How to Prevent Interaction of Functional and Inclusion Dependencies

Mark Levene
Department of Computer Science
University College London
Gower Street
London WC1E 6BT, U.K.
Email: mlevene@cs.ucl.ac.uk

George Loizou
Department of Computer Science
Birkbeck College
Malet Street
London WC1E 7HX, U.K.
Email: george@dcs.bbk.ac.uk

Abstract

Functional dependencies (FDs) and inclusion dependencies (INDs) are the most fundamental integrity constraints that arise in practice in relational databases. A given set of FDs does not interact with a given set of INDs if logical implication of any FD can be determined solely by the given set of FDs, and logical implication of any IND can be determined solely by the given set of INDs. We exhibit a necessary condition and two novel sufficient conditions for a set of FDs and a set of proper circular INDs *not* to interact; these two sufficient conditions are orthogonal to known results in the database literature. We also discuss the difficulty in obtaining a syntactic necessary and sufficient condition for no interaction between FDs and INDs.

Keywords: Relation; Database; Logical implication; Functional dependency; Inclusion dependency

1 Introduction

The *implication problem* for FDs and INDs is the problem of deciding for a given set Σ of FDs and INDs whether Σ logically implies σ, where σ is an FD or an IND. The implication problem is central in data dependency theory and is also utilised in the process of database design, since it can be used to test whether two sets of dependencies are equivalent or to detect whether a dependency in a given set is redundant. The implication problem for FDs and INDs is known to be undecidable in the general case [Mit83, CV85] and can be decided only in exponential time when the INDs are restricted to be noncircular [CK86]. On the other hand, the implication problem for FDs on their own is known to be decidable in linear time [BB79] and the corresponding implication problem for noncircular INDs again on their own is known to be NP-complete [MR92] (for INDs, which may be circular, the implication problem is PSPACE-complete [CA84]). Thus given a set Σ of FDs and INDs and an FD or IND σ, it would be desirable if the set F of FDs and the set I of INDs do *not interact*, in the sense that the implication problem of whether Σ logically implies σ can be decided by F on its own, when σ is an FD, and by I on its own, when σ is an IND. That is, if F and I do not interact then the algorithms in database design that use logical implication can be implemented more efficiently than would otherwise be the case (see [LV96]).

We consider the interaction between FDs and proper circular INDs [Imi91] which extend noncircular INDs to allow a restricted form of cycles, called proper cycles. Our main results are to exhibit a necessary condition for no interaction to occur between FDs and proper circular INDs and two novel sufficient conditions for no such interaction to occur, which also happen to be desirable properties in the design of incomplete information databases [LL98]. Our new sufficient conditions are orthogonal to the well-known results in the literature for no interaction to occur between FDs and noncircular INDs [MR92, Section 10.10]. Although a syntactic necessary and sufficient condition for no interaction to occur between FDs and proper circular INDs is evasive, we identify the two ways of preventing such interaction from occurring. Both of them utilise the fact that the *chase* procedure [MR92, AHV95, LL99] can be used to solve the implication problem for FDs and proper circular INDs. The first way to prevent interaction is to find a condition which prevents any application of the chase FD rule during the invocation of the chase. The known results in the database literature prevent interaction using this first approach. The second way to prevent interaction is to find a condition which prevents any old value present in the original database, prior to the computation of the chase, to be equated to another value during the computation of the chase (new values added to the database during the computation of the chase may be equated to other values). Our new results prevent interaction using this second approach.

The layout of the rest of the paper is as follows. In Section 2 we define FDs, INDs and their subclasses, and introduce the chase procedure as a means of testing satisfaction of a set of FDs and INDs. In Section 3 we give a necessary condition for no interaction to occur between FDs and proper circular INDs and present several novel sufficient conditions for such non-interaction. Finally, in Section 4 we give our concluding remarks and discuss the open problem of finding a syntactic condition that is both necessary and sufficient for no interaction to occur.

2 Functional and inclusion dependencies

Herein we present the preliminary concepts needed to obtain our results. These include the definition of FDs and INDs, the subclasses of these dependencies that we consider and the definition of the chase procedure with respect to a set of FDs and INDs.

We use the notation $|S|$ to denote the cardinality of a set S. If S is a subset of T we write $S \subseteq T$ and if S is a proper subset of T we write $S \subset T$. Furthermore, S and T are *incomparable* if $S \not\subseteq T$ and $T \not\subseteq S$. We often denote the singleton $\{A\}$ simply by A, and the union of two sets S, T, i.e. $S \cup T$, simply by ST.

Definition 2.1 (Database schema and database) Let $\mathcal{U}$ be a finite set of attributes. A *relation schema* R is a finite sequence of distinct attributes from $\mathcal{U}$. A *database schema* is a finite set $\mathbf{R} = \{R_1, \ldots, R_n\}$, such that each $R_i \in \mathbf{R}$ is a relation schema and $\bigcup_i R_i = \mathcal{U}$.

We assume a countably infinite domain of values $\mathcal{D}$; without loss of generality we assume that $\mathcal{D}$ is linearly ordered. An R-tuple (or simply a tuple whenever R is understood from context) is a member of the Cartesian product $\mathcal{D} \times \ldots \times \mathcal{D}$ ($|R|$ times).

A *relation* r over R is a finite (possibly empty) set of R-tuples. A *database* d over $\mathbf{R}$ is a family of n relations $\{r_1, \ldots, r_n\}$ such that each $r_i \in d$ is over $R_i \in \mathbf{R}$.

From now on we let $\mathbf{R}$ be a database schema and d be a database over $\mathbf{R}$. Furthermore, we let $r \in d$ be a relation over the relation schema $R \in \mathbf{R}$.

Definition 2.2 (Projection) The *projection* of an R-tuple t onto a set of attributes $Y \subseteq R$, denoted by $t[Y]$ (also called the Y-value of t), is the restriction of t to Y. The projection of a relation r onto Y, denoted as $\pi_Y(r)$, is defined by $\pi_Y(r) = \{t[Y] \mid t \in r\}$.

