Semantic Considerations on Multivalued Dependencies in Relational Databases

FUMIO NAKAMURA* and PETER P. CHEN**

Data dependencies play an essential role in database design. They provide not only integrity constraints but also structural information about data. Among them, multivalued dependencies, as well as functional dependencies, are fundamental. However, the semantic aspects of multivalued dependencies have not been sufficiently investigated. This paper examines multivalued dependencies from the semantic point of view and clarifies the following: (1) The difference between Boyce-Codd normal form and fourth normal form is clarified. (2) It is shown that the real transitivity condition for multivalued dependencies does not hold in a natural sense. (3) It is also shown that a rule which is a mixture of functional and multivalued dependencies has the same problem as the transitivity rule.

1. Introduction

In the design and analysis of relational databases, data dependencies and database normalization theory have been recognized as central ever since Codd introduced functional dependencies and applied them to the normalization of relations [6, 7]. Based on the concept of functional dependencies, Codd defined three normal forms for relations to eliminate undesirable behavior in data manipulation.

Functional dependencies are a kind of relationships between data elements (i.e., attributes) or sets of data elements. A functional dependency $X \rightarrow Y$, where X and Y are sets of attributes, hold whenever the relationship between X and Y is one-to-one or many-to-one. In other words, functional dependencies can completely represent one-to-one and many-to-one relationships between attributes. This completeness and the close correspondence to the single-valued function $y = f(x)$ have made functional dependencies easy to understand.

As suggested by Schmid and Swenson [19], there may be other kinds of relationships which are not functional. Fagin [10] and, independently, Zaniolo [20] have presented a new type of dependencies called, multivalued dependencies. (Delobel has also presented a similar idea which he called hierarchical dependencies [9].) Based on multivalued dependencies, Fagin defined fourth normal form.

Recently, more work has been done on multivalued dependencies [3, 11, 12, 13, 14, 15, 17]. It seems that those efforts have been dealing more and more syntactic aspects of multivalued dependencies, and the semantic considerations have not been sufficiently examined (an exception is [12]).

*Systems Development Laboratory Hitachi, Ltd., 5030 Totsuka-machi, Totsuka-ku, Yokohama 244, Japan.
**Graduate School of Management, University of California, Los Angeles, California 90024 U.S.A.

The purpose of this paper is to discuss the semantic aspects of multivalued dependencies, to clarify the characteristics of relations which include multivalued dependencies, and to review some known results on multivalued dependencies. The motivation of our work is as follows:

1. Multivalued dependencies, as well as functional dependencies, have an important role in the database theory, since both of them relate to the fundamental properties of relationships between data. Basic types of relationships between X and Y, where X and Y are sets of attributes, are as follows:
   (1) One-to-one correspondence.
   (2) Many-to-one correspondence.
   (3) One-to-many correspondence.
   (4) Many-to-many correspondence.

   The functional dependency $X \rightarrow Y$ holds for any X and Y which belong to type (1) or type (2). Multivalued dependency can represent some of type (3) and type (4), which can not be represented by functional dependencies. However, multivalued dependencies can not cover all the relationships which belong to type (3) or type (4). This makes multivalued dependencies more difficult to understand and treat than functional dependencies.
2. As mentioned before, semantic aspects of multivalued dependencies have not been sufficiently investigated. It should be noticed, however, that semantic considerations are essential for better understanding of data and for fruitful contribution to the problems of data base design and manipulation.
3. The data dependency theory is in the midst of development, and an effort should be made on building a better representation of data dependencies. The first step to accomplish this should be clarifying the powers and limits of functional and multivalued dependencies. Compared with functional dependencies, the properties of multivalued dependencies have not been well examined.

Table 1 A Relation in BCNF but not in 4NF.

Employee	Child	Year	Salary
1	Hans	70	10,000
1	Dick	70	10,000
1	Hans	72	12,000
1	Dick	72	12,000
1	Hans	73	15,000
1	Dick	73	15,000
2	Eva	70	20,000

The main topics discussed later are: (1) the difference between Boyce-Codd normal form and fourth normal form and (2) transitivity of multivalued dependencies. Let us examine the relation R (EMPLOYEE, CHILD, YEAR, SALARY) in Table 1 (which is introduced in [19] and cited by [10]). R is in Boyce-Codd normal form but not in fourth normal form. (The combination of YEAR and SALARY makes up salary history for employees.) The database designer may ask the following fundamental questions: What does R represent? Is it possible to add non-key attributes to R in order to explain R in more detail? In a later section, we shall demonstrate that, in a natural sense, it is only all-key relations that are in Boyce-Codd normal form but not in fourth normal form.

The transitivity rule in multivalued dependencies states that if $X \to\to Y$ and $Y \to\to Z$ hold, then $X \to\to Z$ also holds. However, there is a problem about the condition $X \to\to Y$ and $Y \to\to Z$. Semantically, the multivalued dependency $X \to\to Y$ in R(X, Y, Z) means the following (see Section 2):

(1) Each value of X determines a set of values of Y.

(2) Y and Z are conditionally independent.

$Y \to\to Z$ in R gives the following:

(3) Each value of Y determines a set of values of Z.

(4) X and Z are conditionally independent.

There is a semantic contradiction between (2) and (3). We shall show that the real transitivity condition for multivalues dependencies (i.e., $X \to\to Y$ and $Y \to\to Z$) never holds in a usual sense. On the other hand, we know that transitivity is often applied to prove other inference rules. However, such transitivity does not exist originally but is derived by augmentation. When two given multivalued dependencies originally do not meet transitivity, the following are shown:

(a) It is always possible by augmentation to obtain a transitivity condition from the given multivalued dependencies.

(b) The results obtained by such transitivity are derivable without using transitivity.

We also discuss an inference rule which is a mixture of functional and multivalued dependencies. It was presented in [4] and appears in Section 2 as the FD-MVD2 rule. We show a close resemblance between transitivity and FD-MVD2, and that the assumption of FD-MVD2 is not semantically acceptable either.

The paper is organized into four sections. Section 2 reviews the concepts and formal notations of functional dependencies and multivalued dependencies. Section 3 discusses the problem of applying the transitivity rule for multivalued dependencies. It also identifies what type of relations are in Boyce-Codd normal form but not in fourth normal form. Section 4 contains the conclusions of the paper.

