomatization

An appropriate set of inference rules is always R-complete. However, an R-complete set of inference rules does not need to be neither complementary nor adequate. In this subsection we will show that this is indeed the case for our axiomatization $\mathfrak{W}$. The next lemma shows that $\mathfrak{W}$ is not complementary.

Lemma 3 *There is a relation schema R, an NFS R_s, and a set $\Sigma \cup \{\varphi\}$ of FHDs over R such that $\varphi \in \Sigma^+_{\mathfrak{W}} - \Sigma^+_{\mathfrak{W}-\{\mathcal{C}^R_{\mathrm{H}}\}}$, but there is no inference of φ from Σ by $\mathfrak{W}$ in which the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ is only applied in the last step.*

Proof Let Σ consist of the two FHDs

$$\mathit{Movie} : \{\{\mathit{Actor}\}, \{\mathit{Feature}\}\} \quad \text{and} \quad \mathit{Movie} : \{\{\mathit{Actor}\}, \{\mathit{Language}\}\}.$$

An inspection of the inference rules in $\mathfrak{W} - \{\mathcal{C}^R_{\mathrm{H}}\}$ shows that $\mathit{Movie} : \{\{\mathit{Actor}\}, \{\mathit{Feature,Language}\}\} \notin \Sigma^+_{\mathfrak{W}-\{\mathcal{C}^R_{\mathrm{H}}\}}$. Moreover, Lemma 9 shows that $\mathit{Movie} : \{\{\mathit{Actor}\}, Y\} \notin \Sigma^+_{\mathfrak{W}-\{\mathcal{C}^R_{\mathrm{H}}\}}$ for any Y such that

$$Y - \{\mathit{Actor, Feature, Language}\} \neq \emptyset\ .$$

For DVD=$\{$*Movie, Director, Actor, Feature, Language*$\}$ we have

$$\mathit{Movie} : \{\{\mathit{Actor}\}, \{\mathit{Feature,Language}\}\} \in \Sigma^+_{\mathfrak{W}}.$$

Hence, in any such inference the DVD-complementation rule $\mathcal{C}^{\mathrm{DVD}}_{\mathrm{H}}$ must be applied at least once. However, since

$$\mathrm{DVD} - \{\mathit{Movie,Actor,Feature,Language}\} = \{\mathit{Director}\}$$

$\mathcal{C}^{\mathrm{DVD}}_{\mathrm{H}}$ is not just applied in the last step of the inference.

The next lemma shows that the system $\mathfrak{W}$ is not adequate.

Lemma 4 *There is a relation schema R, an NFS R_s, a set Σ of FDs and FHDs, and an FD φ over R such that $\varphi \in \Sigma^+_{\mathfrak{W}}$ but $\varphi \notin \Sigma^+_{\mathfrak{W}-\{\mathcal{C}^R_{\mathrm{H}}\}}$.*

Proof Let $R = AB$, $R_s = AB$, and $\Sigma = \{\emptyset : \{A\}, B \to A\}$ and $\varphi = \emptyset \to A$. We show first that $\varphi \notin \Sigma^+_{\mathfrak{W}-\{\mathcal{C}^R_{\mathrm{H}}\}}$. We represent the closure $\Sigma^+_{\mathfrak{W}-\{\mathcal{C}^R_{\mathrm{H}}\}}$ of Σ with respect to $\mathfrak{W} - \{\mathcal{C}^R_{\mathrm{H}}\}$ as two tables. The FHD $X : \{Y\}$ (FD $X \to Y$) belongs to $\Sigma^+_{\mathfrak{W}-\{\mathcal{C}^R_{\mathrm{H}}\}}$ if and only if in the :-table ($\to$-table) the entry in row labeled X and column labeled Y is the symbol $\circ$. Due to the definition of FDs and FHDs some entries do not correspond to any dependencies, these are marked by $\times$. The $\rightarrow$-table can be obtained as follows. First, we enter the premise $B \rightarrow A$ from Σ. Then, we enter the FD $\emptyset \rightarrow \emptyset$ which results from an application of the empty-set-axiom $\mathcal{R}_{\mathrm{F}}$. Finally, we enter the FDs $A \rightarrow \emptyset$, $B \rightarrow \emptyset$ and $AB \rightarrow \emptyset$ which results from an application of the augmentation rule $\mathcal{A}_{\mathrm{F}}$ to $\emptyset \rightarrow \emptyset$, respectively. Applications of other rules do not result in new FDs. The :-table can be obtained as follows. First, we apply $\mathcal{I}_{\mathrm{FH}}$ to copy all $\circ$ from the $\rightarrow$-table into the corresponding entries in the :-table. Finally, we enter the premise $\emptyset : \{A\}$ from Σ. This set is closed under inference using $\mathfrak{W} - \{\mathcal{C}_{\mathrm{H}}^{R}\}$. In particular, φ cannot be inferred from Σ by using $\mathfrak{W} - \{\mathcal{C}_{\mathrm{H}}^{R}\}$. 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 $\mathcal{T}_{\mathrm{FH}}$, but in order to apply this rule the R-complementation rule $\mathcal{C}_{\mathrm{H}}^{R}$ must first be applied to $\emptyset : \{A\}$. However, $\mathcal{C}_{\mathrm{H}}^{R}$ is not available in $\mathfrak{W} - \{\mathcal{C}_{\mathrm{H}}^{R}\}$.

$\rightarrow$	$\emptyset$	A	B	AB
$\emptyset$	$\circ$			
A	$\circ$	$\times$		$\times$
B	$\circ$	$\circ$	$\times$	$\times$
AB	$\circ$	$\times$	$\times$	$\times$

:	$\emptyset$	A	B	AB
$\emptyset$	$\circ$	$\circ$		
A	$\circ$	$\times$		$\times$
B	$\circ$	$\circ$	$\times$	$\times$
AB	$\circ$	$\times$	$\times$	$\times$

It remains to verify that $\varphi \in \Sigma^{+}_{\mathfrak{W}}$. First, we apply $\mathcal{C}_{\mathrm{H}}^{R}$ to $\emptyset : \{A\}$ to infer $\emptyset : \{B\}$. Subsequently, we apply $\mathcal{T}_{\mathrm{FH}}$ to $\emptyset : \{B\}$ and $B \rightarrow A$ and infer $\emptyset \rightarrow A$.

Corollary 2 *The system $\mathfrak{W}$ is neither complementary nor adequate for the R-implication of FDs and FHDs in the presence of an NFS.* ∎

Corollary 2 raises the question whether there is any complementary or adequate (or even appropriate) set of inference rules for the R-implication of FDs and FHDs in the presence of an NFS.

5.3 Appropriate Reasoning about FDs and FHDs in the presence of NFSs

In this section we will establish an appropriate inference system for the R-implication of FDs and FHDs in the presence of an NFS.

Lemma 5 *The following inference rules*

$$\frac{X : \{W\} \quad XY : \{Y_1, \dots, Y_k\}}{X : \{Y_1 \cap W, \dots, Y_k \cap W, W - Y_1 \cdots Y_k\}}\; Y \cap W = \emptyset, Y \subseteq R_s$$
(subset rule, $\mathcal{S}_{\mathrm{H}}$)

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

are R-sound for the R-implication of FDs and FHDs in the presence of an NFS.

Proof We show that both rules can be derived from $\mathfrak{W}$, and are therefore R-sound. First, we show that

$$\mathcal{S}^1_{\mathrm{H}}: \qquad \frac{X:\{W\} \quad XY:\{Y_1,\dots,Y_k\}}{X:\{Y_1\cap W,\dots,Y_k\cap W\}}\, Y\cap W=\emptyset, Y\subseteq R_s$$

is derivable from $\mathfrak{W}$:

$$\dfrac{\dfrac{\dfrac{X:\{W\}}{\mathcal{C}^R_{\mathrm{H}}:\ X:\{R-XW\}} \qquad XY:\{Y_1,\dots,Y_k\}}{\mathcal{T}_{\mathrm{H}}:\ X:\{Y_1-(R-XW),\dots,Y_k-(R-XW),R-XW\}}}{\mathcal{O}_{\mathrm{H}}:\quad X:\{\underbrace{Y_1-(R-XW)}_{=Y_1\cap W},\dots,\underbrace{Y_k-(R-XW)}_{=Y_k\cap W}\}} \quad .$$

Next, we show that

$$\mathcal{S}^2_{\mathrm{H}}: \qquad \frac{X:\{W\} \quad XY:\{Y_1,\dots,Y_k\}}{X:\{W-Y_1\cdots Y_k\}}\, Y\cap W=\emptyset, Y\subseteq R_s$$

is derivable from $\mathfrak{W}$:

$$\dfrac{X:\{W\} \quad \dfrac{X:\{W\} \quad \dfrac{XY:\{Y_1,\dots,Y_k\}}{\mathcal{M}_{\mathrm{H}}:\ XY:\{Y_1\cdots Y_k\}}}{\mathcal{S}^1_{\mathrm{H}}:\quad X:\{Y_1\cdots Y_k\cap W\}}}{\mathcal{D}_{\mathrm{H}}:\quad X:\{\underbrace{W-(Y_1\cdots Y_k\cap W)}_{=W-Y_1\cdots Y_k}\}} \quad .$$

Next we show that the subset rule $\mathcal{S}_{\mathrm{H}}$ is derivable from $\mathfrak{W}$:

$$\dfrac{\dfrac{\dfrac{X:\{W\} \qquad XY:\{Y_1,\dots,Y_k\}}{\mathcal{S}^1_{\mathrm{H}}:\ X:\{Y_1\cap W,\dots,Y_k\cap W\}}}{\mathcal{I}_{\emptyset}:\ X:\{Y_1\cap W,\dots,Y_k\cap W,\emptyset\}} \quad \dfrac{X:\{W\} \quad XY:\{Y_1,\dots,Y_k\}}{\mathcal{S}^2_{\mathrm{H}}:\quad X:\{W-Y_1\cdots Y_k\}}}{\mathcal{U}_{\mathrm{H}}:\qquad X:\{Y_1\cap W,\dots,Y_k\cap W,W-Y_1\cdots Y_k\}} \quad .$$

Finally, we show that the mixed subset rule $\mathcal{S}_{\mathrm{FH}}$ is derivable from $\mathfrak{W}$:

$$\dfrac{\dfrac{X:\{W\}}{\mathcal{C}^R_{\mathrm{H}}:\ X:\{R-XW\}} \quad XY\to Z}{\mathcal{T}_{\mathrm{FH}}:\quad X\to \underbrace{Z-(R-XW)}_{=W\cap Z}} \quad .$$

The completes the proof of the lemma.