Definition 2.3 (Functional Dependency) A *functional dependency* (or simply an FD) over a database schema $\mathbf{R}$ is a statement of the form $R{:}X \rightarrow Y$ (or simply $X \rightarrow Y$ whenever R is understood from context), where $R \in \mathbf{R}$ and $X, Y \subseteq R$ are sets of attributes. An FD of the form $R{:}X \rightarrow Y$ is said to be *trivial* if $Y \subseteq X$; it is said to be *p-standard* if $|X| \geq p$ for some natural number $p \geq 1$; when $p = 1$ then we say that $R{:}X \rightarrow Y$ is a standard FD. F is said to be p-standard when all the FDs in F are p-standard.

An FD R:X $\rightarrow$ Y is satisfied in d, denoted by $d \models$ R:X $\rightarrow$ Y, whenever $\forall t_1, t_2 \in r$, if t_1[X] = t_2[X] then t_1[Y] = t_2[Y].

Definition 2.4 (Inclusion Dependency) An *inclusion dependency* (or simply an IND) over a database schema **R** is a statement of the form R_i[X] $\subseteq$ R_j[Y], where $R_i, R_j \in$ **R** and X $\subseteq R_i$, Y $\subseteq R_j$ are sequences of distinct attributes such that |X| = |Y|. An IND is said to be *trivial* if it is of the form R[X] $\subseteq$ R[X], and an IND R[X] $\subseteq$ S[Y] is said to be *p-ary* if |X|$\leq p$ for some natural number $p \geq 1$; when $p = 1$ then we say that R[X] $\subseteq$ S[Y] is a unary IND. I is said to be p-ary when all the INDs in I are p-ary.

An IND R_i[X] $\subseteq$ R_j[Y] over **R** is satisfied in d, denoted by d $\models$ R_i[X] $\subseteq$ R_j[Y], whenever $\pi_{\mathrm{X}}(r_i) \subseteq \pi_{\mathrm{Y}}(r_j)$, where $r_i, r_j \in d$ are the relations over R_i and R_j, respectively.

Definition 2.5 (Noncircular sets of INDs) A set of INDs I over **R** is *circular* if either there exists a nontrivial IND R[X] $\subseteq$ R[Y] $\in$ I, or there exist m distinct relation schemas, $R_1, R_2, R_3, \ldots, R_m$ $\in$ **R**, with m > 1, such that I contains the INDs: $R_1[X_1] \subseteq R_2[Y_2]$, $R_2[X_2] \subseteq R_3[Y_3]$, ..., $R_m[X_m]$ $\subseteq R_1[Y_1]$. A set of INDs I is *noncircular* if it is not circular.

The class of proper circular INDs [Imi91] defined below includes the class of noncircular INDs as a special case.

Definition 2.6 (Proper circular sets of INDs) A set I of INDs over **R** is *proper circular* if it is either noncircular or whenever there exist m distinct relation schemas, $R_1, R_2, R_3, \ldots, R_m \in$ **R**, with $m > 1$, such that I contains the INDs: $R_1[X_1] \subseteq R_2[Y_2]$, $R_2[X_2] \subseteq R_3[Y_3]$, ..., $R_{m-1}[X_{m-1}]$ $\subseteq R_m[Y_m]$, $R_m[X_m] \subseteq R_1[Y_1]$, then for all $i \in \{1, 2, \ldots, m\}$ we have $X_i = Y_i$; such a sequence of INDs is called a *proper cycle*.

We note that it may be the case that a set of INDs, which is not proper circular, may be transformed into a proper circular set by removing from it some INDs which do not alter its closure (see Definition 2.9).

From now on we let F be a set of FDs over **R** and $F_i = \{R_i : \mathrm{X} \rightarrow \mathrm{Y} \in \mathrm{F}\}$, $i \in \{1, \ldots, n\}$, be the set of FDs in F over $R_i \in$ **R**. Furthermore, we let I be a set of INDs over **R** and let Σ = F $\cup$ I.

The following definition originates from [Sci86].

Definition 2.7 (Graph representation of INDs) The graph representation of a set of INDs I over **R** is a directed graph G_I = (N, E), which is constructed as follows. Each relation schema R in **R** has a separate node in N labelled by R; we do not distinguish between nodes and their labels. There is an arc (R, S) $\in$ E if and only if there is a nontrivial IND R[X] $\subseteq$ S[Y] $\in$ I. A relation schema R $\in$ **R** is called a *source* relation schema with respect to I if it has no incoming arcs.

It can easily be verified that there is a path in G_I from R to S if and only if for some IND R[X] $\subseteq$ S[Y] we have I $\models$ R[X] $\subseteq$ S[Y]. Moreover, I is noncircular if and only if G_I is acyclic.

Definition 2.8 (Keys and key-based INDs) A set of attributes X $\subseteq R_i$ is a *superkey* for R_i with respect to F_i if $F_i \models R_i$:X $\rightarrow R_i$ holds; X is a *key* for R_i with respect to F_i if it is a superkey for R_i with respect to F_i and for no proper subset Y $\subset$ X is Y a superkey for R_i with respect to F_i.

A database schema **R** is in *Boyce-Codd Normal Form* (or simply BCNF) with respect to F if for all $R_i \in$ **R**, for all nontrivial FDs R_i : X $\rightarrow$ Y $\in F_i$, X is a superkey for R_i with respect to F_i.

An IND R_i[X] $\subseteq$ R_j[Y] is *key-based* if Y is a key for R_j with respect to F_j.

Definition 2.9 (Logical implication) Σ is *satisfied* in d, denoted by $d \models \Sigma$, if $\forall \sigma \in \Sigma, d \models \sigma$.

Σ *logically implies* an FD or an IND σ, written $\Sigma \models \sigma$, if whenever d is a database over **R** then the following condition is true:

$$\text{if } d \models \Sigma \text{ holds then } d \models \sigma \text{ also holds.}$$

Σ *logically implies* a set Γ of FDs and INDs over **R**, written $\Sigma \models \Gamma$, if $\forall \sigma \in \Gamma, \Sigma \models \sigma$. We let Σ^+, called the *closure* of Σ, denote the set of all FDs and INDs that are logically implied by Σ.

It is well known that Armstrong's axiom system [Arm74] can be used to compute F^+ and that Casanova's et al. axiom system [CA84] can be used to compute I^+. However, when we consider FDs and INDs together computing Σ^+ was shown to be undecidable [Mit83, CV85]. On the other hand, when I is noncircular then Mitchell's axiom system [Mit83] can be used to compute Σ^+ [CK86]. Moreover, in the special case when I is a set of unary INDs then Cosmadakis's et al. axiom system [CKV90] can be used to compute Σ^+.