2. Basic Concepts

In this section we briefly review and define the relational terminology. More detailed discussions are made elsewhere [2, 3, 4, 7, 8, 10, 20].

2.1 Relations

An **attribute** is a symbol taken from a finite set $U = \{A_1, A_2, \cdots, A_n\}$. For each attribute there is a set of possible values called its **domain**, denoted by DOM(A). One domain can be associated with more than one attribute. We will use capital letters from the beginning of the alphabet $(A, B, \cdots)$ for single attributes, and capital letters from the end of the alphabet $(X, Y, \cdots)$ for sets of attributes. For a set of attributes X, let x denote the values assigned to these attributes from their respective domains. The notation XY will be used to represent the union of two arbitrary sets of attributes.

A **relation** on the set of attributes $U = \{A_1, A_2, \cdots, A_n\}$ is a subset of the Cartesian product $DOM(A_1) \times \cdots \times DOM(A_n)$. The elements (rows) of R are called **tuples.** A relation R on $\{A_1, \cdots, A_n\}$ will be denoted by $R(A_1, \cdots, A_n)$. Similarly, if R is defined on the union of sets $X_1, \cdots, X_m$, then the notation $R(X_1, \cdots, X_m)$ will be used. The word **relation scheme** denotes the structure (description) of the relation. A relation is time-varing because of the insertion, deletion and modification of tuples, while the relation scheme is not unless the structure itself is changed. **Data dependencies are specified on relational schemes.** When we say that a data dependency holds for a relation scheme, we mean that **every relation which is an iustance of the scheme obeys that dependency.** Thus we will use the word "relation" instead of "relation scheme" hereafter when there is no confusion.

Let u be a tuple in R(U). If Y is a subset of U, then u[Y] is the tuple which contains the components of u corresponding to the elements of Y. The **projection** of R on Y, denoted by R[Y], is defined by

$$R[Y] = \{u[Y] \mid u \in R\}.$$

Similarly the **conditional projection** of R on Y by a value x for attribute X, where $X \subseteq U$, is defined as follows:

$$R[x, Y] = \{u[Y] \mid u \in R \text{ and } u[X] = x\}.$$

Let R(X, Y) and S(Y, Z) be relations where X, Y and Z are disjoint sets of attributes. The **natural join** of R and S, denoted by R*S, is the relation T(X, Y, Z) whose attributes are XYZ, and is defined by:

$$T(X, Y, Z) = R(X, Y)*S(Y, Z) = \{(x, y, z) \mid (x, y) \in R \text{ and } (y, z) \in S\}$$

Let R(U) be a relation and let X be a subset of U. We say that X is a **superkey** of R if every attribute in R functionally depends on X. If X is a superkey and no proper subset of X is a superkey, X is a **key** of R. We say that R is an **all-key** relation if the key X is equal to U.

2.2 Data Dependencies and Normal Forms

Data dependencies provide two kinds of information: integrity constraints and structural information of data.

Functional dependency (FD) is a statement f: $X \to Y$ where X and Y are sets of attributes. When we say that the FD f holds for a relation $R(X, Y, \cdots)$, every two tuples of R that have the same X-value also have the same Y-value. When f holds, we say that Y is **functionally dependent** on X or that X **functionally determines** Y, and we usually write $X \to Y$ for simplicity.

Let X and Y be subsets of U, where U is the set of attributes of a relation R(U). And let $Z = U - XY$. We say that there is a **multivalued dependency** (MVD) from X to Y in R(U), denoted by g: $X \to\to Y$, if and only if (iff) for every XZ-value xz in R(U),

$$R[xz, Y] = R[x, Y].$$

When g holds, we say that X **multidetermines** Y. Informally speaking, Y and Z are independent under X. When $X \cap Y \neq \phi$, $X \to\to Y$ holds for R iff $X \to\to Y - X$. This is a direct consequence of the definition. By $X \to\to Y_1 | Y_2 | \cdots | Y_n$, we mean that $X \to\to Y_i$ holds for each set Y_i $(i = 1, 2, \cdots, n)$.

From the definitions, an FD $X \to Y$ is defined by the sets of attributes X and Y alone and independent of other attributes in the relation. On the other hand, the validity of an MVD $X \to\to Y$ in R(X, Y, Z) depends on the existence of Z and thus cannot be determined by X and Y alone. It is possible that $X \to\to Y$ is not valid in R(X, Y, Z) but is valid in R(X, Y, Z'), where $Z' \subset Z$. Thus, MVD's cannot represent all one-to-many and many-to-many relationships, and they are **context sensitive.** This context sensitivity makes it difficult to specify proper MVD's especially when there are many attributes in a relation.

An FD $X \to Y$ is said to be **trivial** if $Y \subseteq X$. A trivial MVD $X \to\to Y$ in R(X, Y, Z) is defined when $Y \subseteq X$ or $Z = \phi$. An FD $X \to Y$ and an MVD $X \to\to Y$ are called **full** if for any proper subset X' of X, $X' \not\to Y$ and $X' \not\to\to Y$ respectively. Y and Z are said to be **conditionally independent** to each other under X if $X \to\to Y$ and $X \to\to Z$ (shortly $X \to\to Y|Z$) hold in R(X, Y, Z). If $X = \phi$ (namely $\phi \to\to Y|Z$), then they are called **completely independent.**

A nontrivial MVD $X \to\to Y$ is called **strong** unless it is an FD. In the discussion of the semantics of MVD's in Section 3, we are concerned with only strong MVD's.

As a complete set of inference rules for FD's, Armstrong [2] presented the following rules:

FD1 (reflexivity): If $Y \subseteq X$ then $X \to Y$.
FD2 (augmentation): If $Z \subseteq W$ and $X \to Y$ then $XW \to YZ$.
FD3 (transitivity): If $X \to Y$ and $Y \to Z$ then $X \to Z$.

Redundant rules:
FD4 (pseudo-transitivity): If $X \to Y$ and $YW \to Z$ then $XW \to Z$.
FD5 (union): If $X \to Y$ and $X \to Z$ then $X \to YZ$.
FD6 (decomposition): If $X \to YZ$ then $X \to Y$ and $X \to Z$.