We will first show that $\mathcal{S}_{\mathrm{FH}}$ is independent of many other rules.

Lemma 6 *The* mixed subset rule $\mathcal{S}_{\mathrm{FH}}$ *is independent of* $\mathfrak{W}-\{\mathcal{C}^R_H\}$.

Proof Let $R = \{A, B\}$, $R_s = AB$, $\Sigma = \{\emptyset : \{A\}, B \to A\}$ and $\varphi = \emptyset \to A$. The proof of Lemma 4 shows that $\varphi \notin \Sigma^+_{\mathfrak{W}-\{\mathcal{C}^R_{\mathrm{H}}\}}$ but a single application of the mixed subset rule $\mathcal{S}_{\mathrm{FH}}$ to the premises in Σ shows that $\varphi \in \Sigma^+_{(\mathfrak{W}-\{\mathcal{C}^R_{\mathrm{H}}\})\cup\{\mathcal{S}_{\mathrm{FH}}\}}$.

Lemma 7 *The* mixed transitivity rule $\mathcal{T}_{\mathrm{FH}}$ *is derivable from* $\{\mathcal{T}_{\mathrm{H}}, \mathcal{O}_{\mathrm{H}}, \mathcal{I}_{\mathrm{FH}}, \mathcal{S}_{\mathrm{FH}}\}$.

Proof

$$
\dfrac{\dfrac{\dfrac{X : \{W\} \qquad \dfrac{XY \to Z}{\mathcal{I}_{\mathrm{FH}} :\ XY : \{Z\}}}{\mathcal{T}_{\mathrm{H}} :\quad X : \{Z - W, W\}}}{\mathcal{O}_{\mathrm{H}} :\quad X : \{Z - W\}} \qquad XY \to Z}{\mathcal{S}_{\mathrm{FH}} :\quad X \to Z - W}
$$

Note that $Y \cap (Z - W) = \emptyset$ holds, in particular. This completes the proof.

Using Theorem 3, Lemma 7 enables us to obtain another axiomatization for the R-implication of FDs and FHDs in the presence of an NFS: just replace $\mathcal{T}_{\mathrm{FH}}$ in $\mathfrak{W}$ by $\mathcal{S}_{\mathrm{FH}}$.

Corollary 3 *The set* $\{\mathcal{R}_{\mathrm{F}}, \mathcal{A}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{A}_{\mathrm{H}}, \mathcal{T}_{\mathrm{H}}, \mathcal{O}_{\mathrm{H}}, \mathcal{C}^R_{\mathrm{H}}, \mathcal{I}_{\mathrm{FH}}, \mathcal{S}_{\mathrm{FH}}\}$ *of inference rules forms a finite axiomatization for the R-implication of FDs and FHDs in the presence of an NFS.* ∎

We will now formally establish an appropriate axiomatization for the combined class of functional and full hierarchical dependencies in the presence of an NFS.

Theorem 5 *Let R be a relation schema, Σ a set of FDs and FHDs, and R_s an NFS over R. For every inference γ from Σ by the system*

$$\mathfrak{W} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{A}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{A}_{\mathrm{H}}, \mathcal{O}_{\mathrm{H}}, \mathcal{T}_{\mathrm{H}}, \mathcal{C}^R_{\mathrm{H}}, \mathcal{I}_{\mathrm{FH}}, \mathcal{T}_{\mathrm{FH}}\},$$

there is an inference ξ from Σ by the system

$$\mathfrak{W}_{\mathcal{C}} = (\mathfrak{W}-\{\mathcal{T}_{\mathrm{FH}}\})\cup\{\mathcal{M}_{\mathrm{H}}, \mathcal{S}_{\mathrm{H}}, \mathcal{S}_{\mathrm{FH}}\} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{A}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{A}_{\mathrm{H}}, \mathcal{O}_{\mathrm{H}}, \mathcal{T}_{\mathrm{H}}, \mathcal{S}_{\mathrm{H}}, \mathcal{M}_{\mathrm{H}}, \mathcal{C}^R_{\mathrm{H}}, \mathcal{I}_{\mathrm{FH}}, \mathcal{S}_{\mathrm{FH}}\},$$

with the following properties:

1. *if γ infers an FHD, then*
   - *γ and ξ infer the same FHD,*
   - *in ξ the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ is applied at most once, and*
   - *if $\mathcal{C}^R_{\mathrm{H}}$ is applied in ξ, then $\mathcal{C}^R_{\mathrm{H}}$ 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 $\mathcal{C}^R_{\mathrm{H}}$ is not applied at all.*

Proof We proceed by induction on the length l of 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 $\mathfrak{W}$ which has length l. We consider ten cases according to which inference rule in $\mathfrak{W}$ was applied to infer σ_l from $[\sigma_1, \ldots, \sigma_{l-1}]$.

Case 1. In this case, σ_l is either an element of Σ or the FD $\emptyset \to \emptyset$ obtained by an application of the *empty-set-axiom* $\mathcal{R}_{\mathrm{F}}$. It follows immediately that $\xi = [\sigma_l]$ has the desired properties.

Cases 2-4. In these cases σ_l has been inferred by an application of one of the inference rules that deal with FDs only, i.e., the *augmentation rule* $\mathcal{A}_{\mathrm{F}}$, the *decomposition rule* $\mathcal{D}_{\mathrm{F}}$ or the *FD union rule* $\mathcal{U}_{\mathrm{F}}$ to one or two premises σ_i and σ_j where $i, j < l$. Let ξ_i respectively ξ_j be obtained by applying the induction hypothesis to $\gamma_i = [\sigma_1, \ldots, \sigma_i]$ respectively $\gamma_j = [\sigma_1, \ldots, \sigma_j]$. It follows that $\xi = [\gamma_i, \gamma_j, \sigma_l]$ has the desired properties.

Cases 5-10. Whenever the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ is applied to an FHD $X : S$ where the union over X and the elements of S covers R, the newly introduced complement set $R - XY_1 ... Y_k$ is empty and thus, following our global conditions, immediately removed. Accordingly, we can always apply the omission rule $\mathcal{O}_{\mathrm{H}}$ to infer the same conclusion. Consequently, we will assume for the remainder of the proof that for every FHD $X : S$ to which the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ is applied the union over X and the elements of S does not cover R.

Case 5. We infer σ_l by applying the *augmentation rule* $\mathcal{A}_{\mathrm{H}}$ to the premise σ_i with $i < l$. Let ξ_i be obtained by using the induction hypothesis for $\gamma_i := [\sigma_1, \ldots, \sigma_i]$.

Consider the inference $\xi := [\xi_i, \sigma_l]$. If $\mathcal{C}^R_{\mathrm{H}}$ is not applied in ξ_i, then ξ has the desired properties. If $\mathcal{C}^R_{\mathrm{H}}$ is applied in ξ_i (as the last rule), then the last two steps of ξ are of the following form:

$$\cfrac{\cfrac{X : \{Y_1, \ldots, Y_k\}}{\mathcal{C}^R_{\mathrm{H}} :\ X : \{Y_1, \ldots, Y_{k-1}, R - XY_1 \cdots Y_k\}}}{\mathcal{A}_{\mathrm{H}} :\ XZ : \{Y_1 - Z, \ldots, Y_{k-1} - Z, R - XZY_1 \cdots Y_k\}} .$$

However, these steps can be replaced as follows:

$$\cfrac{\cfrac{X : \{Y_1, \ldots, Y_k\}}{\mathcal{A}_{\mathrm{H}} :\ XZ : \{Y_1 - Z, \ldots, Y_k - Z\}}}{\mathcal{C}^R_{\mathrm{H}} :\ XZ : \{Y_1 - Z, \ldots, Y_{k-1} - Z, \underbrace{R - XZ(Y_1 - Z) \cdots (Y_k - Z)}_{=R - XZY_1 \cdots Y_k}\}} .$$

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

Case 6. We infer σ_l by applying the *omission rule* $\mathcal{O}_{\mathrm{H}}$ to the premise σ_i with $i < l$. Let ξ_i be obtained by using the induction hypothesis for $\gamma_i := [\sigma_1, \ldots, \sigma_i]$.

Case 6.1. Consider the inference $\xi := [\xi_i, \sigma_l]$. If $\mathcal{C}^R_{\mathrm{H}}$ is not applied in ξ_i, then ξ has the desired properties.