The *pseudo-transitivity* inference rule for FDs [Ull88, LL99] and the *pullback* inference rule for FDs and INDs [Mit83, CA84], which are utilised below, are stated in the next two lemmas.

Proposition 2.1 If F $\models$ R : X $\rightarrow$ Y and F $\models$ R : WY $\rightarrow$ Z, then F $\models$ R : XW $\rightarrow$ Z. $\square$

Proposition 2.2 If $\Sigma \models$ {R[XY] $\subseteq$ S[WZ], S:W $\rightarrow$ Z} and |X|=|W|, then $\Sigma \models$ R:X $\rightarrow$ Y. $\square$

We next define two subclasses of FDs which have been very useful in characterising desirable properties in the design of incomplete information databases [LL98].

Definition 2.10 (Intersection property) Two nontrivial FDs of the forms R:X $\rightarrow$ A and R:Y $\rightarrow$ A are said to be *incomparable* if X and Y are incomparable.

A set of FDs F over **R** satisfies the *intersection property* if $\forall F_j \in$ F, $\forall$ A $\in R_j$, whenever there exist incomparable FDs, R_j:X $\rightarrow$ A, R_j:Y $\rightarrow$ A $\in F_j^+$, then R_j:X $\cap$ Y $\rightarrow$ A $\in F_j^+$.

Definition 2.11 (Split-freeness property) Two nontrivial FDs of the forms R:XB $\rightarrow$ A and R:YA $\rightarrow$ B are said to be *cyclic*.

A set of FDs F over **R** satisfies the *split-freeness property* if $\forall F_j \in$ F, whenever there exist cyclic FDs, R_j:XB $\rightarrow$ A, YA $\rightarrow$ B $\in F_j^+$, then either R_j:Y $\rightarrow$ B $\in F_j^+$ or R_j:(X $\cap$ Y)A $\rightarrow$ B $\in F_j^+$.

Definition 2.12 (Reduced set of FDs and INDs) The projection of a set of FDs F_i over R_i onto a set of attributes Y $\subseteq R_i$, denoted by F_i[Y], is given by F_i[Y] = {R_i:W $\rightarrow$ Z | R_i:W $\rightarrow$ Z $\in F_i^+$ and WZ $\subseteq$ Y}.

A set of attributes Y $\subseteq R_i$ is said to be *reduced* with respect to R_i and a set of FDs F_i over R_i (or simply reduced with respect to F_i if R_i is understood from context) if F_i[Y] contains only trivial FDs. A set of FDs and INDs Σ = F $\cup$ I is said to be *reduced* if $\forall R_i$[X] $\subseteq R_j$[Y] $\in$ I, Y is reduced with respect to F_j.

Proposition 2.3 It can be decided in polynomial time in the size of Σ whether Σ is reduced or not.

Proof. The condition that Y is reduced with respect to F_j is true if and only if $\forall$ A $\in$ Y, (Y$-$A) $\rightarrow$ A $\notin F_j^+$. The result now follows, since (Y$-$A) $\rightarrow$ A $\notin F_j^+$ can be checked in polynomial time in the size of F_j [BB79]. □

The chase procedure provides us with a very useful algorithm which forces a database to satisfy a set of FDs and INDs.

Definition 2.13 (The chase procedure for INDs) The chase of d with respect to Σ, denoted by CHASE(d, Σ), is the result of applying the following chase rules, namely the FD and the IND rules, to the current state of d as long as possible. (The current state of d prior to the first application of either of the chase rules is its state upon input to the chase procedure.)

FD rule: If R_j:X $\rightarrow$ Y $\in F_j$ and $\exists t_1, t_2 \in r_j$ such that t_1[X] $= t_2$[X] but t_1[Y] $\neq t_2$[Y], then $\forall$A $\in$ Y, change all the occurrences in d of the larger of the values of t_1[A] and t_2[A] to the smaller of the values of t_1[A] and t_2[A].

IND rule: If R_i[X] $\subseteq R_j$[Y] $\in$ I and $\exists t \in r_i$ such that t[X] $\notin \pi_{\mathrm{Y}}(r_j)$, then add a tuple u over R_j to r_j, where u[Y] $= t$[X] and $\forall$A $\in R_j-$Y, u[A] is assigned a new value greater than any other current value occurring in the tuples of the relations in the current state of d.

Given the initial state of d prior to the computation of CHASE(d, Σ) we call the values in the tuples of the relations in d *old values* and call the values which are newly introduced to d during the computation of CHASE(d, Σ) as a result of the IND rule adding tuples to relations in d *new values*.

We observe that there is no loss of generality to consider an FD rule for R_j : X $\rightarrow$ Y as an FD rule for the FDs R_j : X $\rightarrow$ A, with A $\in$ Y$-$X such that t_1[A] $\neq t_2$[A]. We will utilise this observation in proofs which use the chase procedure. We also observe that, in general, if we allow I to be circular then the chase procedure does not always terminate [JK84]. The following theorem is a consequence of results in [MR92, Chapter 10] and [Imi91]. It shows that when I is a set of proper circular INDs, then the chase procedure terminates and satisfies Σ. It also shows that in this case the chase can be decoupled into two distinct stages. At the first stage the IND rule is applied to the current state of d exhaustively and at the second stage the FD rule is applied exhaustively to the current state of d, after the first stage has been computed, terminating with the final result.

Theorem 2.4 Let Σ = F $\cup$ I be a set of FDs and proper circular INDs over a database schema **R**. Then the following three statements are true:

(i) CHASE(d, Σ) $\models \Sigma$.

(ii) CHASE(d, Σ) is identical to CHASE(CHASE(d, I), F) up to renaming of new values.

(iii) CHASE(d, Σ) terminates after a finite number of applications of the IND and FD rules to the current state of d. □

3 Interaction between FDs and INDs

Herein we show that the condition that a set Σ, comprising a set of FDs F and a set of proper circular INDs I, both over **R**, be reduced is a necessary condition for no interaction to occur between F and I but it is not sufficient. Various sufficient conditions for no interaction between F and I are presented; the main new conditions that we present are that if Σ is reduced and F satisfies either the intersection property or the spit-freeness property then F and I do not interact.

We motivate our definition of interaction by an example.