Beeri et al. [4] axiomatized both FD's and MVD's and presented the following inference rules:

MVD0 (complementation): If $XYZ = U$ and $Y \cap Z \subseteq X$, then $X \to\to Y$ iff $X \to\to Z$.
MVD1 (reflexivity): If $Y \subseteq X$ then $X \to\to Y$.
MVD2 (augmentation): If $Z \subseteq W$ and $X \to\to Y$ then $XW \to\to YZ$.
MVD3 (transitivity): If $X \to\to Y$ and $Y \to\to Z$ then $X \to\to Z - Y$.

Redundant rules:
MVD4 (pseudo-transitivity): If $X \to\to Y$ and $YW \to\to Z$ then $XW \to\to Z - YW$.
MVD5 (union): If $X \to\to Y$ and $X \to\to Z$ then $X \to\to YZ$.
MVD6 (decomposition): If $X \to\to Y$ and $X \to\to Z$ then $X \to\to Y \cap Z$, $X \to\to Y - Z$ and $X \to\to Z - Y$.

FD-MVD rules:
FD-MVD1: If $X \to Y$ then $X \to\to Y$.
FD-MVD2: If $X \to\to Z$ and $Y \to Z'$, where $Z' \subseteq Z$ and Y and Z are disjoint, then $X \to Z'$.

We will discuss in Section 3 the problems in MVD3, MVD4 and FD-MVD2.

A relation scheme R is in **Boyce-Codd normal form** (BCNF) if, whenever a nontrivial FD: $X \to Y$ holds for R, then so does the FD: $X \to A$ for every attribute A of R [8]. In other words, X is a superkey of R. BCNF is a stronger version of **third normal form** (3NF) [7]. A relation R is in **fourth normal form** (4NF) if, whenever a nontrivial MVD $X \to\to Y$ holds for R, then so does the FD $X \to A$ for every attribute A of R [10]. 4NF is stronger than BCNF. If R is in 4NF, then it is in BCNF.

3. Semantic Considerations on Multivalued Dependencies

3.1 Difference between 4NF and BCNF

In this section we discuss the difference between 4NF and BCNF. What kind of BCNF relations are not in 4NF? In order to examine this, let us go back to Table 1. The relation R(EMPLOYEE, CHILD, YEAR, SALARY) is in BCNF but not in 4NF. The strong MVD: EMPLOYEE $\to\to$ CHILD|{YEAR, SALARY} holds for R. Apparently, R is an all-key relation.

Assume that the strong MVD: $X \to\to Y|Z$ holds for the BCNF relation R(X, Y, Z). It is obvious that X is not a key for R since X does not uniquely determine Y and Z. By the definition of MVD, if (x, y_1, z_1) and (x, y_2, z_2) are the tuples of R, then so are (x, y_1, z_2) and (x, y_2, z_1). Therefore, XY and XZ cannot be a key, either. Then, the question is: is it only all-key relations that satisfy the condition of BCNF and non-4NF? From the semantic point of view, we believe the answer is yes. Syntactically, however, there is an exception. We clarify it by the following lemma:

Lemma 1: Assume that the nontrivial MVD $X_1X_2\rightarrow\rightarrow Y|Z$ holds for $R(X_1, X_2, Y, Z)$, where X_1, X_2, Y, and Z are disjoint. Also assume $X_1\not\rightarrow X_2$. Then the FD $X_1YZ\rightarrow X_2$ holds iff the following condition is satisfied:

For every two X_1X_2-values x_1x_2 and x_1x_2' $(x_2\neq x_2')$,

$Y_1\cap Y_2=\phi$ or $Z_1\cap Z_2=\phi$, where $Y_1=R[x_1x_2, Y]$,

$Z_1=R[x_1x_2, Z]$, $Y_2=R[x_1x_2', Y]$ and $Z_2=R[x_1x_2', Z]$.

Proof: (1) (if-part) Let $S_1=\{x_1\}\times Y_1\times Z_1$ and $S_2=\{x_1\}\times Y_2\times Z_2$. Then $S_1=R[x_1x_2, X_1, Y, Z]$ and $S_2=R[x_1x_2', X_1, Y, Z]$ because $X_1X_2\rightarrow\rightarrow Y|Z$ holds for R. By the assumption, $Y_1\cap Y_2=\phi$ or $Z_1\cap Z_2=\phi$. Hence, $S_1\cap S_2=\phi$. This implies that no X_1YZ-value is associated with more than one X_2-value, which is equivalent to $X_1YZ\rightarrow X_2$.

(2) (only-if-part) Assume $Y_1\cap Y_2\neq\phi$ and $Z_1\cap Z_2\neq\phi$; we derive a contradiction. Let $y\in Y_1\cap Y_2$ and $z\in Z_1\cap Z_2$. Then both (x_1, x_2, y, z) and (x_1, x_2', y, z) are tuples of R because $X_1X_2\rightarrow\rightarrow Y|Z$. This implies that $X_1YZ\rightarrow X_2$ does not hold. This is a contradiction. □

We introduce the following theorem:

Theorem 1: If a relation R is in BCNF but not in 4NF, then R is an all-key relation except the condition of Lemma 1.

Proof: Let R(X, Y, Z) be in BCNF but not in 4NF, where X, Y and Z are disjoint. Assume that the strong MVD $X\rightarrow\rightarrow Y|Z$ holds for R. Let K be a key of R. There are three cases to be considered.

(1) Case 1: Assume $K\not\supseteq Y$. Let $Y=Y_1Y_2$ such that $K\supseteq Y_1$ and $K\cap Y_2=\phi$ $(Y_2\neq\phi)$. By the definition of MVD's,

$$R[x, Y, Z]=R[x, Y]\times R[x, Z]. \qquad (1)$$

Let $R[x, Y]=Y_0$ and $R[x, Z]=Z_0$. If y_1y_2 and y_1y_2' $(y_2\neq y_2')$ are elements of Y_0, then both (x, y_1y_2, z) and (x, y_1y_2', z) are tuples of R by eq. (1), where $z\in Z_0$. This violates the FD $XY_1Z\rightarrow Y_2$ because there are two Y_2 values, y_2 and y_2', for the same XY_1Z-value xy_1z. ($XY_1Z\rightarrow Y_2$ is derived by $K\rightarrow Y_2$ and $K\subseteq XY_1Z$.) Hence, if $y_1y_2\in Y_0$ then $y_1y_2'\notin Y_0$. This implies $XY_1\rightarrow Y_2$. $XY_1\not\rightarrow Z$ is obvious by eq. (1). (E.g. if $R[x, Z]=\{z_1, z_2\}$ and $y_1y_2\in R[x, Y]$, then $(x, y_1y_2, z_1)\in R$ and $(x, y_1y_2, z_2)\in R$.) This contradicts the assumption that R is in BCNF. Hence we get $K\supseteq Y$.