Case 6.2. If $\mathcal{C}^R_{\mathrm{H}}$ is applied in ξ_i as the last rule, then the last two steps of ξ have either the form

$$\cfrac{\cfrac{X : \{Y_1, \ldots, Y_k, Y\}}{\mathcal{C}^R_{\mathrm{H}} :\ X : \{Y_1, \ldots, Y_k, R - XYY_1 \cdots Y_k\}}}{\mathcal{O}_{\mathrm{H}} :\qquad X : \{Y_1, \ldots, Y_k\}}$$

or the form

$$\cfrac{\cfrac{X:\{Y_1,\ldots,Y_k,Y\}}{\mathcal{C}_{\mathrm{H}}^R:\ X:\{Y_1,\ldots,Y_k,R-XYY_1\cdots Y_k\}}}{\mathcal{O}_{\mathrm{H}}:\ XZ:\{Y_1,\ldots,Y_{i-1},Y_{i+1},\ldots,Y_k,R-XYY_1\cdots Y_k\}}\ .$$

In the first case these steps may be simply replaced by

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

In the second case, these steps can be replaced as follows:

$$\cfrac{\cfrac{X:\{Y_1,\ldots,Y_k,Y\}}{\mathcal{M}_{\mathrm{H}}:\ X:\{Y_1,\ldots,Y_{i-1},Y_{i+1},\ldots,Y_k,Y_iY\}}}{\mathcal{C}_{\mathrm{H}}^R:\ X:\{Y_1,\ldots,Y_{i-1},Y_{i+1},\ldots,Y_k,R-XYY_1\cdots Y_k\}}\ .$$

In both cases the result of these replacements is an inference with the desired properties.

Case 7. We infer σ_l by applying the *transitivity rule* $\mathcal{T}_{\mathrm{H}}$ 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 four cases according to the occurrence of the R-complementation rule $\mathcal{C}_{\mathrm{H}}^R$ in ξ_i and ξ_j.

Case 7.1. If $\mathcal{C}_{\mathrm{H}}^R$ is applied neither in ξ_i nor in ξ_j, then ξ has the desired properties.

Case 7.2. If $\mathcal{C}_{\mathrm{H}}^R$ is not applied in ξ_i but is applied in ξ_j (as the last rule), then the last step of ξ_j and the last step of ξ are of the following form:

$$\cfrac{X:\{W\}\qquad \cfrac{XY:\{Y_1,\ldots,Y_k\}}{\mathcal{C}_{\mathrm{H}}^R:\ XY:\{Y_1,\ldots,Y_{k-1},R-XYY_1\cdots Y_k\}}}{\mathcal{T}_{\mathrm{H}}:\ X:\{Y_1-W,\ldots,Y_{k-1}-W,R-XWY_1\cdots Y_k,W\}}$$

where $Y\subseteq W\cap R_s$ holds. However, these steps can be replaced as follows:

$$\cfrac{\cfrac{X:\{W\}\qquad XY:\{Y_1,\ldots,Y_k\}}{\mathcal{T}_{\mathrm{H}}:\quad X:\{Y_1-W,\ldots,Y_k-W,W\}}}{\mathcal{C}_{\mathrm{H}}^R:\ X:\{Y_1-W,\ldots,Y_{k-1}-W,R-XWY_1\cdots Y_k,W\}}\ .$$

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

Case 7.3. If $\mathcal{C}_{\mathrm{H}}^R$ is applied in ξ_i (as the last rule) but not in ξ_j, then the last step of ξ_i and the last step of ξ are of the following form:

$$\cfrac{\cfrac{X:\{W\}}{\mathcal{C}_{\mathrm{H}}^R:\ X:\{R-XW\}}\qquad XY:\{Y_1,\ldots,Y_k\}}{\mathcal{T}_{\mathrm{H}}:\ X:\{\underbrace{Y_1-(R-XW)}_{=Y_1\cap W},\ldots,\underbrace{Y_k-(R-XW)}_{=Y_k\cap W},R-XW\}}$$

where $Y \subseteq (R - XW) \cap R_s$. In particular, it follows that $Y \cap W = \emptyset$ holds. Consequently, these steps can be replaced as follows:

$$\dfrac{\dfrac{X : \{W\} \qquad\qquad XY : \{Y_1, \ldots, Y_k\}}{\mathcal{S}_{\mathrm{H}} : \quad X : \{Y_1 \cap W, \ldots, Y_k \cap W, W - Y_1 \cdots Y_k\}}}{\mathcal{C}^R_{\mathrm{H}} : \; X : \{Y_1 \cap W, \ldots, Y_k \cap W, \underbrace{R - X(Y_1 \cap W) \cdots (Y_k \cap W)(W - Y_1 \cdots Y_k)}_{=R-XW}\}} .$$

Note that X and W are disjoint. The result of this replacement is an inference with the desired properties.

Case 7.4. If $\mathcal{C}^R_{\mathrm{H}}$ is applied in both ξ_i and ξ_j (as the last rule), then the last steps of ξ_i, ξ_j and ξ are of the following form:

$$\dfrac{\dfrac{X : \{W\}}{\mathcal{C}^R_{\mathrm{H}} : X : \{R - XW\}} \qquad \dfrac{XY : \{Y_1, \ldots, Y_k\}}{\mathcal{C}^R_{\mathrm{H}} : XY : \{Y_1, \ldots, Y_{k-1}, R - XYY_1 \cdots Y_k\}}}{\mathcal{T}_{\mathrm{H}} : \; X : \{\underbrace{Y_1 - (R - XW)}_{=Y_1 \cap W}, \ldots, \underbrace{Y_{k-1} - (R - XW)}_{=Y_{k-1} \cap W}, \underbrace{(R - XYY_1 \cdots Y_k) - (R - XW)}_{=W - Y_1 \cdots Y_k}, R - XW\}} .$$

Note that $X \cap W = \emptyset$ and $Y \subseteq (R - XW) \cap R_s$, i.e. $Y \cap W = \emptyset$. Hence, these steps can be replaced as follows:

$$\dfrac{\dfrac{X : \{W\} \qquad\qquad XY : \{Y_1, \ldots, Y_k\}}{\mathcal{S}_{\mathrm{H}} : \quad X : \{Y_1 \cap W, \ldots, Y_k \cap W, W - Y_1 \ldots Y_k\}}}{\mathcal{C}^R_{\mathrm{H}} : \; X : \{Y_1 \cap W, \ldots, Y_{k-1} \cap W, W - Y_1 \ldots Y_k, \underbrace{R - X(Y_1 \cap W) \cdots (Y_k \cap W)(W - Y_1 \cdots Y_k)}_{=R-XW}\}} .$$

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

Case 8. We infer σ_l by applying the *R-complementation rule* $\mathcal{C}^R_{\mathrm{H}}$ to the premise σ_i with $i < l$. Let ξ_i be obtained by using the induction hypothesis for $\gamma_i := [\sigma_1, \ldots, \sigma_i]$. Consider the inference $\xi := [\xi_i, \sigma_l]$.

Case 8.1. If $\mathcal{C}^R_{\mathrm{H}}$ is not applied in ξ_i, then ξ has the desired properties.

Case 8.2. If $\mathcal{C}^R_{\mathrm{H}}$ is applied in ξ_i as the last rule, then the last two steps of ξ are either of the following form:

$$\dfrac{\dfrac{X : \{Y_1, \ldots, Y_k\}}{\mathcal{C}^R_{\mathrm{H}} : \; X : \{Y_1, \ldots, Y_{k-1}, R - XY_1 \cdots Y_k\}}}{\mathcal{C}^R_{\mathrm{H}} : \; X : \{Y_1, \ldots, Y_{k-1}, \underbrace{R - XY_1 \cdots Y_{k-1}(R - XY_1 \cdots Y_k)}_{=Y_k}\}}$$

or of the form:

$$\dfrac{\dfrac{X : \{Y_1, \ldots, Y_k\}}{\mathcal{C}^R_{\mathrm{H}} : \; X : \{Y_1, \ldots, Y_{k-1}, R - XY_1 \cdots Y_k\}}}{\mathcal{C}^R_{\mathrm{H}} : \; X : \{Y_1, \ldots, Y_{k-2}, \underbrace{R - XY_1 \cdots Y_{k-1}(R - XY_1 \cdots Y_k)}_{=Y_k}, R - XY_1 \cdots Y_k\}} .$$

In the first case, the inference obtained by deleting these two steps from ξ has the desired properties. In the second case, the inference can be replaced by the following step:

$$\frac{X : \{Y_1, \ldots, Y_{k-2}, Y_k, Y_{k-1}\}}{\mathcal{C}^R_{\mathrm{H}} :\; X : \{Y_1, \ldots, Y_{k-2}, Y_k, R - XY_1 \cdots Y_k\}} \quad .$$

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

Case 9. In this case σ_l has been inferred by an application of the *implication rule* $\mathcal{I}_{\mathrm{FH}}$ to the premise σ_i where $i < l$. This case follows the same structure of *Cases 2-5.* Let ξ_i be obtained by applying the induction hypothesis to $\gamma_i = [\sigma_1, \ldots, \sigma_i]$. It follows that the inference $\xi = [\gamma_i, \sigma_l]$ meets the desired properties.

Case 10. In this case σ_l has been inferred by an application of the *mixed transitivity rule* $\mathcal{T}_{\mathrm{FH}}$ to the premises σ_i and σ_j where $i, j < l$. Let ξ_i respectively ξ_j be obtained by applying the induction hypothesis for $\gamma_i = [\sigma_1, \ldots, \sigma_i]$ respectively $\gamma_j = [\sigma_1, \ldots, \sigma_j]$. 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 $\mathcal{C}^R_{\mathrm{H}}$ in ξ_i (assuming that ξ_j infers the FD in the premise).

Case 10.1. If $\mathcal{C}^R_{\mathrm{H}}$ is not applied in ξ_i, then ξ has the desired properties.