Example 3.1 Let Σ = F $\cup$ I, where F = {R:X $\rightarrow$ Y, S:W $\rightarrow$ Z}, with |X|=|W|, and I = {R[XY] $\subseteq$ S[WZ]}. By Proposition 2.2 we have {S:W $\rightarrow$ Z} $\cup$ I $\models$ R:X $\rightarrow$ Y and trivially we have F $\models$ R:X $\rightarrow$ Y. Therefore, although F $\models$ R:X $\rightarrow$ Y, there is a subset G = {S:W $\rightarrow$ Z} of F such that G $\not\models$ R:X $\rightarrow$ Y but we have that G $\cup$ I $\models$ R:X $\rightarrow$ Y. Thus we consider F and I to interact, since we can derive an FD from a subset of F together with I that cannot be derived from that subset of F on its own.

Definition 3.1 (Interaction between FDs and INDs) A set of FDs F over **R** is said *not to interact* with of set of INDs I over **R**, if

1) for all FDs α over **R**, for all subsets G $\subseteq$ F, G $\cup$ I $\models \alpha$ if and only if G $\models \alpha$, and

2) for all INDs β over **R**, for all subsets J $\subseteq$ I, F $\cup$ J $\models \beta$ if and only if J $\models \beta$.

We observe that our definition of interaction is stricter than the corresponding definition in [MR92, Section 10.10], where the only subsets of F and I that are considered are F and I themselves. Obviously our definition implies that of [MR92] but, in general, their definition does not imply ours as is shown in Example 3.1.

Definition 3.2 (Traced value) Let r_j be the current state of the relation over R_j in d during the computation of CHASE(d, I). A tuple $t \in r_j$ is said to be *traced* from a tuple $u \in r_i$, where r_i is the original state of the relation over R_i in d, if t was inserted into r_j as a result of the following sequence of applications of the IND rule, $m \geq 0$: (1) u_1 was inserted into r_{h1} over R_{h1} as a result of the IND rule for R_i[Z] $\subseteq R_{h1}[Z_1]$, with u[Z] $= u_1[Z_1]$; (2) u_2 was inserted into r_{h2} over R_{h2} as a result of the IND rule for $R_{h1}[Z_2] \subseteq R_{h2}[Z_3]$ with $u_1[Z_2] = u_2[Z_3]$; ...; $(m+1)$ t was inserted into r_j over R_j as a result of the IND rule for $R_{hm}[Z_{2m}] \subseteq R_j$[V] with $u_m[Z_{2m}] = t$[V]. (We do not exclude the case when $i = j$, and when $m = 0$ the sequence of the applications of the IND rule reduces to: t was inserted into r_j over R_j as a result of the IND rule for R_i[Z] $\subseteq R_j$[V], with u[Z] $= t$[V].)

If $t \in r_j$ is traced from some tuple $u \in r_i$ and t[A] $= u$[B], for some A $\in R_j$ and some B $\in R_i$, then we say that the A-value t[A] is *traced* from r_i.

The next lemma is pivotal in proving our main result (recall Definition 2.7 of a source relation schema).

Lemma 3.1 Let d be a database over **R** such that $d \models$ F and let R_i be a source relation schema with respect to a noncircular set of INDs I over **R**. If the FD rule applies an FD R_j:X $\rightarrow$ A (A $\notin$ X) to the current state of the relation $r_j \in d$ over R_j during the computation of CHASE(d, Σ) and either t_1[A] or t_2[A] is traced from the relation $r_i \in d$ over R_i, then either Σ is not reduced or F violates both the intersection property and the split-freeness property.

Proof. We assume without loss of generality that the left-hand side of X $\rightarrow$ A is an appropriate minimal cardinality set, i.e. that there does not exist a proper subset X_1 of X such that $X_1 \rightarrow$ A $\in F_j^+$. We also assume without loss of generality that no relation $r_h \in d$ is redundant in the following sense: r_h is *redundant* if the state of r_j, during the computation of the chase just prior to the application of the FD rule for X $\rightarrow$ A referred to in the statement of the lemma, is not affected when r_h is removed from d before the chase is invoked.

Now, assume for the rest of the proof that t_1[A] is traced from r_i. Consider the first application of the FD rule referred to in the statement of the lemma. By the assumption that t_1[A] is traced from r_i, it follows that there is a path in G_I from R_i to R_j. We prove the lemma by induction on the length of the longest path, say q, in G_I from R_i to any other relation schema in **R**.

(*Basis*): Suppose that $q = 1$; $q = 0$ is not possible since I is acyclic. We prove the lemma by a second induction on the number of times, say k, the FD rule was applied to d prior to the application of the FD rule for R_j:X $\rightarrow$ A, referred to in the statement of the lemma.

(*Basis*): If $k = 0$ then there were no previous applications of the FD rule. It follows that t_1[XA] $\in \pi_{\rm ZB}(r_i)$ for some ZB $\subseteq R_i$, due to the fact that the IND rule does not equate values, t_1[X] = t_2[X] and t_1[A] is traced from r_i. Moreover, since $q = 1$ there must be an IND R_i[ZB] $\subseteq R_j$[XA], which is obtained from I by the projection and permutation inference rule for INDs [CA84]. The result that Σ is not reduced follows, since X $\rightarrow$ A $\in F_j^+$ is a non-trivial FD.

(*Induction*): Assume that the result holds when the minimum number of times the FD rule was applied to d prior to the application of the FD rule for R_j:X $\rightarrow$ A, referred to in the statement of the lemma, is k, where $k \geq 0$; we then need to prove that the result holds when the minimum number of times the FD rule was applied is $k + 1$. Consider the $(k + 1)th$ application of the FD rule, which applies, say the FD R_h:W $\rightarrow$ C $\in F_h$. Now, if this FD rule does not equate t_1[C] and t_2[C], for some C $\in$ X, then the $(k + 1)th$ application is redundant and the result follows by inductive hypothesis. So, the FD rule must have equated t_1[C] and t_2[C], for some C $\in$ X; in addition, neither t_1[C] nor t_2[C] is traced from any relation $r' \in d$ over R', otherwise again the result follows by inductive hypothesis since the statement of the lemma would be satisfied with this FD rule. (We note that if R' is not a source relation schema with respect to I then we can transform d and I in such a manner that R' will satisfy this requirement as in the inductive step below for $q \geq 1$.) We observe that $r_i \neq r_j$, since neither t_1[C] nor t_2[C] is traced from r_i and $d \models$ F.