(2) Case 2: Assume $K\not\supseteq Z$. As Y and Z are symmetric in R, we can directly conclude $K\supseteq Z$ from the above discussion by exchanging Y and Z.

(3) Case 3: Assume $K\not\supseteq X$. Let $X=X_1X_2$ such that $K\supseteq X_1$ and $K\cap X_2=\phi$ $(X_2\neq\phi)$. Let $R[x_1x_2, Y]=Y_1$, $R[x_1x_2, Z]=Z_1$, $R[x_1x_2', Y]=Y_2$ and $R[x_1x_2', Z]=Z_2$, where x_1x_2 and $x_1x_2'\in X_1X_2$ and $x_2\neq x_2'$. Since the condition of Lemma 1 is excluded, there may be two X_1X_2-values, x_1x_2 and x_1x_2', such that $Y_1\cap Y_2\neq\phi$ and $Z_1\cap Z_2\neq\phi$. This implies $X_1YZ\not\rightarrow X_2$ and we obtain $K\supseteq X$.

From (1), (2) and (3), we get $K\supseteq XYZ$. Since $K\subseteq XYZ$, the desired result $K=XYZ$ is obtained. □

Before discussing the result of Theorem 1, we have to examine the possibility that the condition of Lemma 1 takes place. If such a condition is a usual one, Theorem 1 becomes meaningless. Semantically, the constraint that $Y_1\cap Y_2=\phi$ or $Z_1\cap Z_2=\phi$ is strange and does not seem to be controllable in an actual environment because the FD: $X_1YZ\rightarrow X_2$ produces more problems than it solves. If we decompose R into two 4NF relations, $R_1(X_1, X_2, Y)$ and $R_2(X_1, X_2, Z)$, it becomes extremely difficult to preserve the constraint $X_1YZ\rightarrow X_2$. Each time R_1 or R_2 is updated (inserted, deleted, or replaced), it must be confirmed that the constraint $Y_1\cap Y_2=\phi$ or $Z_1\cap Z_2=\phi$ is satisfied for appropriate X_1X_2-values or that X_1YZ is unique by joining R_1 and R_2. (According to [18], R_1 and R_2 are not "independent".) Of course, other problems may occur unless we decompose R [10, 19, 20]. Zaniolo and Melkanoff [21] have presented an example which belongs to Lemma 1. WS (DAY, TIME, GROUP) is the weekly schedule of occupancy of a conference room. There are two constraints on WS.

(1) Only one group meets at any given day and time in the room.
(2) A group must follow the same time schedule for any day when it uses the room.

The first constraint implies {DAY, TIME}→GROUP and the second gives GROUP→→DAY|TIME. This is the case where $X_1=\phi$ in Lemma 1. The second constraint, however, is unnatural and does not seem to be the case in the real world. Even if it were the case, the decomposition into WS1 (DAY, GROUP) and WS2 (TIME, GROUP) causes the difficult problems mentioned above.

From our experience, it is unlikely to have relations which satisfy the condition of Lemma 1 unless unnatural assumptions like the relation WS are made. Furthermore, even if there exist such relations that satisfy the condition of Lemma 1, they never increase the usefulness of 4NF because the decomposition into 4NF relations creates difficulties while it solves other problems as mentioned above. (In the example of WS, the decomposition should not be made because assumption (1) is essential.) Therefore, **we exclude the case of Lemma 1 in the discussion of Theorem 1.**

Theorem 1 tells us that the relations which are in BCNF but not in 4NF are only all-key relations. The main usage of all-key relations is to represent relationships between other relations. For example, R(EMP#, PROJ#) represents the relationship between "Employee" and "Project" relations. Similarly, S(PROJ#, PART#, SUPP#) represents the relationship among "Project," "Part" and "Supplier" relations. Those are all-key relations, but they are in 4NF. There is a significant difference, however, between R (or S) and BCNF but non-4NF relations. The semantics of R is clear and we can add non-key attributes to R if necessary (e.g., PERCENT-OF-TIME and START-DATE).

3.2 Problems of Transitivity

The inference rules for MVD's are analogous to those for FD's. There is a significant difference, however, between FD's and MVD's. MVD's are sensitive to context while FD's are not. This sensitivity produces problems in the transitivity rule (MVD3) for MVD's.

If both X→Y and Y→Z are valid in a relation, X→Y, Y→Z and, transitively, X→Z always hold for any relation whose attributes include X, Y, and Z. For example, if EMP→DEPT and DEPT→MANAGER are valid in S(EMP, DEPT, AGE) and T(DEPT, MANAGER, BUDGET) respectively, these FD's and, transitively, EMP→MANAGER are valid in R(EMP, DEPT, MANAGER) and in any other relation which includes EMP, DEPT and MANAGER.

However, the same arguments can not be made for MVD's. Let's assume the existence of the MVD: PROJ→→EMP in S(PROJ, EMP, PART) and the MVD: EMP→→CHILD in T(EMP, CHILD, PROD). Then someone may ask; does PROJ→→CHILD (also PROJ→→EMP) hold in R(PROJ, EMP, CHILD)? As will be demonstrated later, the answer is no. (On the other hand, EMP→→PROJ |CHILD holds for R.) This example suggests that the transitivity condition X→→Y and Y→→Z in R(X, Y, Z) will not hold. As mentioned in Section 1, X→→Y semantically tells us the following:

(1) Each value of X determines a set of values of Y.
(2) Y and Z are conditionally independent.

Y→→Z gives the following:

(3) Each value of Y determines a set of values of Z.
(4) X and Z are conditionally independent.

There is a semantic contradiction between (2) and (3). Therefore, we have the following theorem: (Note that unless X→→Y and Y→→Z are valid in S[X, Y, Z] which is the projection of S(X, Y, Z, W), they are not valid in S, either [10]. Therefore, we discuss the relations which include only X, Y and Z.)