Case 10.2. If $\mathcal{C}^R_{\mathrm{H}}$ is applied in ξ_i as the last rule, then the last step of ξ_i and the last step of ξ are of the following form:

$$\frac{\dfrac{X : \{W\}}{\mathcal{C}^R_{\mathrm{H}} :\; X : \{R - XW\}} \quad XY \to Z}{\mathcal{T}_{\mathrm{FH}} : \quad X \to \underbrace{Z - (R - XW)}_{= Z \cap W}}$$

where $Y \subseteq (R - XW) \cap R_s$ and $Z \cap X = \emptyset$ hold. In particular, $Y \cap W = \emptyset$. Applying the mixed subset rule $\mathcal{S}_{\mathrm{FH}}$ instead one may infer the same FHD by the following inference steps:

$$\frac{X : \{W\} \quad XY \to Z}{\mathcal{S}_{\mathrm{FH}} :\; X \to W \cap Z} \quad .$$

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

Corollary 4 *The set $\mathfrak{W}_{\mathcal{C}}$ of inference rules forms a finite appropriate axiomatization for the R-implication of FDs and FHDs in the presence of an NFS.* ∎

Example 11 *Consider again Example 10 where $R = MDAFL$, $R_s = AD$, and $\Sigma = \{M : \{DF\}, MD : \{FL\}, MA \to D\}$. The following is an inference of $M : \{A\}$ where $\mathcal{C}^R_{\mathrm{H}}$ is applied in the last step only:*

$$\frac{\dfrac{\dfrac{M : \{DF\} \quad MD : \{FL\}}{\mathcal{T}_{\mathrm{H}} : \quad M : \{L, DF\}}}{\mathcal{M}_{\mathrm{H}} : \quad M : \{DFL\}}}{\mathcal{C}^R_{\mathrm{H}} : \quad M : \{A\}} \quad .$$

Moreover, the following is an inference of $MA \to D$ without an application of $\mathcal{C}^R_{\mathrm{H}}$:

$$\cfrac{\cfrac{\cfrac{M:\{DF\} \quad MD:\{FL\}}{\mathcal{T}_{\mathrm{H}}: \quad M:\{L,DF\}}}{\mathcal{M}_{\mathrm{H}}: \quad M:\{DFL\}} \qquad MA \to D}{\mathcal{S}_{\mathrm{FH}}: \qquad M \to D}\ .$$

Note that $DA \subseteq R_s$ and $A \cap LDS = \emptyset$. The examples highlight how the new rules can be applied to guarantee appropriate inferences. In particular, the inference confirms our intuition that the FD $M \to D$ is already implied by Σ in the presence of R_s without fixing the relation schema R. However, the question remains whether this is also the case for the FHD $M : \{A\}$. $\blacksquare$

Figure 2 illustrates the connection between the different inference systems and their semantic properties. In summary, one gains complementarity by including the subset rule $\mathcal{S}_{\mathrm{H}}$ and merging rule $\mathcal{M}_{\mathrm{H}}$, and adequacy by replacing the mixed transitivity rule $\mathcal{T}_{\mathrm{FH}}$ by the mixed subset rule $\mathcal{S}_{\mathrm{FH}}$.

$\mathfrak{W}_{\mathcal{C}} = (\mathfrak{W}' - \{\mathcal{T}_{\mathrm{FH}}\}) \cup \{\mathcal{S}_{\mathrm{FH}}\}$
appropriate: adequate and complementary

$\mathfrak{W}' = \mathfrak{W} \cup \{\mathcal{S}_{\mathrm{H}}, \mathcal{M}_{\mathrm{H}}\}$
complementary

$\mathfrak{W} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{A}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{A}_{\mathrm{H}}, \mathcal{T}_{\mathrm{H}}, \mathcal{U}_{\mathrm{H}}, \mathcal{C}^R_{\mathrm{H}}, \mathcal{I}_{\mathrm{FH}}, \mathcal{T}_{\mathrm{FH}}\}$

Figure 2: Axiomatizations for FDs and FHDs in the presence of an NFS and their properties

5.4 Nearly complete reasoning in fixed universes

Among others Theorem 5 shows that

$$\mathfrak{U} = \mathfrak{W}_{\mathcal{C}} - \{\mathcal{C}^R_{\mathrm{H}}\} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{A}_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}, \mathcal{U}_{\mathrm{F}}, \mathcal{A}_{\mathrm{H}}, \mathcal{O}_{\mathrm{H}}, \mathcal{T}_{\mathrm{H}}, \mathcal{S}_{\mathrm{H}}, \mathcal{M}_{\mathrm{H}}, \mathcal{I}_{\mathrm{FH}}, \mathcal{S}_{\mathrm{FH}}\}$$

is nearly R-complete for the R-implication of FDs and FHDs in the presence of an NFS over any relation schema R. Indeed, $\mathfrak{U}$ enables us to infer every R-implied FD. Moreover, for every R-implied FHD $X : \{Y_1, \ldots, Y_k\}$ the system $\mathfrak{U}$ enables us to infer $X : \{Y_1, \ldots, Y_k\}$ itself or $X : \{Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k, R - XY_1 \cdots Y_k\}$ for some $i \in \{1, \ldots, k\}$.

Corollary 5 *Let $\Sigma \cup \{\varphi\}$ be a set of FDs and FHDs over relation schema R. Then*

- *If φ denotes an FD, then $\varphi \in \Sigma^+_{\mathfrak{W}_{\mathcal{C}}}$ if and only if $\varphi \in \Sigma^+_{\mathfrak{U}}$.*

- *If φ denotes the FHD $X : \{Y_1, \ldots, Y_k\}$, then $X : \{Y_1, \ldots, Y_k\} \in \Sigma^+_{\mathfrak{W}_\mathcal{C}}$ if and only if $X : \{Y_1, \ldots, Y_k\} \in \Sigma^+_{\mathfrak{U}}$ or there is some i such that $1 \leq i \leq k$ and $X : \{Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k, R - XY_1 \cdots Y_k\} \in \Sigma^+_{\mathfrak{U}}$.* ∎

Another interpretation of Corollary 5 is the following: if $\mathfrak{U}$ is utilized to infer FDs, then the underlying universe does not need to be fixed at all; and if $\mathfrak{U}$ is utilized to infer FHDs, then the fixing of a universe can be deferred until the very last step of the inference.

Example 12 *Consider again Example 11 where $R_s = AD$ and $\Sigma = \{M : \{DF\}, MD : \{FL\}, MA \to D\}$. The inference of $MA \to D$ from Σ and R_s does not require us to fix any underlying relation schema, and for the inference of the FHD $M : \{A\}$ from Σ and R_s we fix the underlying relation schema to be $R = MDAFL$ in the very last step.* ∎

6 Reasoning about FDs and FHDs in undetermined universes

We have just seen that the system $\mathfrak{U}$ is almost complete for the R-implication of FDs and FHDs in the presence of an NFS. The notion of R-implication takes into account the underlying relation schema R over which the FDs, FHDs and NFSs are defined. We will show in this section that the system $\mathfrak{U}$ is actually complete for a notion of implication in which the underlying set of attributes remains undetermined. Consequently, the system $\mathfrak{U}$ allows inferences of exactly those data dependencies that are implied by a given set of FDs and FHDs in the presence of an NFS only.

FDs, FHDs and NFSs are syntactical expressions as before, but their attribute sets are finite subsets of our countably infinite set $\mathfrak{A}$. Let $Dom(r)$ denote the domain of a relation r, i.e., the set of attributes over which the relation is defined. A relation r is said to satisfy the FD $X \to Y$ if $XY \subseteq Dom(r)$ and for all tuples $t_1, t_2 \in r$ the following holds: if $X \subseteq ag^s(t_1, t_2)$, then $Y \subseteq ag(t_1, t_2)$. A relation r is said to satisfy the FHD $X : \{Y_1, \ldots, Y_k\}$ if $XY_1 \cdots Y_k \subseteq Dom(r)$ and $r = r_X[XY_1] \bowtie \cdots \bowtie r_X[XY_k] \bowtie r_X[X(R - XY_1 \cdots Y_k)]$ holds. Finally, a relation r is said to satisfy the NFS R_s if $R_s \subseteq Dom(r)$ and r is R_s-total.

Definition 7 *Let $\Sigma \cup \{\varphi\}$ be a finite set of FDs and FHDs and R_s an NFS. We say that* Σ implies φ *in the presence of R_s, denoted by $\Sigma \models_{R_s} \varphi$, if and only if every relation r satisfies the following condition: if $\cup_{\sigma \in \Sigma} Attr(\sigma) \cup Attr(\varphi) \cup R_s \subseteq Dom(r)$ and r satisfies Σ and the NFS R_s, then r 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 $\mathcal{R}_{\mathrm{F}}$, $\mathcal{A}_{\mathrm{F}}$, $\mathcal{T}_{\mathrm{F}}$, $\mathcal{D}_{\mathrm{F}}$, $\mathcal{U}_{\mathrm{F}}$, $\mathcal{A}_{\mathrm{H}}$, $\mathcal{T}_{\mathrm{H}}$, $\mathcal{O}_{\mathrm{H}}$, $\mathcal{S}_{\mathrm{H}}$, $\mathcal{M}_{\mathrm{H}}$, $\mathcal{D}_{\mathrm{H}}$, $\mathcal{I}_{\mathrm{H}}$, $\mathcal{I}_{\mathrm{FH}}$, $\mathcal{S}_{\mathrm{FH}}$, $\mathcal{T}_{\mathrm{FH}}$ are all sound inference rules (since they are R-sound for all R), the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$ and R-axiom are both R-sound but neither of them is sound. The following example illustrates this for the R-complementation rule $\mathcal{C}^R_{\mathrm{H}}$.