Now, since $q = 1$ it must be the case that $h = j$ and therefore the $(k + 1)th$ FD rule applies the FD R_j:W $\rightarrow$ C $\in F_j$. A fragment of the state of r_j prior to the application of the $(k + 1)th$ FD rule is shown in Table 1, where X' = X$-$C, W' = W$-$A and REST = schema(R_j)$-$XW; '?' can be any value and the zeros can be replaced by any other domain value.

W'	X'	C	A	REST
?... ?	0...0	1	3	?... ?
0...0	0...0	2	4	?... ?
0...0	?... ?	1	4	?... ?

Table 1: The current state of the relation r_j

We now prove that F violates the split-freeness property, assuming that Σ is reduced. We first claim that there is some nontrivial FD R_j:Z $\rightarrow$ C $\in F_j^+$ for some C $\in$ X, with A $\in$ Z, such that (Z$-$A) $\rightarrow$ C $\notin F_j^+$. Now, since t_1[A] is traced from r_i and $q = 1$, there is an IND R_i[VB] $\subseteq R_j$[UA] in I that caused the insertion of part of the original state of t_1 into a previous state of r_j. Suppose that the said claim is false. Let X$-$U = $\{C_1, C_2, \ldots, C_m\}$. Then, due to the fact that t_1[X] = t_2[X] in r_j and all previous applications of the FD rule must have equated two new values, we have the sequence of FDs: U $\rightarrow C_1 \in F_j^+$, U $C_1 \rightarrow C_2 \in F_j^+$, ..., U $C_1C_2 \ldots C_{m-1} \rightarrow C_m \in F_j^+$. Thus by repeated applications of the pseudo-transitivity rule for FDs (Proposition 2.1) and the union rule for FDs we can deduce that U $\rightarrow$ X$-$U $\in F_j^+$, implying that U $\rightarrow$ X $\in F_j^+$. Therefore, U $\rightarrow$ A $\in F_j^+$, leading to a contradiction of our assumption that Σ is reduced thus proving the claim.

Let X = X'C and Z = Z'A, We claim that $X' \neq Z'$ implying that F violates the split-freeness property. Suppose that the claim is false. Consider the FD rule for R_j:Z $\rightarrow$ C, which must have been applied to a current state of r_j previous to the application of the FD rule for R_j:X $\rightarrow$ A referred to in the statement of the lemma. Now, since t_1[A] $\neq$ t_2[A] there must exist a third tuple t_3, which is distinct from t_1 and t_2, which was involved in equating t_1[C] and t_2[C] via the application of the FD rule for R_j:Z $\rightarrow$ C. Suppose that the tuples involved in this FD rule are t_2 and t_3, with t_2[A] = t_3[A], and that t_2[C] is equated with t_3[C] by this FD rule application, implying that t_1[C] = t_3[C] but t_1[A] $\neq$ t_3[A]. Moreover, $t_1[X'] = t_2[X'] = t_3[X']$, since we have assumed that $X' = Z'$. The result now follows by inductive hypothesis, since the FD rule for R_j:Z $\rightarrow$ C is redundant due to the fact that t_1[A] and t_2[A] can be equated by applying the FD rule for R_j:X $\rightarrow$ A to the tuples t_1 and t_3 without needing to apply the FD rule for R_j:Z $\rightarrow$ C.

We next prove that F violates the intersection property, assuming that Σ is reduced. If A $\notin$ W then by the pseudo-transitivity inference rule R_j:W(X$-$C) $\rightarrow$ A $\in F_j^+$ and by the intersection property R_j:X$-$C $\rightarrow$ A $\in F_j^+$ implying that the left-hand side of R_j:X $\rightarrow$ A is not minimal as assumed; thus the $(k+1)th$ chase rule is redundant implying by inductive hypothesis that the result holds. Otherwise, if A $\in$ W, we claim that there must have been a previous application of the FD rule for an FD R_j:Z $\rightarrow$ C $\in F_j$, with A $\notin$ Z, concluding the result as above by the intersection property.

Now, since t_1[A] $\neq t_2$[A] there must be a third tuple t_3, which is distinct from t_1 and t_2, and was involved in equating t_1[C] and t_2[C] via the application of the FD rule for R_j:W $\rightarrow$ C. Suppose that the tuples involved in this FD rule are t_2 and t_3, with t_2[A] $= t_3$[A], and that t_2[C] is equated with t_3[C] by this FD rule application, implying that t_1[C] $= t_3$[C] but t_1[A] $\neq t_3$[A]. (A symmetric argument is made when t_1 and t_3 are the tuples involved in this FD rule.) Then, prior to the application of the FD rule for R_j:W $\rightarrow$ C, t_3[C] must have been equated with t_1[C] by an FD rule for some FD R_j:Z $\rightarrow$ C $\in F_j$, since as argued earlier neither t_1[C] nor t_2[C] is traced from any relation $r' \in d$. The proof is complete since it must be the case that A $\notin$ Z, otherwise t_3[C] could not have been equated with t_1[C] due to the fact that t_1[A] $\neq t_3$[A] in the state of r_j prior to the application of the $(k+1)th$ FD rule and in all previous states of r_j during the computation of the chase.

(*Induction*): Assume that the result holds when the length of the longest path in G_I from R_i to any other relation schema in **R** is q, where $q \geq 1$; we then need to prove that the result holds when the length of the longest path in G_I from R_i to any other relation schema in **R** is $q+1$.

To conclude the proof we transform the original database d into a database d' as follows. Let I' be the subset of INDs in I such that R_i[U] $\subseteq R_h$[V], for some $R_h \in$ **R** and let d' = CHASE(d, I' $\cup$ F) $-$ $\{r_i'\}$, where r_i' is the current state of r_i in this chase of d. Let u be the tuple in r_i such that t_1 is traced from u, and let u_1 be a tuple in the relation r_h' over R_h in d' such that t_1 is traced from u_1 as a result of an IND $R_i[U'\mathrm{B}] \subseteq R_h[V'B'] \in I'$, which inserted u_1 into r_h' with $u[U'\mathrm{B}] = u_1[V'B']$. It follows that t_1[A] $= u$[B] $= u_1[B']$ and thus t_1[A] is traced from r_h'. Moreover, R_h is a source relation schema with respect to I $-I'$, $d' \models$ F and the current state of the computation of CHASE(d', (I $-I'$) $\cup$ F) $\cup\{r_i\}$ is equal to the current state of the computation of CHASE(d, Σ) prior to the application of the FD rule referred to in the statement of the lemma, due to the fact that I is acyclic, R_i is a source relation schema with respect to I and initially $d \models$ F implying that the state of r_i did not change until this FD rule application. The result now follows by inductive hypothesis, since by the construction of d' and I $-I'$ the length of the longest path in $G_{I-I'}$ from R_h to any other relation schema in **R** is q. In the statement of the lemma we replace d by d', R_i by R_h and I by I $-I'$. $\square$

A counterexample to the lemma, without the condition that F satisfies the intersection and split-freeness properties, is provided below.