Theorem 2: Let R(X, Y, Z), S(X, Y, V), and T(Y, Z, W) be the projections of U(X, Y, Z, V, W), where X, Y, Z, V, and W are disjoint. Assume the following:

(1) X→→Y holds for S and Y→→Z holds for T, where X→→Y and Y→→Z are strong MVD's.
(2) In U, Y determines Z-values independently of X (i.e. there can be y_1 and y_2 which satisfy $U[x, y_1, Z] \neq U[x, y_2, Z]$).

Then X→→Y does not hold for R.

Proof: Suppose that X→→Y holds in R, and we proceed to derive a contradiction. Let $S[x, Y] = Y_0$. Let $T[y_1, Z] = Z_1$ and $T[y_2, Z] = Z_2$, where y_1 and $y_2 \in Y_0$. It is possible to choose y_1 and y_2 such that $Z_1 - Z_2 \neq \phi$ because a different Y-value can determine a different set of Z-values independently of X. Let $z_1 \in (Z_1 - Z_2)$ and $z_2 \in Z_2$. Then (x, y_1, z_1) and (x, y_2, z_2) are tuples in R. Since X→→Y, the tuples (x, y_1, z_2) and (x, y_2, z_1) must also be in R. Therefore, $R[y_2, Z]$ includes z_1 which is not an element of Z_2. Hence, $R[Y, Z] \neq T[Y, Z]$. This contradicts the fact that both R and T are the projections of U. □

What about Y→→Z in R? Contrary to X→→Y, Y→→Z holds in R as long as X and Z are independent. Such cases are common. As was shown in the previous example, EMP→→PROJ|CHILD holds for R(PROJ, EMP, CHILD). However, if we select PROD from T(EMP, CHILD, PROD) instead of CHILD, then EMP→→PROJ|PROD does not hold for R′(PROJ, EMP, PROD) because PROJ and PROD are not independent. (No MVD's hold for R′; except trivial ones.)

The following are examples of Theorem 2:

(1) R(PROJ, PART, SUPP), S(PROJ, PART, EMP), and T(PART, SUPP, SUBPART).
(2) R(TEACHER, STUDENT, HOBBY), S(TEACHER, STUDENT, PHONE), and T(STUDENT, HOBBY, FRIEND).
(3) R(LIB, BOOK, AUTHOR), S(LIB, BOOK, LIBRARIAN), and T(BOOK, AUTHOR, KEYWORD).
(4) R(MAKER, DEALER, CAR), S(MAKER, DEALER, BANK), and T(DEALER, CAR, WORKER).
(5) R(STATE, CITY, PARK), S(STATE, CITY, HIST-OF-GOVERNOR), and T(CITY, PARK, MUSEUM).

Assumption (2) of Theorem 2 is so natural that we could not find counter examples to the assumption. (Of course, if ϕ→→X|Y|Z holds for R(X, Y, Z), all of the following MVD's hold: X→→Y|Z, Y→→Z|X and Z→→X|Y. However, such a relation is neither a common nor a useful one.) Therefore, we examine transitivity from another viewpoint. Namely, we clarify the condition in which transitivity holds.

Theorem 3: Let R(X, Y, Z) be a relation, where X, Y, and Z are disjoint. X→→Y and Y→→Z hold for R iff, for every XY-value xy in R,

$$R[x, Z] = R[x, y, Z] = R[y, Z]. \quad (2)$$

Proof: (1) (if-part) By the definition of MVD, X→→Z is obtained from R[x, Z] = R[x, y, Z]. By complementation, X→→Y is obtained. We obtain Y→→Z from R[x, y, Z] = R[y, Z].

(2) (only-if-part) Since X→→Y is equivalent to X→→Z in R, eq. (2) is directly derived by X→→Y and Y→→Z. □

We obtain the following two corollaries from Theorem 3.

Corollary 1: Let R(X, Y, Z, W) be a relation, where X, Y, Z, and W are disjoint. If X→→Y and Y→→Z hold for R, then R must satisfy the following:

(1) Every Y-value y in R[x, Y] determines the same set of Z-values as R[x, Z].
(2) Every X-value x in R[y, X] determines the same set of Z-values as R[y, Z].

Proof: Statement (1) is equivalent to R[y, Z] = R[x, Z], where y∈R[x, Y]. Statement (2) is equivalent to R[x, Z] = R[y, Z], where x∈R[y, X]. Since X→→Y and Y→→Z still hold for R[X, Y, Z], the proof is direct from Theorem 3. □

The following is presented by Kambayashi et al. [12].

Corollary 2: Let R(X, Y, Z, W) be a relation, where X, Y, Z and W are disjoint. If $X \rightarrow\rightarrow Y$ and $Y \rightarrow\rightarrow Z$ hold for R, then R must satisfy the following:

(1) If $R[x, Y] \cap R[x', Y] \neq \phi$, where $x \neq x'$, then $R[x, Z] = R[x', Z]$.

(2) If $R[y, X] \cap R[y', X] \neq \phi$, where $y \neq y'$, then $R[y, Z] = R[y', Z]$.

Proof: Since $X \rightarrow\rightarrow Y$ and $Y \rightarrow\rightarrow Z$ hold for R[X, Y, Z], the proof can be directly made by eq. (2) as follows:

(1) $R[x, Z] = R[x, y, Z] = R[y, Z] = R[x', y, Z] = R[x', Z]$ where, $y \in R[x, Y] \cap R[x', Y]$.

(2) $R[y, Z] = R[x, y, Z] = R[x, Z] = R[x, y', Z] = R[y', Z]$ where, $x \in R[y, X] \cap R[y', X]$. □

The results of Corollary 1 and Corollary 2 can not be semantically accepted and do not seem to occur in the real world. Therefore, we may say that assumption (2) of Theorem 2 restricts virtually nothing on the attributes X, Y and Z, and that transitivity condition ($X \rightarrow\rightarrow Y$ and $Y \rightarrow\rightarrow Z$) never occurs in a natural sense. In addition, the transitivity rule is misleading. The database designer might specify the wrong MVD: $X \rightarrow\rightarrow Y$ in R(X, Y, Z) when $X \rightarrow\rightarrow Y$ and $Y \rightarrow\rightarrow Z$ hold in S(X, Y, V) and T(Y, Z, W) respectively.