Example 13 *Let Σ consist of the single FHD* Title : {{Actor},{Feature}}*, and let R be the relation schema with the four attributes* Title, Actor, Feature, Language *and let $R_s = R$. Then* Title : $\{\{$Actor$\}, \{$Language$\}\}$ *is R-implied by Σ in the presence of R_s by the soundness of $\mathcal{C}^R_{\mathrm{H}}$. However,* Title : $\{\{$Actor$\}, \{$Language$\}\}$ *is not implied by Σ in the presence of R_s as the following counterexample relation r shows.*

Title	*Actor*	*Feature*	*Language*	*Crew*
Miyamoto Musashi	*T. Mifune*	*Trailer*	*English*	*H. Hinagaki*
Miyamoto Musashi	*T. Mifune*	*Trailer*	*Japanese*	*H. Hojo*

While $r = r$[Title Actor] $\bowtie$ *r*[Title Feature] $\bowtie$ *r*[Title Language Crew] *we have $r \neq r$*[Title Actor] $\bowtie$ *r*[Title Language] $\bowtie$ *r*[Title Feature Crew]. ∎

Let $\Sigma \cup \{\varphi\}$ be a set of FDs and FHDs, R_s an NFS, and let R be some relation schema such that $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \cup R_s \subseteq R$ holds. Based on Definitions 5 and 7 it follows that the implication of φ by Σ in the presence of R_s entails the R-implication of φ by Σ in the presence of R_s.

Lemma 8 *Let $\Sigma \cup \{\varphi\}$ be a finite set of FDs and FHDs, and R_s an NFS such that $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \cup R_s \subseteq R$ holds. Then $\Sigma \models^R_{R_s} \varphi$ whenever $\Sigma \models_{R_s} \varphi$.* ∎

The reverse direction of Lemma 8 also holds when φ is an FD, but Example 13 illustrates that the reverse direction does not hold when φ is an FHD.

Before we show that $\mathfrak{U}$ is a finite axiomatization for the implication of FDs and FHDs in the presence of an NFS, we prove two lemmata. The correctness of the first lemma can easily be observed by inspecting the inference rules in $\mathfrak{U}$. For each of the rules, the right-hand side of the conclusion does not contain any attribute that did not already occur in the right-hand side of at least one of the premises. Accordingly, by induction, this property is preserved by an inference of any length.

Lemma 9 *Let $\Sigma \cup \{\varphi\}$ be a finite set of FDs and FHDs, and R_s an NFS. If $\varphi \in \Sigma^+_{\mathfrak{U}}$, then $rhs(\varphi) \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma)$.* ∎

For the next lemma one may notice that attributes outside of $T := \cup_{\sigma\in\Sigma} Attr(\sigma)$ can always be introduced only in the last step of the inference by utilizing the *augmentation rules* $\mathcal{A}_{\mathrm{F}}$ and $\mathcal{A}_{\mathrm{H}}$, respectively.

Lemma 10 *Let $\Sigma \cup \{\varphi\}$ be a finite set of FDs and FHDs, and R_s an NFS. If $\varphi \in \Sigma^+_{\mathfrak{U}}$, then there is some inference $\gamma = [\psi_1, \ldots, \psi_l]$ of φ from Σ and R_s by $\mathfrak{U}$ such that*

$$Attr(\psi_i) \subseteq \cup_{\sigma\in\Sigma} Attr(\sigma)$$

holds for all $i = 1, \ldots, l-1$.

Proof For convenience let us define $T := \cup_{\sigma\in\Sigma} Attr(\sigma)$. Moreover, $\psi \cap T$ denotes the FHD $X \cap T : \{Y_1 \cap T, \ldots, Y_k \cap T\}$ if ψ denotes the FHD $X : \{Y_1, \ldots, Y_k\}$, and $\psi \cap T$ denotes the FD $X \cap T \to Y \cap T$ if ψ denotes the FD $X \to Y$.

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

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

We claim that ξ is an inference of $\varphi \cap T$ from Σ by $\mathfrak{U}$. 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 in

$$\mathfrak{U} = \{\mathcal{R}_{\rm F}, \mathcal{A}_{\rm F}, \mathcal{T}_{\rm F}, \mathcal{D}_{\rm F}, \mathcal{U}_{\rm F}, \mathcal{A}_{\rm H}, \mathcal{T}_{\rm H}, \mathcal{O}_{\rm H}, \mathcal{S}_{\rm H}, \mathcal{M}_{\rm H}, \mathcal{I}_{\rm FH}, \mathcal{S}_{\rm FH}\},$$

then $\psi_i \cap T$ is the result of applying the same rule to the corresponding premises in ξ.

According to Lemma 9 we have $rhs(\varphi) \subseteq \cup_{\sigma\in\Sigma} rhs(\sigma) \subseteq T$. We now distinguish between two cases. First, let φ denote the FHD $X : \{Y_1, \ldots, Y_k\}$. It follows that $Y_1 \cdots Y_k \subseteq T$ holds, in particular $Y_i \subseteq T$ for all $i = 1, \ldots, k$. Consequently, $Y_i \cap T = Y_i$ for $i = 1, \ldots, k$ and this implies that we can infer φ from $X \cap T : \{Y_1 \cap T, \ldots, Y_k \cap T\}$ by a single application of the augmentation rule $\mathcal{A}_{\rm H}$:

$$\frac{X \cap T : \{Y_1 \cap T, \ldots, Y_k \cap T\}}{\underbrace{(X \cap T) \cup (X - T)}_{=X} : \{Y_1 - (X - T), \ldots, \underbrace{Y_i - (X - T)}_{=Y_i}, \ldots, Y_k - (X - T)\}}.$$

Note that $Y_i - (X - T) = Y_i$ since Y_i and X are disjoint. Hence, the inference $[\xi, X : \{Y_1, \ldots, Y_k\}]$ has the desired properties. It remains to consider the case where φ denotes the FD $X \to Y$. Given that $Y \subseteq T$ we can infer φ from $X \cap T \to Y \cap T$ by a single application of the augmentation rule $\mathcal{A}_{\rm F}$:

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

Note that $Y - (X - T) = Y$ since X and Y are disjoint. Hence, the inference $[\xi, X \to Y]$ has the desired properties.

Theorem 6 *The set $\mathfrak{U}$ of inference rules is a finite axiomatization for the implication of FDs and FHDs in the presence of an NFS.*

Proof For the soundness of $\mathfrak{U}$ we need to show that every $\varphi \in \Sigma^+_{\mathfrak{U}}$ is implied by Σ in the presence of R_s. That is, every relation r that satisfies $T := \cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \cup R_s \subseteq Dom(r)$, $\models_r \sigma$ for all $\sigma \in \Sigma$ and r is R_s-total also satisfies $\models_r \varphi$. According to Lemma 10 there is an inference γ of φ from Σ by $\mathfrak{U}$ such that $Attr(\psi) \subseteq T \subseteq Dom(r)$ holds for every ψ occurring in γ. Since each rule of $\mathfrak{U}$ 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 $\mathfrak{U}$ we assume that $\varphi \notin \Sigma^+_{\mathfrak{U}}$. Let $R \subseteq \mathfrak{A}$ be a finite set of attributes such that T is a proper subset of R, i.e., $T \subset R$.

If φ denotes an FD, then Corollary 5 shows that $\varphi \notin \Sigma^+_{\mathfrak{W}_\mathcal{C}}$. However, $\mathfrak{W}_\mathcal{C}$ is R-complete for the R-implication of FDs and FHDs in the presence of R_s. Hence, it follows that Σ does not R-imply φ in the presence of R_s. Consequently, Σ does not imply φ in the presence of R_s by Lemma 8.

If φ denotes the FHD $X : \{Y_1, \ldots, Y_k\}$, then Lemma 9 shows that $X : \{Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k, R - XY_1 \cdots Y_k\} \notin \Sigma^+_{\mathfrak{U}}$ for all $i = 1, \ldots, k$ since $R - XY_1 \cdots Y_k$ is not a subset of T. From $X : \{Y_1, \ldots, Y_k\} \notin \Sigma^+_{\mathfrak{U}}$ and $X : \{Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k, R - XY_1 \cdots Y_k\} \notin \Sigma^+_{\mathfrak{U}}$ for all $i = 1, \ldots, k$ we conclude that $X : \{Y_1, \ldots, Y_k\} \notin \Sigma^+_{\mathfrak{W}_\mathcal{C}}$ by Corollary 5. However, $\mathfrak{W}_\mathcal{C}$ is R-complete for the R-implication of FDs and FHDs in the presence of R_s. Hence, it follows that Σ does not R-imply φ in the presence of R_s. Consequently, Σ does not imply φ in the presence of R_s by Lemma 8.

Example 14 *Consider again Example 11 where $R = MDAFL$, $R_s = AD$, and $\Sigma = \{M : \{DF\}, MD : \{FL\}, MA \to D\}$. Since $A \notin DFL$ it follows from Lemma 9 that $M : \{A\} \notin \Sigma^+_{\mathfrak{U}}$. According to Theorem 6 we conclude that $\Sigma \not\models_{R_s} M : \{A\}$. Recall that $\Sigma \models^R_{R_s} M : \{A\}$, cf. Example 11.* ∎

7 Value unknown at present

Codd's original proposal [24] to handle incomplete information suggested the addition to the database domains of an unmarked null value `unk`, whose meaning is "value unknown at present". Following Codd's proposal, incomplete information is represented in SQL by using `unk` as a distinguished null value [27]. We will discuss in this section how the results from the previous sections carry over to this approach towards handling incomplete information.