Example 3.2 Let **R** $= \{R_1, R_2\}$ be a database schema, with $R_1 = \{B_1, B_2, B_3, \mathrm{A}\}$ and $R_2 = R_1 \cup$ {C}. Also, let $d = \{r_1, r_2\}$ be a database over **R**, with $r_1 = \{< 1, 2, 3, 0 >\}$ and $r_2 = \emptyset$. Finally, let Σ = F $\cup$ I be a set of FDs and noncircular INDs. The set of INDs is given by I $= \{R_1[B_2B_3] \subseteq R_2[B_2B_3], R_1[B_1B_3\mathrm{A}] \subseteq R_2[B_1B_3\mathrm{A}], R_1[B_1B_2\mathrm{A}] \subseteq R_2[B_1B_2\mathrm{A}]\}$. The set of FDs is given by F $= F_2 = \{B_1\mathrm{A} \rightarrow \mathrm{C}, B_2 \rightarrow \mathrm{C}, B_3\mathrm{C} \rightarrow \mathrm{A}\}$. It can be verified that $\Sigma \models R_1[B_2B_3\mathrm{A}] \subseteq R_2[B_2B_3\mathrm{A}]$, since CHASE$(d, \Sigma)$ produces a tuple t in r_2, with $t[B_2B_3\mathrm{A}] = < 2, 3, 0 >$. So there is interaction between F and I but Σ is reduced. It is interesting to note that the closure of Σ is *not* reduced, since F $\models B_2B_3 \rightarrow$ A.

The next counterexample shows that even if the closure of Σ is reduced there may still be interaction between F and I.

Example 3.3 Let $\mathbf{R} = \{R_1, R_2, R_3\}$ be a database schema, with $R_1 = R_2 = \{$A, B, C$\}$ and R_3 $\{$A, $B_1, B_2, C_1, C_2\}$. Also, let $d = \{r_1, r_2, r_3\}$ be a database over $\mathbf{R}$, with $r_1 = \{< 0, 1, 2 >\}$ and $r_2 = r_3 = \emptyset$. Finally, let Σ = F $\cup$ I be a set of FDs and noncircular INDs. The set of INDs is given by I = $\{R_1$[AC] $\subseteq$ R_2[BC], R_1[AC] $\subseteq$ R_3[AB_1], R_1[AC] $\subseteq$ R_3[AB_2], R_1[BC] $\subseteq$ R_3[AB_1], R_1[BC] $\subseteq$ R_3[AB_2]$\}$. The set of FDs is given by F = F_3 = $\{$A $\rightarrow C_1C_2$, $B_1 \rightarrow C_1$, $B_2 \rightarrow C_2$, $C_1C_2 \rightarrow$ A$\}$. It can be verified that $\Sigma \models R_1$[BC] $\subseteq$ R_2[BC], since in CHASE(d, Σ) the values 0 and 1 are equated. So there is interaction between F and I but it can be verified that the closure of Σ is reduced.

In the above example the interaction is due to the IND R_1[AB] $\subseteq$ R_3[AA] being logically implied by Σ (see [Mit83, CA84] for more on INDs with repeated attributes on their right-hand side, which are also called *repeating dependencies*).

We now extend Lemma 3.1 to the case where I is a set of proper circular INDs. Let relation schemas R and S be in the same *cyclic* equivalence class, if in G_I there is a cycle which contains both R and S; if a relation schema R does not participate in any proper cycle then its cyclic equivalence class is a singleton. We construct a directed graph P_I = (N, E), whose nodes represent cyclic equivalence classes of relation schemas and there is an arc $(n_1, n_2) \in$ E if and only if there is an arc in G_I from a relation schema in the cyclic equivalence class represented by n_1 to a relation schema in the cyclic equivalence class represented by n_2. By the definition of a proper circular set of INDs it follows that P_I is acyclic. Let us call a relation schema R a *cyclic-source* if its cyclic equivalence class is a source node in P_I.

Lemma 3.2 Let d be a database over $\mathbf{R}$ such that $d \models$ F and let R_i be a cyclic-source relation schema with respect to a proper circular set of INDs I over $\mathbf{R}$. If the FD rule applies an FD R_j:X $\rightarrow$ A (A $\notin$ X) to the current state of the relation $r_j \in d$ over R_j during the computation of CHASE(d, Σ) and either t_1[A] or t_2[A] is traced from the relation $r_i \in d$ over R_i, then either Σ is not reduced or F violates both the intersection property and the split-freeness property.

Proof. We observe that if R[X] $\subseteq$ S[Y] is an IND in a proper cycle, then there is no other IND in I of the form R[X'] $\subseteq$ S[Y']. This observation, which is crucial to the proof, holds since otherwise we could replace R[X] $\subseteq$S[Y] in the said cycle by R[X'] $\subseteq$ S[Y'] and thus I would no longer be proper circular as assumed.

Case 1. R_i and R_j are in different cyclic equivalence classes. We can now preprocess d as follows. For each cyclic equivalence class, say C_k, which is not a singleton, let I_{C_k} be the set of all INDs in I of the form R[V] $\subseteq$ S[U], where R and S are in C_k. Moreover, let I_C denote the union of all sets of INDs, I_{C_k}, where C_k is a cyclic equivalence class. We now let d' = CHASE(d, I_C $\cup$ F) be our initial database. Finally, we break the proper cycles in I by removing exactly one IND from each such cycle, ensuring that R_i is a source relation schema in the resulting set I' of noncircular INDs. (We enforce this condition by removing the IND R_h[V] $\subseteq$ R_i[U] from proper cycles that involve R_i.) The result now follows by Lemma 3.1; in the statement of the lemma we replace d by d' and I by I $-$ I_C.