Such an example is shown in Figure 1. PROJ→→EMP does not hold for R(PROJ, EMP, CHILD), while EMP→→CHILD holds, even if PROJ→→EMP and EMP→→CHILD hold in S(PROJ, EMP, PART) and T(EMP, CHILD, PROD) respectively. If the database designer mistakenly specifies PROJ→→EMP in R in addition to EMP→→CHILD, undesirable tuples depicted in Figure 1 appear. In Figure 1, the correct data is depicted in a hierarchical structure. Figure 1-(a) shows the relation R when the tuple (P2, E1, C1) does not exist, and Figure 1-(b) represents R after the tuple was inserted. As shown in Figure 1-(a), there are three unallowable tuples caused by Corollary 1-(1) or Corollary 2-(2). In addition, eleven unallowable tuples have appeared in Figure 1-(b) by the addition of (P2, E1, C1),

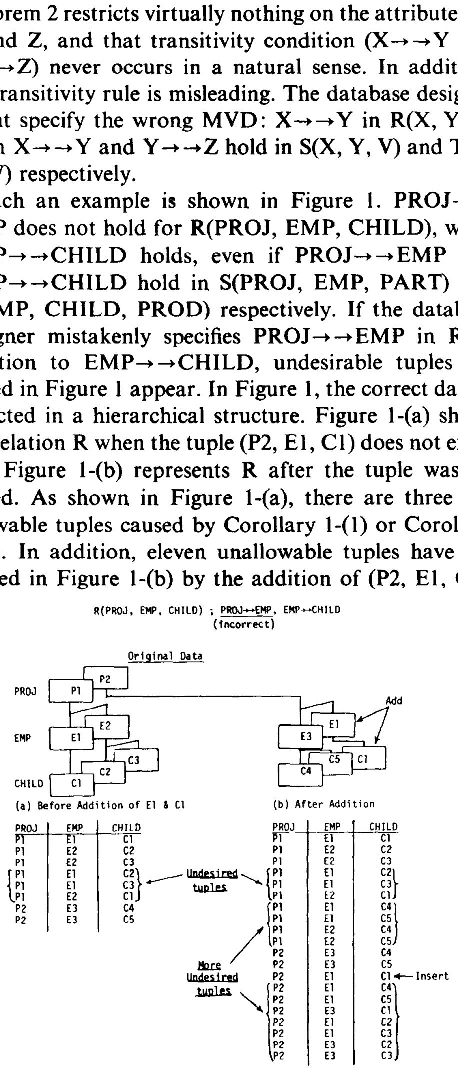

Fig. 1 Anomalies caused by incorrect transitive MVD's.

which is the result of Corollary 1-(2) or Corollary 2-(1).

On the other hand, we have to refer to the fact that transitivity is often applied especially when other inference rules are proved. For example, the union rule (MVD5) is derived as follows: Assume $X \rightarrow\rightarrow Y$ and $X \rightarrow\rightarrow Z$ in R(X, Y, Z, W), where X, Y, Z, and W are disjoint.

(1) $X \rightarrow\rightarrow Y$; given
(2) $X \rightarrow\rightarrow Z$; given
(3) $X \rightarrow\rightarrow XY$; (1), MVD2
(4) $XY \rightarrow\rightarrow YZ$; (2), MVD2
(5) $XY \rightarrow\rightarrow W$; (4), MVD0
(6) $X \rightarrow\rightarrow W$; (3), (5), **MVD3**
(7) $X \rightarrow\rightarrow YZ$; (6), MVD0

We have to distinguish the above case, however, from Theorem 2. In the above example, originally there was no transitivity condition in R. We will show that we can always obtain a transitivity condition when two MVD's are given, and that the results obtained by such transitivity can be derived by other rules.

Lemma 2: Given two MVD's. It is always possible to create a transitivity condition by MVD2.

Proof: Let R(U) be a relation. Assume that $X \rightarrow\rightarrow Y$ and $Z \rightarrow\rightarrow V$ hold for R(U), where $XYZV \subseteq U$. It is general enough to prove that we can derive a transitivity condition in R(X, Y, Z, V, W), where $W = U - XYZV$. We augment $X \rightarrow\rightarrow Y$ by Z to obtain $XZ \rightarrow\rightarrow YZ$. Augmenting $Z \rightarrow\rightarrow V$ by Y, we obtain $YZ \rightarrow\rightarrow YV$. Thus we have the transitivity condition: $XZ \rightarrow\rightarrow YZ$ and $YZ \rightarrow\rightarrow YV$. □

Lemma 3: Assume that $X \rightarrow\rightarrow Y$ and $Z \rightarrow\rightarrow V$ hold for R(X, Y, Z, V, W), where $X \cap Y = \phi$, $Z \cap V = \phi$ and $W \cap XYZV = \phi$. Neither $X \rightarrow\rightarrow Y$ and $Z \rightarrow\rightarrow V$ nor any projection of them meets MVD3 and MVD4 iff $Y \cap ZV = \phi$ and $V \cap XY = \phi$.

Proof: Since the if-part is obvious, we prove the only-if-part. Assume $Y \cap ZV \neq \phi$ or $V \cap XY \neq \phi$. There are three cases to be examined:

(1) Assume $Y \cap Z \neq \phi$. Let $A \subseteq Y \cap Z$, $Y' = Y - A$ and $Z' = Z - A$. Then $X \rightarrow\rightarrow A$ holds for R[X, A, Z', V, W]. Since $Z \rightarrow\rightarrow V$ is equivalent to $AZ' \rightarrow\rightarrow V$, we obtain $X \rightarrow\rightarrow A$ and $AZ' \rightarrow\rightarrow V$ which meet MVD4. Hence, $Y \cap Z$ must be null.

(2) Assume $V \cap X \neq \phi$. This follows (1). Hence, $V \cap X$ must be null.

(3) Assume $Y \cap V \neq \phi$. Let $A \subseteq Y \cap V$. Then $X \rightarrow\rightarrow A$ and $Z \rightarrow\rightarrow A$ hold for R[X, A, Z], which meets MVD3 since $X \rightarrow\rightarrow A$ implies $X \rightarrow\rightarrow Z$ in R[X, A, Z]. Hence, $Y \cap V$ must be null. □

Theorem 4: Given two MVD's. Assume that those MVD's and any projection of them do not meet MVD3 and MVD4. Then the results obtained by any transitivity which is created from Lemma 2 can be derived by MVD0, MVD2 and MVD5.