Levene and Loizou introduced and axiomatized strong and weak FDs (WFDs) with respect to a possible world semantics [68]. We will start by summarizing their approach towards defining WFDs. For this purpose, we assume that the domains of all attributes contain the distinguished value `unk` (and no longer the distinguished value `ni`). With this change in mind, we re-apply the definitions of an X-total tuple and relation as before. The set of all *possible worlds* relative to a relation r over R, denoted by $Poss(r)$, is defined by

$$\begin{aligned} Poss(r) \; := \; & \{s \mid s \text{ is a relation over } R \text{ and there is a total and onto mapping} \\ & f : r \to s \text{ such that } \forall t \in r, t \text{ is subsumed by } f(t) \text{ and } f(t) \text{ is } R\text{-total}\}. \end{aligned}$$

This definition of possible worlds embodies the *closed world assumption* (CWA) [60, 81], since $Poss(r)$ allows only R-total tuples from the relation r to be present in $Poss(r)$.

A *weak functional dependency* (WFD) over a relation schema R is a statement of the form $\Diamond(X \to Y)$, where $XY \subseteq R$ and $X \cap Y = \emptyset$. A relation r over R is said to *satisfy* the WFD $\Diamond(X \to Y)$ over R, if there is some $p \in Poss(r)$ such that for all $t_1, t_2 \in p$, if $t_1[X] = t_2[X]$, then $t_1[Y] = t_2[Y]$. We note that the definition of satisfaction of a WFD in a relation reduces to the standard definition of the satisfaction of an FD when the relation is R-total (in this case there is exactly one $p \in Poss(r)$ and $\forall p \in Poss(r)$ is equivalent to $\exists p \in Poss(r)$). We observe that $\Diamond$ can be viewed as representing the modal operator *possibly* of a normal system of propositional modal logic [25]. Finally we remark that the weak approach to satisfaction of an FD by a partial relation allows a higher degree of uncertainty to be represented in the database than the strong approach (where an FD must be satisfied in all possible worlds) [68]. The disadvantage of the weak over the strong approach is that strongly satisfied FDs are easier to maintain [68]. Hence, both approaches complement one another.

It is known that WFDs in the absence of an NFS enjoy the same axiomatization as "no information" FDs (NFDs) [6, 70]. However, WFDs are different from NFDs. First of all, WFDs are defined with respect to Codd's null value `unk`. Under this interpretation we know that a value exists, whereas under the "no information" interpretation it may also be the case that no value exists at all. Moreover, WFDs and NFDs also behave differently. For example, the relation r over $R = MDA$ with the two tuples (The Seven Samurai, A. Kurosawa, T. Mifune) and (The Seven Samurai, `unk`, T. Shimura) satisfies the WFD $\Diamond(M \to D)$. However, the NFD $M \to D$ is violated by the relation consisting of (The Seven Samurai, A. Kurosawa, T. Mifune) and (The Seven Samurai, `ni`, T. Shimura). That is, we have two distinct tuples which have an information on the attribute M and the information is the same, but the first tuple has some information for D while the second tuple has "no information" for D.

In the context of NFDs we defined the weak agree set of two tuples as $ag^w(t_1, t_2) = \{A \in R \mid t_1(A) = \texttt{ni} = t_2(A)\}$. For WFDs we re-define this to be $ag^w(t_1, t_2) := \{A \in R \mid t_1(A) = \texttt{unk} \text{ or } t_2(A) = \texttt{unk}\}$. Intuitively, this makes perfect sense in this context: two tuples weakly agree on an attribute if there is a possible world on which they agree on A. The definition of a strong agree set $ag^s(t_1, t_2) := \{A \in R \mid t_1(A) = t_2(A) \text{ and } t_1(A) \neq \texttt{unk} \neq t_2(A)\}$ requires no adjustment apart from the notation of the null value, and $ag(t_1, t_2) := ag^s(t_1, t_2) \cup ag^w(t_1, t_2)$ as before. The next proposition, which gives a syntactic characterization of satisfaction of a WFD, follows from the definition of satisfaction.

Proposition 1 *Let r be a relation over relation schema R. Then r satisfies the WFD $\Diamond(X \to Y)$ over R if and only if for all $t_1, t_2 \in r$, if $X \subseteq ag^s(t_1, t_2)$, then $Y \subseteq ag(t_1, t_2)$.* ∎

A *weak full hierarchical dependency* (WFHD) over R is a statement $\Diamond(X : S)$, where $X \subseteq R$ and S is a set of mutually disjoint non-empty subsets of R that are also disjoint from X, i.e., for all $Y \in S$ we have $\emptyset \neq Y \subseteq R$ and for all $Y, Z \in S \cup \{X\}$ we have $Y \cap Z = \emptyset$. A relation r over R is said to *satisfy* the WFHD $\Diamond(X : \{Y_1, \ldots, Y_k\})$ over R, if there is some $p \in Poss(r)$ such that for all $t_1, \ldots, t_{k+1} \in p$ 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 $t \in p$ such that $t[XY_i] = t_i[XY_i]$ for all $i = 1, \ldots, k$ and $t[X(R - XY_1 \cdots Y_k)] = t_{k+1}[X(R - XY_1 \cdots Y_k)]$.

Similar to the case of functional dependencies, WFHDs behave quite differently from FHDs in the "no information" context. For example, the following relation r over $R = ASLC$:

Movie	Director	Actor	Feature
The Seven Samurai	A. Kurosawa	T. Mifune	Deleted Scene
Rashomon	A. Kurosawa	T. Shimura	Subtitle
Rashomon	`unk`	T. Mifune	Subtitle
The Seven Samurai	`unk`	T. Shimura	Deleted Scene

satisfies the WFHD $\Diamond(D : \{\{A\}\})$. However, under the "no information" interpretation the FHD $D : \{\{A\}\}$ is violated by the following relation.

Movie	Director	Actor	Feature
The Seven Samurai	A. Kurosawa	T. Mifune	Deleted Scene
Rashomon	A. Kurosawa	T. Shimura	Subtitle
Rashomon	`ni`	T. Mifune	Subtitle
The Seven Samurai	`ni`	T. Shimura	Deleted Scene

For WFHDs we obtain the following syntactic characterization for their satisfaction by a partial relation.

Proposition 2 *Let r be a relation over relation schema R. Then r satisfies the WFHD $\Diamond(X : \{Y_1, \ldots, Y_k\})$ over R if and only if for all $t_1, \ldots, t_{k+1} \in r$ the following condition is satisfied: if $X \subseteq ag^s(t_i, t_j)$ for all $1 \leq i, j \leq k+1$, then there is some $t \in r$ such that $X \subseteq ag^s(t, t_i)$ for $i = 1, \ldots, k+1$, $Y_i \subseteq ag(t, t_i)$ for all $i = 1, \ldots, k$, and $R - XY_1 \cdots Y_k \subseteq ag(t, t_{k+1})$.* ∎

For an inference system $\mathfrak{S}$ for FDs and FHDs, let $\mathfrak{S}'$ denote the set of inference rules obtained from replacing the FDs and FHDs in $\mathfrak{S}$ by WFDs and WFHDs, respectively. Using Propositions 1 and 2 it is not difficult to show that the inference rules of $\mathfrak{W}'$ are R-sound for the R-implication of WFDs and WFHDs in the presence of an NFS. Following the same line of arguments as in Section 4 it can be shown that the system $\mathfrak{W}'$ forms a finite axiomatization for the R-implication of the combined class of WFDs and WFHDs in the presence of an NFS. In particular, the two-tuple relation r_φ

	$X(X^+_\Sigma \cap R_s)$	$(X^+_\Sigma - X) - R_s$	$W_1 \cap R_s$	$W_1 - R_s$	.	W_i	.	$W_k \cap R_s$	$W_k - R_s$
t_1	$0 \cdots 0$	$\texttt{unk} \cdots \texttt{unk}$	$0 \cdots 0$	$\texttt{unk} \cdots \texttt{unk}$		$0 \cdots 0$		$0 \cdots 0$	$\texttt{unk} \cdots \texttt{unk}$
t_2	$0 \cdots 0$	$0 \cdots 0$	$0 \cdots 0$	$0 \cdots 0$		$1 \cdots 1$		$0 \cdots 0$	$0 \cdots 0$

shows that a WFD or WFHD φ is not R-implied by a set Σ of WFDs and WFHDs in the presence of the NFS R_s whenever φ cannot be inferred from Σ by $\mathfrak{W}'$ in the context of WFDs and WFHDs, cf. the proof of Theorem 3. The major results of our article, i.e. Theorems 5 and 6, carry over to WFDs and WFHDs. We summarize these results in the following theorem.

Theorem 7 *The following hold:*

1. *For all relation schemata R, the sets $\mathfrak{W}'$ and $\mathfrak{W}'_0$ form finite axiomatizations for the R-implication of WFDs and WFHDs in the presence of an NFS.*

2. *For all relation schemata R, the set $\mathfrak{W}'_C$ forms a finite appropriate axiomatization for the R-implication of WFDs and WFHDs in the presence of an NFS.*

3. *The set $\mathfrak{U}'$ forms a finite axiomatization for the implication of WFDs and WFHDs in the presence of an NFS.* ∎

8 Conclusion and Future Work

We have investigated implication problems for expressive classes of data dependencies over partial relations. For a database administrator to have full control over the degree of partiality in relations we studied the implication problems in the presence of a null-free subschema. The null-free subschema amounts to the set of attributes that are declared `NOT NULL` in SQL table definitions [27]. We have established the first axiomatization for the R-implication of FDs and FHDs in the presence of an NFS. Moreover, we have extended previous research on the appropriateness of inference systems for the R-implication of FDs and MVDs over total relations. That is, we have established an appropriate axiomatization for the R-implication of FDs and FHDs in the presence of an NFS. Our axiomatization is appropriate in the following sense: 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. This result demonstrates that the complementation rule is a mere means for achieving database normalization. Furthermore, we have established an axiomatization for the implication of FDs and FHDs in the presence of an NFS where the underlying relation schema is left undetermined. This unburdens the theory of the strong assumption that the complete set of attributes is already known to the database designers before they can start to think about the data dependencies that are meaningful to the relation schema.