Case 2. R_i and R_j are in the same cyclic equivalence class and thus both R_i and R_j are cyclic-source relation schemas. In this case we only need to invoke an induction on the number of times, say k, the FD rule was applied to d prior to the application of the FD rule for R_j:X $\rightarrow$ A referred to in the statement of the lemma. The basis step, for $k = 0$, is essentially the same as in Lemma 3.1; again consider the IND, R_i[ZB] $\subseteq$ R_j[XA], which was responsible for inserting t_1 into r_j. This IND is obtained from the set of INDs in the cyclic equivalence class of R_i by employing the transitivity and the projection and permutation inference rules for INDs [CA84]. The inductive step for $k \geq 0$ follows along the same lines as the inductive step in Lemma 3.1. We note that in this case we can assume without loss of generality that $h = j$, since we are free to choose R_i and R_j to be any of the relation schemas in their equivalence class due to the fact that R_i is a cyclic-source and the observation at the beginning of the proof implying a symmetry between R_i and R_j with respect to the INDs in the cyclic-equivalence class. $\square$

The next theorem represents the most general result, which we are able to obtain, for no interaction to occur between F and I, apart from Theorems 3.5 and 3.6 which are extensions of well-known results.

Theorem 3.3 If I is proper circular, F satisfies either the intersection property or the split-freeness property and Σ is reduced then F and I do not interact.

Proof. We prove the result by contraposition. On assuming that F and I do interact, there are two cases to consider.

Case 1. For some FD R_i:W $\rightarrow$ Z and for some subset G $\subseteq$ F, G $\cup$ I $\models$ R_i:W $\rightarrow$ Z but G $\not\models$ R_i:W $\rightarrow$ Z. We need to show that either Σ is not reduced or F violates both the intersection property and the split-freeness property.

Let Γ = G $\cup$ I. It follows that for some B $\in$ Z, G $\not\models$ R_i:W $\rightarrow$ B. Thus there is some database d, where apart from the relation $r_i \in d$ over R_i all other relations are empty and such that $d \models$ G but $d \not\models$ R_i:W $\rightarrow$ B. Without any loss of generality, we can consider R_i to be a cyclic-source relation schema since due to the construction of d we can eliminate from I all INDs of the form R_k[U] $\subseteq$ R_j[V], where in P_I there is a path from the cyclic equivalence class of R_k to the cyclic equivalence class of R_i, and R_j is in the same cyclic equivalence class as R_i, without affecting the computation of CHASE(d, Σ). Thus the conditions of Lemma 3.2 are now satisfied, since $d \not\models$ R_i:W $\rightarrow$B but CHASE$(d, \Gamma) \models$ R_i:W $\rightarrow$ B; hence the result follows.

Case 2. For some IND R_i[X] $\subseteq$ R_j[Y] and for some subset J $\subseteq$ I, F $\cup$ J $\models$ R_i[X] $\subseteq$ R_j[Y] but J $\not\models$ R_i[X] $\subseteq$ R_j[Y]. Again we need to show that either Σ is not reduced or F violates both the intersection property and the split-freeness property.

Let Γ = F $\cup$ J. Also, let d be a database, where apart from r_i all the relations in d are empty, and let r_i contain a single tuple t_i such that for all distinct attributes A, B $\in$ R_i, t_i[A] and t_i[B] are pairwise distinct values. By the construction of d we have $d \models$ F but $d \not\models$ R_i[X] $\subseteq$ R_j[Y]. As in Case 1 we can consider R_i to be a cyclic-source relation schema without any loss of generality. Thus the conditions of Lemma 3.2 are satisfied, since CHASE(d, J) $\not\models$ R_i[X] $\subseteq$ R_j[Y] but CHASE$(d, \Gamma) \models$ R_i[X] $\subseteq$ R_j[Y]; hence the result follows. $\square$

The next lemma shows that although Σ being reduced is not a sufficient condition it is a necessary one. On the other hand, it is easy to see that neither the intersection property nor the split-freeness property is a necessary condition for no interaction; for example, consider the case when the set I of INDs over **R** is empty.

Lemma 3.4 If F and I do not interact then Σ is reduced.

Proof. We prove the result by contraposition. Assume that Σ is not reduced and thus for some IND $R_i[Z_i] \subseteq R_j[Z_j] \in$ I, Z_j is not reduced with respect to R_j and F_j. It now follows that $F_j[Z_j]$ contains a nontrivial FD, say $R_j : X_j \rightarrow Y_j$, with $X_jY_j \subseteq R_j$. Furthermore, we have that I $\models$ $R_i[X_iY_i] \subseteq R_j[X_jY_j]$ for some subset $X_iY_i \subseteq Z_i$, with $|X_i|$=$|X_j|$, since $X_jY_j \subseteq Z_j$. Therefore, by Proposition 2.2, $\Sigma \models R_i : X_i \rightarrow Y_i$, where $R_i : X_i \rightarrow Y_i$ is a nontrivial FD. The result follows, since $F_j \cup$ I $\models R_i : X_i \rightarrow Y_i$ but $F_j \not\models R_i : X_i \rightarrow Y_i$. $\square$

The next two theorems extend Theorems 10.20 and 10.21 in [MR92] to broader classes of dependencies. Their proofs, which utilise the results of Theorem 2.4 for proper circular INDs, are ommited since they are essentially the same as those in [MR92, Section 10.10].

Theorem 3.5 If F is a set of p-standard FDs and I is a set of proper circular p-ary INDs then F and I do not interact. $\square$

Theorem 3.6 If **R** is in BCNF with respect to F, I is a proper circular set of INDs and Σ = F $\cup$ I is reduced then F and I do not interact. $\Box$

The next theorem, which follows from a straightforward simplification of Lemma 3.1, shows that a stronger syntactic restriction on Σ than it just being reduced would be sufficient for no interaction to occur between F and I.

Theorem 3.7 Let F be a set of FDs over **R** and I be a set of proper circular INDs over **R**. If $\forall R_i$[V] $\subseteq$ R_j[U] $\in$ I and $\forall$ Y $\rightarrow$ Z $\in$ F_j we have Y $\not\subseteq$ U, or $\forall R_i$[V] $\subseteq$ R_j[U] $\in$ I and $\forall$ Y $\rightarrow$ Z $\in$ F_j, with B $\in$ Z$-$Y, we have B $\notin$ U, then there is no interaction between F and I. $\Box$

The next example shows that we cannot extend Theorems 3.3, 3.5 and 3.6 to the case when the set of INDs is circular.