Proof: Let $X \rightarrow\rightarrow Y$ and $Z \rightarrow\rightarrow V$ be MVD's in R(X, Y, Z, V, W), where $X \cap Y = \phi$, $Z \cap V = \phi$ and $W \cap XYZV = \phi$. By Lemma 3, we obtain $Y \cap XZVW = \phi$ and

$V \cap XYZW = \phi$. For the proof of Theorem 4, it is sufficient to show that any result obtained by transitivities which are created through Lemma 2 is also derivable by MVD0, MVD2 and MVD5. Starting from $X \to\to Y$ and $Z \to\to V$, we obtain the following transitivities:

(1) $X \to\to Y$; given
(2) $Z \to\to V$; given
(3) $XZ \to\to YZ$; (1), MVD2
(4) $YZ \to\to YV$; (2), MVD2
(5) $XZ \to\to XV$; (2), MVD2
(6) $XV \to\to YV$; (1), MVD2

(3) and (4) are minimum augmentation which produces a transitivity. (5) and (6) are another minimum augmentation, and there is no other minimum augmentation. From (3) and (4), we obtain the following results:

*(7) $XZ \to\to V$; (3), (4), MVD3, $V \cap XYZ = \phi$
(8) $YZ \to\to XW$; (4), MVD0
(9) $XZ \to\to XW - Z$; (3), (8), MVD3
*(10) $XZ \to\to W$; (9), $W \cap XZ = \phi$, definition

(5) and (6) give the following:

*(11) $XZ \to\to Y$; (5), (6), MVD3, $Y \cap XZV = \phi$
(12) $XV \to\to ZW$; (6), MVD0
(13) $XZ \to\to ZW - Z$; (5), (12), MVD3
(14) $XZ \to\to W$; (13), $W \cap XZ = \phi$, definition

Operations (8)–(9)–(10) are reversible. Operations (12)–(13)–(14) are also reversible. Since (10) and (14) are identical, we must prove that we can derive (7), (10) and (11) (marked with *'s) without transitivity. The proof is as follows:

(15) $XZ \to\to Y$; (1), MVD2
(16) $XZ \to\to V$; (2), MVD2
(17) $XZ \to\to YV$; (15), (16), MVD5
(18) $XZ \to\to W$; (17), MVD0, $YV \cap W = \phi$

(15), (16), and (18) give the results.

Redundant augmentation gives the only possibility to obtain other transitivities from (1) and (2). Let Q be an arbitrary set of attributes in R. Then redundant augmentation gives the following transitivities:

(19) $XZQ \to\to YZQ$; (1), MVD2
(20) $YZQ \to\to YVQ$; (2), MVD2
(21) $XZQ \to\to XVQ$; (2), MVD2
(22) $XVQ \to\to YVQ$; (1), MVD2

It is enough to prove that the results obtained by (19) and (20) can be derived without transitivity since similar arguments can be made on (21) and (22).

*(23) $XZQ \to\to V - Q$; (19), (20), MVD3, $V \cap XYZ = \phi$
(24) $YZQ \to\to XW$; (20), MVD0
(25) $XZQ \to\to XW - ZQ$; (19), (24), MVD3
*(26) $XZQ \to\to W - Q$; (25), $W \cap XZ = \phi$

The derivation of (23) and (26) is direct from (16) and (18) by MVD2 and the definition of MVD. □

We have shown by Theorem 2 and Theorem 3 that the transitivity condition never occurs in a natural sense. We have also shown that it is always possible by augmentation to obtain the transitivity condition when two MVD's are given which originally do not meet transitivity and pseudo-transitivity, and that the results obtained by such transitivity are derivable without transitivity. Therefore, the transitivity rule becomes meaningless in an actual environment. The same arguments can be made on pseudo-transitivity (MVD4) because it is one variation of transitivity.

Now we discuss FD-MVD2. It may appear that there is no close resemblance between transitivity and FD-MVD2. Actually, we can make almost the same arguments about FD-MVD2 as those for transitivity. For convenience, we assume that $Z' = Z$ and X, Y, and Z are all the attributes in the relation R, where X, Y, and Z are disjoint. (Such condition is always obtainable by projection.) Then, FD-MVD2 can be written as follows:

FD-MVD2′: If $X \to\to Z$ and $Y \to Z$ then $X \to Z$.

Since $X \to\to Z$ is equivalent to $X \to\to Y$ in R(X, Y, Z), the following is obtained:

FD-MVD2″: If $X \to\to Y$ and $Y \to Z$ then $X \to Z$.

Note the resemblance between transitivity and FD-MVD2″. The problem of FD-MVD2″ is even clearer because $X \to\to Y$ and $Y \to Z$ give the following:

(1) Y and Z are conditionally independent.
(2) Y functionally determines Z. (Every Y-value uniquely determines a Z-value.)

It goes without saying that there is a serious contradiction between statements (1) and (2). Similarly to transitivity, the following theorem holds:

Theorem 5: Let R(X, Y, Z), S(X, Y, V) and T(Y, Z, W) be the projections of U(X, Y, Z, V, W), where X, Y, Z, V, and W are disjoint. Assume the following:

(1) The strong MVD: $X \to\to Y$ holds for S and $Y \to Z$ holds for T.
(2) In U, Y determines Z-values independently of X. (The meaning is the same as Theorem 2.)

Then $X \to\to Y$ does not hold for R.

Proof: Let $Z_2 = \{z_2\}$ and $Z_1 = \{z_1\}$ hold in the proof of Theorem 2. Then the proof of Theorem 5 is the same as that of Theorem 2. □

When one specifies the MVD: $X \to\to Z$, he has to confirm that Z is independent of the attributes other than X and Z. The confirmation fails, however, because the FD: $Y \to Z$ already exists. From another viewpoint, if $X \to Z$ realy holds, not $X \to\to Z$ but $X \to Z$ must always be specified because the existence of FD's is much more definite than that of MVD's, and because $X \to\to Z$ gives less correct information than $X \to Z$. Once $X \to Z$ has been decided not to hold, the relationship between X and Z is neither one-to-one nor many-to-one, and $X \to Z$ should never be derived from $X \to\to Z$ and $Y \to Z$. ($X \to\to Z$ is wrong if $Y \to Z$ holds.) Therefore, one should not expect to get a new FD by FD-MVD2 which he did not initially specify. Instead, he must examine his initial specification for errors if such an FD appears.