We conclude this article by listing some related problems that may be of interest for future research. The (mixed) *subset rule* plays a key role in achieving complementarity and adequacy of inference systems. It would be interesting to see whether there are any axiomatizations that do not feature either or both of the subset rules.

Levene and Loizou [68] have established an axiomatization for the combined class of strong and weak functional dependencies. It would be interesting to study whether this axiomatization can be extended to cover subclasses of strong and weak full hierarchical dependencies as well, both in the absence or presence of a null-free subschema.

Embedded multivalued dependencies are multivalued dependencies that hold in the projection of a relation. It has been shown that the implication problem for embedded multivalued dependencies is not finitely axiomatizable by a Hilbert-style axiomatization [86, 82]. Moreover, the implication and finite implication problems for the class of embedded multivalued dependencies are both undecidable [57, 58]. Full hierarchical dependencies are equivalent to multivalued dependencies, and hierarchical dependencies, called first-order hierarchical decomposition in [31], are equivalent to embedded multivalued dependencies. Note that in the case of implication over undetermined universes, multivalued dependencies are defined over any *full* set of attributes that includes those occurring in the dependencies. This is in contrast to embedded multivalued dependencies which are defined over projections of the full set of attributes. This difference is the deciding factor for the (non-)axiomatizability of (embedded) multivalued dependencies. It is an interesting problem for future work to investigate the properties of embedded multivalued dependencies over partial relations, with respect to different approaches to partiality. The class of so-called *conflict-free* embedded multivalued dependencies is of particular interest as it enjoys a finite axiomatization [79].

There are equivalences between the logical R-implication of classes of relational dependencies and classes of conditional independencies in Bayesian networks [62, 98]. 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.

A very interesting treatment of MVDs and FHDs in the context of Entity-Relationship modeling can be found in [90]. 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.

It would be a rewarding exercise to provide the foundations for extending design aids available for total relations [29, 30, 76, 84]. It seems intuitive that design teams find it more difficult to understand the interaction of FDs and FHDs in the presence of an NFS. Hence, Armstrong databases [35] might be of even bigger value than reported for the case of total relations [67]. Computational and structural properties of Armstrong tables for FDs in the presence of an NFS have recently been investigated [50].

Finally, we mention that the class of full hierarchical dependencies has largely been unexplored for XML, except for [83, 94]. This is somewhat surprising since the body of research on functional dependencies over XML data is rather substantial, and full hierarchical dependencies aim to explore the lossless decompositions of documents in which they are exhibited.

9 Acknowledgement

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

References

[1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases.* Addison-Wesley, Boston, MA, USA, 1995.

[2] M. Arenas, P. Barcelo, L. Libkin, and F. Murlak. *Relational and XML data exchange.* Morgan & Claypool Publishers, 2010.

[3] M. Arenas, W. Fan, and L. Libkin. On the complexity of verifying consistency of XML specifications. *SIAM J. Comput.*, 38(3):841–880, 2008.

[4] M. Arenas and L. Libkin. A normal form for XML documents. *ACM Trans. Database Syst.*, 29(1):195–232, 2004.

[5] W. W. Armstrong. Dependency structures of database relationships. *Information Processing*, 74:580–583, 1974.

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

[7] C. Beeri. On the membership problem for functional and multivalued dependencies in relational databases. *ACM Trans. Database Syst.*, 5(3):241–259, 1980.

[8] C. Beeri and P. Bernstein. Computational problems related to the design of normal form relational schemas. *ACM Trans. Database Syst.*, 4(1):30–59, 1979.

[9] C. Beeri, R. Fagin, and J. H. Howard. A complete axiomatization for functional and multivalued dependencies in database relations. In *Proceedings of the SIGMOD International Conference on Management of Data*, pages 47–61, Toronto, Canada, 1977. ACM.

[10] C. Beeri, R. Fagin, D. Maier, and M. Yannakakis. On the desirability of acyclic database schemes. *J. ACM*, 30(3):479–513, 1983.

[11] C. Beeri and M. Vardi. Formal systems for tuple and equality generating dependencies. *SIAM J. Comput.*, 13(1):76–98, 1984.

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

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

[14] J. Biskup. Achievement of relational database schema design theory revisited. In *Semantic in databases*, volume 1358 of *Lecture Notes in Computer Science*, pages 29–54, Prague, Czech Republic, 1995. Springer.

[15] J. Biskup, U. Dayal, and P. Bernstein. Synthesizing independent database schemas. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 143–151, Boston, USA, May 30 - June 1 1979. ACM.

[16] J. Biskup, D. Embley, and J. Lochner. Reducing inference control to access control for normalized database schemas. *Inf. Proc. Letters*, 106(1):8–12, 2008.

[17] J. Biskup and S. Link. Appropriate inferences of data dependencies in relational databases. *Ann. Math. Artif. Intell.*, 63(3-4):213–255, 2012.

[18] M. Bojanczyk, A. Muscholl, T. Schwentick, and L. Segoufin. Two-variable logic on data trees and XML reasoning. *J. ACM*, 56(3):Article 13, 2009.

[19] A. Cali, D. Calvanese, G. De Giacomo, and M. Lenzerini. Data integration under integrity constraints. *Inf. Syst.*, 29(2):147–163, 2004.

[20] A. Chandra, H. Lewis, and J. Makowsky. Embedded implicational dependencies and their inference problem. In *Proceedings of the 13th Annual ACM Symposium on Theory of Computing (STOC)*, pages 342–354, Milwaukee, USA, May 11-13 1981. ACM.

[21] J. Chomicki. Consistent query answering: Five easy pieces. In *Proceedings of the 11th International Conference on Database Theory (ICDT)*, volume 4353 of *Lecture Notes in Computer Science*, pages 1–17, Barcelona, Spain, January 10-12 2007. Springer.

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

[23] E. F. Codd. Further normalization of the database relational model. In *Courant Computer Science Symposia 6: Data Base Systems*, pages 33–64, 1972.

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

[25] M. Cresswell and G. Hughes. *A new introduction to modal logic*. Routledge, London and New York, 1996.

[26] N. Dalvi, C. Re, and D. Suciu. Probabilistic databases: diamonds in the dirt. *Commun. ACM*, 52(7):86–94, 2009.

[27] C. Date and H. Darwen. *A guide to the SQL standard*. Addison-Wesley Professional, Reading, MA, USA, 1997.

[28] S. Davidson, W. Fan, and C. Hara. Propagating XML constraints to relations. *J. Comput. Syst. Sci.*, 73(3):316–361, 2007.

[29] F. De Marchi, S. Lopes, J.-M. Petit, and F. Toumani. Analysis of existing databases at the logical level: the DBA companion project. *SIGMOD Record*, 32(1):47–52, 2003.

[30] F. De Marchi and J.-M. Petit. Semantic sampling of existing databases through informative Armstrong databases. *Inf. Syst.*, 32(3):446–457, 2007.

[31] C. Delobel. Normalization and hierarchical dependencies in the relational data model. *ACM Trans. Database Syst.*, 3(3):201–222, 1978.

[32] C. Delobel and M. Adiba. *Relational database systems*. Elsevier North Holland, New York, NY, USA, 1985.

[33] A. Deutsch, B. Ludäscher, and A. Nash. Rewriting queries using views with access patterns under integrity constraints. *Theor. Comput. Sci.*, 371(3):200–226, 2007.

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

[35] R. Fagin. Armstrong databases. Technical Report RJ3440(40926), IBM Research Laboratory, San Jose, California, USA, 1982.

[36] R. Fagin. Horn clauses and database dependencies. *J. ACM*, 29(4):952–985, 1982.

[37] R. Fagin, P. Kolaitis, R. Miller, and L. Popa. Data exchange: semantics and query answering. *Theor. Comput. Sci.*, 336(1):89–124, 2005.

[38] R. Fagin, P. Kolaitis, L. Popa, and W. Tan. Reverse data exchange: coping with nulls. In *Proceedings of the Twenty-Eigth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, pages 23–32, Providence, USA, June 19 - July 1 2009. ACM.

[39] R. Fagin and M. Vardi. The theory of data dependencies - a survey. In *Mathematics of Information Processing*, volume 34 of *Proceedings of Symposia in Applied Mathematics*, pages 19–72, Louisville, USA, 1986. American Mathematical Society.

[40] W. Fan, F. Geerts, X. Jia, and A. Kementsietsidis. Conditional functional dependencies for capturing data inconsistencies. *ACM Trans. Database Syst.*, 33(2):Article 6, 2008.

[41] F. Ferrarotti, S. Hartmann, and S. Link. On the role of the complementation rule for data dependencies over incomplete relations. In *Proceedings of the Seventeenth International Workshop on Logic, Language, Information and Computation (WoLLIC)*, volume 6188 of *Lecture Notes in Artificial Intelligence*, pages 136–147, Brasilia, Brazil, 2010. Springer.

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

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

[44] G. Gottlob, R. Pichler, and F. Wei. Tractable database design and datalog abduction through bounded treewidth. *Inf. Syst.*, 35(3):278–298, 2010.