Example 3.4 Consider a database schema **R** = {R}, where R = AB, and a set of FDs and INDs Σ = {R:A $\rightarrow$ B, R[A] $\subseteq$ R[B]} over **R**. It can easily be verified that Σ is reduced and F is a standard set of FDs that satisfies the intersection and split-freeness properties, I is a unary set of INDs and **R** is in BCNF with respect to F. Despite all these conditions being satisfied F and I interact, since $\Sigma \models$ {R:B $\rightarrow$ A, R[B] $\subseteq$ R[A]} [CA84, CKV90].

We close this section with an example showing that we *cannot* combine Theorems 3.3, 3.5, 3.6 and 3.7 to obtain a necessary condition for no interaction to occur between F and I when I is proper circular.

Example 3.5 Let **R** = $\{R_1, R_2\}$ be a database schema, with R_1 = {A, B} and R_2 = {A, B, C, D}. Let Σ = F $\cup$ I be a set of FDs and noncircular INDs, where I = $\{R_1$[AB] $\subseteq$ R_2[AB]} and F = F_2 = {B $\rightarrow$ C, AC $\rightarrow$ B, D $\rightarrow$ B}. It can be verified that F and I have no interaction, despite the fact that F violates both the intersection and split-freeness properties, **R** is not in BCNF with respect to F, F is not 2-standard while I is 2-ary and finally the conditions of Theorem 3.7 is also violated.

We finally note that in Example 3.5 the FD D $\rightarrow$ B is redundant with respect to the chase in the sense that the FD rule can never apply this FD no matter what the state of the initial database is, assuming this initial state satisfies F, since there is no FD in F_2 with D on its right-hand side and, in addition, there is no IND in I which includes D in its right-hand side.

4 Concluding Remarks

A sufficient and necessary condition for no interaction to occur, when I is proper circular, is still an open problem. We have proved in Lemma 3.4 that Σ being reduced is a necessary condition to prevent such interaction but it is not a sufficient condition as is shown in Examples 3.2 and 3.3. Still, we have made definite progress towards a solution to this problem by identifying two new sufficient conditions that are orthognal to previous conditions given in the database literature, i.e. that Σ be reduced and F satisfy either the the intersection property or the split-freeness property. Moreover, we have also extended known results to broader classes of dependencies in Theorems 3.5 and 3.6.

Our main contribution can be viewed via Lemmas 3.1 and 3.2 which imply that, when I is proper circular, interaction can be prevented in one of two ways by adding an extra condition to that of Σ being reduced. The first way is to add a condition that prevents any application of the FD rule during the computation of the chase for a database satisfying F. Examples of such extra conditions are that **R** be in BCNF or that F be p-standard and I be p-ary. The second way is to add a condition that prevents the application of the FD rule only in the case where an old value is equated with another value, i.e. in this case the FD rule is still allowed to equate two new valeus. An example of such a condition is that F satisfy the intersection property or F satisfy the split-freeness property. The problem in characterising no interaction is to find a sufficient and necessary condition which covers both these situations. The condition that Σ be reduced essentially prevents the application of the pullback inference rule for FDs and INDs. On the other hand, this condition does not preclude the logical implication of a repeating dependency such as R[AB] $\subseteq$R[CC], which is the result of another form of interaction (see Example 3.3). We conjecture that together with the conditions we have presented a further syntactic restriction on the set of INDs is needed, in order to obtain a necessary and sufficient conditions for no interaction to occur.

References

[AHV95] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison-Wesley, Reading, Ma., 1995.

[Arm74] W.W. Armstrong. Dependency structures of data base relationships. In *Proceedings of IFIP Congress*, pages 580–583, Stockholm, 1974.

[BB79] C. Beeri and P.A. Bernstein. Computational problems related to the design of normal form relational schemas. *ACM Transactions on Database Systems*, 4:30–59, 1979.

[CA84] M.A. Casanova and J.E. Amaral de Sa. Mapping uninterpreted schemes into entity-relationship diagrams: Two applications to conceptual schema design. *IBM Journal of Research and Development*, 28:82–94, 1984.

[CK86] S.S. Cosmadakis and P.C. Kanellakis. Functional and inclusion dependencies: A graph theoretic approach. In P.C. Kanellakis and F. Preparata, editors, *Advances in Computing Research*, volume 3, pages 163–184. JAI Press, Greenwich, 1986.

[CKV90] S.S. Cosmadakis, P.C. Kanellakis, and M.Y. Vardi. Polynomial-time implication problems for unary inclusion dependencies. *Journal of the ACM*, 37:15–46, 1990.

[CV85] A.K. Chandra and M.Y. Vardi. The implication problem for functional and inclusion dependencies is undecidable. *SIAM Journal on Computing*, 14:671–677, 1985.

[Imi91] T. Imielinski. Abstraction in query processing. *Journal of the ACM*, 38:534–558, 1991.

[JK84] D.S. Johnson and A. Klug. Testing containment of conjunctive queries under functional and inclusion dependencies. *Journal of Computer and System Sciences*, 28:167–189, 1984.

[LL98] M. Levene and G. Loizou. Database design of incomplete relations. *ACM Transactions on Database Systems*, 1998. To appear.

[LL99] M. Levene and G. Loizou. *A Guided Tour of Relational Databases and Beyond*. Springer-Verlag, London, 1999. To appear.

[LV96] M. Levene and M.W. Vincent. Justification for inclusion dependency normal form. Research Note RN/96/53, Department of Computer Science, University College London, 1996.

[Mit83] J.C. Mitchell. The implication problem for functional and inclusion dependencies. *Information and Control*, 56:154–173, 1983.

[MR92] H. Mannila and K.-J. Räihä. *The Design of Relational Databases.* Addison-Wesley, Reading, Ma., 1992.

[Sci86] E. Sciore. Comparing the universal instance and relational data models. In P.C. Kanellakis and F. Preparata, editors, *Advances in Computing Research*, volume 3, pages 139–162. JAI Press, Greenwich, 1986.

[Ull88] J.D. Ullman. *Principles of Database and Knowledge-Base Systems*, volume 1. Computer Science Press, Rockville, Md., 1988.