For example, if one specifies DEPT→→PHONE and EMP→PHONE (or DEPT→→EMP and EMP→PHONE) in R(DEPT, EMP, PHONE, ···), he gets the invalid FD: DEPT→PHONE which states that there is only one telephone in each department. (DEPT→→PHONE and DEPT→→EMP are not valid in R.) FD-

MVD2 is more misleading than transitivity because such an invalid specification tends to occur especially when a relation has a large number of attributes.

As another example, let us examine the relation R(LANDLORD, ADDR, OCCUPANT,···) and the data dependencies: LANDLORD→→ADDR, OCCUPANT→ADDR, etc. which appears in [3]. From these data dependencies, the authors in [3] derived the FD: LANDLORD→ADDR. However, the same authors also concluded, based on their analysis of the real world situation, that each landlord may own many buildings. As long as this analysis is true, LANDLORD→ADDR never holds. Such an invalid derivation comes from the incorrect MVD: LANDLOAD→→ADDR.

4. Conclusions

We have discussed multivalued dependencies from the semantic point of view and made clear the following items:

(1) With a minor exception, the relations which are in BCNF but not 4NF are **all key relations.**

(2) The real transitivity condition for multivalued dependencies never occurs in a natural sense.

(3) When two given MVD's do not originally meet transitivity, it is always possible, by augmentation, to derive the transitivity condition. Since the results obtained by such derived transitivity can be obtained by other MVD rules, there is thus no need to explicitly derive the transitivity condition.

(4) The assumption of FD-MVD2 rule never holds in a natural sense.

We have shown that the exceptions of (1), (2) and (4) can not be semantically accepted. In practical database design, the results of (2), (3) and (4) make transitivity and FD-MVD2 meaningless and harmful as inference rules. Their usefulness remains only as error chekers. If the database designer find the condition of transitivity or FD-MVD2 in his initial specification, he must re-examine it to find errors.

The above problems closely relate to the fact that MVD's can not cover all one-to-many and many-to-many relationships between attributes, and they are context sensitive. Therefore, to extend our initial motivation, more research is needed for seeking a better representation of data dependencies.

Acknowledgements

One of the authors, Mr. Nakamura, would like to thank Dr. T. Miura, Dr. J. Kawasaki, Dr. T. Mitsumaki and Mr. I. Yoshida of Hitachi Systems Development Laboratory who gave him the opportunity to study at UCLA where this work was done. The authors also wish to thank Mr. Ilchoo Chung of UCLA for his very valuable comments. The many valuable comments of the referees led to a more understandable paper.

Reference

1. AHO, A. V., BEERI, C. and ULLMAN, J. D., The Theory of Joins in Relational Databases, *ACM Trans. on Database Systems*, **4**, 3 (September 1979).
2. ARMSTRONG, W. W., Dependency Structures of Database Relationships, *Proc. IFIP '74*, North Holland (1974).
3. BEERI, C., BERNSTEIN, P. A. and GOODMAN, N., A Sophisticate's Introduction to Database Normalization Theory, *Proc. 4th VLDB Conf.* (October 1978).
4. BEERI, C., FAGIN, R. and HOWARD, J. H., A Complete Axiomatization for Functional and Multivalued Dependencies in Database Relations, *Proc. SIGMOD Conf.* (August 1977).
5. CHEN, P. P., The Entity-Relationship Model—Toward a Unified View of Data, *ACM Trans. on Database Systems*, **1**, 1 (March 1976).
6. CODD, E. F., A Relational Model for Large Shared Data Banks, *Comm. ACM*, **13**, 6 (June 1970).
7. CODD, E. F., Further Normalization of the Data Base Relational Model, in *Data Base Systems* (Courant Computer Science Symposium 6), Prentice-Hall (1972).
8. CODD, E. F., Recent Investigations in Relational Data Base Systems, *Proc. IFIP '74*, North Holland (1974).
9. DELOBEL, C., Normalization and Hierarchical Dependencies in the Relational Data Model, *ACM Trans. on Database Systems*, 3, 3 (September 1978).
10. FAGIN, R., Multivalued Dependencies and a New Normal Form for Relational Databases, *ACM Trans. on Database Systems*, 2, 3 (September 1977).
11. FAGIN, R., Normal Forms and Relational Database Operators, *Proc. SIGMOD Conf.* (May 1979).
12. KAMBAYASHI, Y., TANAKA, K. and YAJIMA, S., Semantic Aspects of Data Dependencies and Their Application to Relational Database Design, *Proc. COMPSAC '79* (November 1979).
13. KAMBAYASHI, Y., TANAKA, T. and YAJIMA, S., Relational Database Design, *Proc. 13th IBM Comp. Sci. Symposium* (November 1979).
14. LIEN, Y. E., Multivalued Dependencies with Null Values in Relational Database, *Proc. VLDB Conf.* (October 1979).
15. MENDELZON, A. O., On Axiomatizing Multivalued Dependencies in Relational Databases, *J. ACM*, **26**, 1 (January 1979).
16. NICOLAS, J. M., Mutual Dependencies and Some Results on Undecomposable Relations, *Proc. VLDB Conf.* (October 1978).
17. PARKER, D. S. and DELOBEL, C., Algorithmic Applications for a New Result on Multivalued Dependencies, *Proc. VLDB Conf.* (October 1979).
18. RISSANEN, J., Independent Components of Relations, *ACM Trans. on Database Systems*, **2**, 4 (December 1977).
19. SCHMID, H. A. and SWENSON, J. R., On the Semantics of the Relational Model, *Proc. SIGMOD Conf.* (May 1975).
20. ZANIOLO, C., Analysis and Design of Relational Schemata for Database Systems, Ph. D. Dissertation, Computer Science Department, UCLA, Tech. Rep. UCLA-ENG-7669 (July 1976).
21. ZANIOLO, C. and MELKANOFF, M. A., A Formal Approach to the Definition and the Design of Conceputal Schemas for Database Systems (1979).