[45] G. Gottlob and R. Zicari. Closed world databases opened through null values. In *Proceedings of the Fourteenth International Conference on Very Large Databases (VLDB)*, pages 50–61, Los Angeles, USA, August 29 - September 1 1988. IEEE Computer Society.

[46] G. Graetzer. *General lattice theory.* Birkhäuser, Boston, USA, 1998.

[47] G. Grahne. Dependency satisfaction in databases with incomplete information. In *Proceedings of the Tenth International Conference on Very Large Databases (VLDB)*, pages 37–45, Singapore, August 27-31 1984. IEEE Computer Society.

[48] J. Grant. Null values in a relational data base. *Inf. Process. Lett.*, 6(5):156–157, 1977.

[49] C. Hara and S. Davidson. Reasoning about nested functional dependencies. In *Proceedings of the Eighteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, pages 91–100, Philadelphia, U.S.A., May 31 - June 2 1999. ACM Press.

[50] S. Hartmann, M. Kirchberg, and S. Link. Design by example for SQL table definitions with functional dependencies. *The VLDB Journal*, 21(1):121–144, 2012.

[51] S. Hartmann and S. Link. Characterising nested database dependencies by fragments of propositional logic. *Ann. Pure Appl. Logic*, 152(1-3):84–106, 2008.

[52] S. Hartmann and S. Link. Efficient reasoning about a robust XML key fragment. *ACM Trans. Database Syst.*, 34(2):Article 10, 2009.

[53] S. Hartmann and S. Link. Numerical constraints on XML data. *Inf. Comp.*, 208(5):521–544, 2010.

[54] S. Hartmann and S. Link. When data dependencies over SQL tables meet the Logics of Paradox and S-3. In *Proceedings of the 29th ACM SIGMOD-SIGART-SIGACT Symposium on Principles of Database Systems (PoDS)*, pages 317–326, Indianapolis, USA, 2010. ACM.

[55] S. Hartmann and S. Link. The implication problem of data dependencies over SQL table definitions: axiomatic, algorithmic and logical characterizations. *ACM Trans. Database Syst.*, 37(2):Article 14, 2012.

[56] S. Hartmann, S. Link, and K.-D. Schewe. Axiomatizations of functional dependencies in the presence of records, lists, sets and multisets. *Theor. Comput. Sci.*, 355(2):167–196, 2006.

[57] C. Herrmann. On the undecidability of implications between embedded multivalued database dependencies. *Inf. Comput.*, 122(2):221–235, 1995.

[58] C. Herrmann. Corrigendum to "on the undecidability of implications between embedded multivalued database dependencies". *Inf. Comput.*, 204(12):1847–1851, 2006.

[59] T. Imielinski. Incomplete information in logical databases. *IEEE Data Eng. Bull.*, 12(2):29–40, 1989.

[60] T. Imielinski and W. Lipski Jr. Incomplete information in relational databases. *J. ACM*, 31(4):761–791, 1984.

[61] C. Jensen, R. Snodgrass, and M. Soo. Extending existing dependency theory to temporal databases. *IEEE Trans. Knowl. Data Eng.*, 8(4):563–582, 1996.

[62] M. Karny and T. Kroupa. Axiomatisation of fully probabilistic design. *Inf. Sci.*, 186(1):105 – 113, 2012.

[63] A. Klug and R. Price. Determining view dependencies using tableaux. *ACM Trans. Database Syst.*, 7(3):361–380, 1982.

[64] H. Koehler, S. Hartmann, and S. Link. Full hierarchical dependencies in fixed and undetermined universes. *Ann. Math. Artif. Intell.*, 50(1-2):195–226, 2007.

[65] H. Koehler and S. Link. Armstrong axioms and Boyce-Codd-Heath normal form under bag semantics. *Inf. Process. Lett.*, 110(16):717–724, 2010.

[66] S. Kolahi. Dependency-preserving normalization of relational and XML data. *J. Comput. Syst. Sci.*, 73(4):636–647, 2007.

[67] W. Langeveldt and S. Link. Empirical evidence for the usefulness of Armstrong relations on the acquisition of meaningful functional dependencies. *Inf. Syst.*, 35(3):352–374, 2010.

[68] M. Levene and G. Loizou. Axiomatisation of functional dependencies in incomplete relations. *Theor. Comput. Sci.*, 206(1-2):283–300, 1998.

[69] M. Levene and G. Loizou. Database design for incomplete relations. *ACM Trans. Database Syst.*, 24(1):80–125, 1999.

[70] E. Lien. On the equivalence of database models. *J. ACM*, 29(2):333–362, 1982.

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

[72] S. Link. On the implication of multivalued dependencies in partial database relations. *Int. J. Found. Comput. Sci.*, 19(3):691–715, 2008.

[73] S. Link. Characterizations of multivalued dependency implication over undetermined universes. *J. Comput. Syst. Sci.*, 78(4):1026–1044, 2012.

[74] A. Makinouchi. A consideration on normal form of not-necessarily-normalized relation in the relational data model. In *Proceedings of the Third International Conference on Very Large Databases (VLDB)*, pages 447–453, Tokyo, Japan, October 6-8 1977. IEEE Computer Society.

[75] J. Makowsky and M. Vardi. On the expressive power of data dependencies. *Acta Inf.*, 23(3):231–244, 1986.

[76] H. Mannila and K.-J. Räihä. Design by example: An application of Armstrong relations. *J. Comput. Syst. Sci.*, 33(2):126–141, 1986.

[77] H. Mannila and K.-J. Räihä. Algorithms for inferring functional dependencies from relations. *Data Knowl. Eng.*, 12(1):83–99, 1994.

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

[79] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Franciso, U.S.A., 1988.

[80] S. Y. Petrov. Finite axiomatization of languages for representation of system properties: Axiomatization of dependencies. *Inf. Sci.*, 47:339–372, 1989.

[81] R. Reiter. On closed world data bases. In *Logic and Data Bases*, pages 119–140, New York, USA, 1978. Plenum Press.

[82] Y. Sagiv and S. Walecka. Subset dependencies and a completeness result for a subclass of embedded multivalued dependencies. *J. ACM*, 29(1):103–117, 1982.

[83] L. Saxton and X. Tang. Tree multivalued dependencies for XML datasets. In *Proceedings of the Fifth International Conference on Advances in Web-Age Information Management (WAIM)*, volume 3129 of *Lecture Notes in Computer Science*, pages 357–367, Dalian, China, July 15-17 2004. Springer.

[84] A. Silva and M. Melkanoff. A method for helping discover the dependencies of a relation. In *Proceedings of the Workshop on Formal Bases for Data Bases - Advances in Data Base Theory*, pages 115–133, Toulouse, France, December 12-14 1979. Plemum Press.

[85] M. Sözat and A. Yazici. A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations. *ACM Fuzzy Sets and Systems*, 117(2):161–181, 2001.

[86] D. Stott Parker Jr. and K. Parsaye-Ghomi. Inferences involving embedded multivalued dependencies and transitive dependencies. In *Proceedings of the 1980 ACM SIGMOD International Conference on Management of Data*, pages 52–57, Santa Monica, U.S.A., may 14-16 1980. ACM Press.

[87] Z. Tan and L. Zhang. Repairing XML functional dependency violations. *Inf. Sci.*, 181(23):5304–5320, 2011.

[88] Z. Tari, J. Stokes, and S. Spaccapietra. Object normal forms and dependency constraints for object-oriented schemata. *ACM Trans. Database Syst.*, 22:513–569, 1997.

[89] B. Thalheim. *Dependencies in relational databases*. Teubner, 1991.

[90] B. Thalheim. Conceptual treatment of multivalued dependencies. In *Proceedings of the 22nd International Conference on Conceptual Modeling (ER)*, volume 2813 of *Lecture Notes in Computer Science*, pages 363–375, Chicago, USA, October, 13-16 2003. Springer.

[91] D. Toman and G. Weddell. On keys and functional dependencies as first-class citizens in description logics. *J. Autom. Reasoning*, 40(2-3):117–132, 2008.

[92] K. Vadaparty and S. Naqvi. Using constraints for efficient query processing in nondeterministic databases. *IEEE Trans. Knowl. Data Eng.*, 7(6):850–864, 1995.

[93] M. Vincent. Semantic foundations of 4NF in relational database design. *Acta Inf.*, 36(3):173–213, 1999.

[94] M. Vincent and J. Liu. Multivalued dependencies and a 4NF for XML. In *Proceedings of the 15th International Conference on Advanced Information Systems Engineering (CaISE)*, volume 2681 of *Lecture Notes in Computer Science*, pages 14–29, Klagenfurt, Austria, June 16-18 2003. Springer.

[95] M. Vincent, J. Liu, and C. Liu. Strong functional dependencies and their application to normal forms in XML. *ACM Trans. Database Syst.*, 29(3):445–462, 2004.

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

[97] J. Wijsen. Temporal FDs on complex objects. *ACM Trans. Database Syst.*, 24(1):127–176, 1999.

[98] S. Wong, C. Butz, and D. Wu. On the implication problem for probabilistic conditional independency. *Trans. Systems, Man, and Cybernetics, Part A: Systems and Humans*, 30(6):785–805, 2000.

[99] M. Wu. The practical need for fourth normal form. In *Proceedings of the Twenty-third ACM SIGCSE Technical Symposium on Computer Science Education*, pages 19–23, Kansas City, USA, 1992. ACM.

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

[101] W. Ziarko. The discovery, analysis, and representation of data dependencies in databases. In *Knowledge Discovery in Databases*, pages 195–212, Cambridge, USA, 1991. MIT Press.