Available online at www.sciencedirect.com
ScienceDirect
Information and Computation 204 (2006) 1847–1851
Information and Computation
www.elsevier.com/locate/ic

Corrigendum

Corrigendum to "On the undecidability of implications between embedded multivalued database dependencies" [Inform. and Comput. 122 (1995) 221–235]

Christian Herrmann *

Technische Universität Darmstadt, Fachbereich Mathematik, Schlossgartenstrasse 7, D64289 Darmstadt, Germany

Available online 30 October 2006

By an *implication* for database dependencies we mean an expression $H \Rightarrow F$, where H is a conjunction of dependencies and F a single dependency. Fixing a class of such implications, a solution of the (*finite*) *implication problem* consists in an algorithmic procedure deciding for every implication in the class whether or not it holds in all (finite) databases (in which it is to be interpreted). In [3] this problem was studied for dependencies which are functional (fd) or embedded multivalued (emvd). As pointed out by Luc Segoufin, what was really shown is the following.

Theorem 1. *The implication problem and the finite implication problem for implications $H \Rightarrow F$, where F is an emvd and H a conjunction of emvds and fds, are unsolvable.*

The claimed extension to emvds, alone, relied on an elimination of fds from the problem which was attributed to Beeri and Vardi [1, Lemma 4]. However, this reference does not supply the elimination claimed in Theorem 16 of [3]. On the other hand, the arguments given in [3] do not provide a proof—those for Lemma 18 are vacuous. Both facts have been observed by Luc Segoufin. The purpose of this Corrigendum is to provide a proof for this elimination and so for the result stated in [3]:

Theorem 2. *The implication problem and the finite implication problem for emvds are unsolvable.*

The proof of Theorem 2 will be self contained but relying on Theorem 1 and on Theorem 4, below, which recalls the result of Beeri and Vardi [2], crucial for the elimination of fds.

DOI of original article: 10.1006/inco.1995.1148
* Fax: +49 9151 193317.
E-mail address: herrmann@mathematik.tu-darmstadt.de

doi:10.1016/j.ic.2006.09.002

A, B, C will be variables to denote pairwise distinct attributes (also the singleton sets), X, Y variables for sets of attributes and $XY = X \cup Y$. We refer to the relational database model with a single relation I over a finite universe U of attributes (cf. Beeri and Vardi [2]). As such, we call a U-database. By convention, U will always denote a finite set so that it makes sense to consider U the attribute set of a database. Given $t \in I$ and $X \subseteq U$ we denote by $t[X]$ the restriction of t to X. Fds and emvds are written in the form $X \to Y$ and $[X, Y]$, respectively. If $X, Y \subseteq U$, then $[X, Y]$ holds in I if and only if for all $t_1, t_2 \in I$ such that $t_1[X \cap Y] = t_2[X \cap Y]$ there is $t \in I$ such that $t[X] = t_1[X]$ and $t[Y] = t_2[Y]$. By a U-dependency, respectively, implication we mean one with all attributes in U. A U-mvd is an emvd $[X, Y]$ such that $XY = U$.

For conjunctions H and G of dependencies in the attribute set U we say that H *U-implies* G if G holds in all U-databases in which H holds.

Given $U^{\#} \supseteq U$, we say that a U-implication $H \Rightarrow F$ is $U - U^{\#}$-similar to the $U^{\#}$ implication $H^{\#} \Rightarrow F$ provided that $H \Rightarrow F$ holds in all U-databases if and only if $H^{\#} \Rightarrow F$ holds in all $U^{\#}$-databases and provided that the same takes place for finite databases. The following result relies heavily on Beeri and Vardi [2] and has been known to them, yet we were not able to find an explicit reference.

Theorem 3. *With each U-implication $H \Rightarrow F$, where F is an emvd and H a conjunction of fds and emvds, one can effectively associate $U^{\#} \supseteq U$ and a $U - U^{\#}$-similar $U^{\#}$-implication $H^{\#} \Rightarrow F$, where $H^{\#}$ is a conjunction of emvds.*

Actually, $U^{\#}$ will depend on U, only, and $H^{\#}$ will arise from H replacing the fds by conjunctions of $U^{\#}$-mvds.

Proof of Theorem 2. Any decision procedure for the emvd implication problem could be converted into one solving the problem referred to in Theorem 1. Indeed, given an emvd F and conjunction H of emvds and fds, choose U to comprise all attributes in $H \Rightarrow F$, form $U^{\#}$ and $H^{\#}$ according to Theorem 3, and apply the decision procedure. □

It remains to prove Theorem 3. In [2], Beeri and Vardi associate with an fd $B \to A$ over the universe U the conjunction $(B \to 4)^*_U$ of two total tuple generating dependencies (ttgds). Let $C_1, \ldots, C_m$ be a listing of $U - AB$ with no repetitions. Choose pairwise distinct variables for values: a_i, b_i, c_{ij} for $i = 0, 1, 2$ and $j = 1, \ldots, m$. Let the two ttgds be given by Table 1. The first tuple stands for the conclusion, the others for the premise.

By definition, the first ttgd is valid in a relation I over U if the statement described in Table 2 is valid for each map h from the set of variables in the set of values of I. Similarly, for the second ttgd.

Table 1
Translating an fd into ttgds

A	B	C_1	...	C_m	A	B	C_1	...	C_m
a_0	b_1	c_{21}	...	c_{2m}	a_1	b_1	c_{21}	...	c_{2m}
a_0	b_0	c_{01}	...	c_{0m}	a_0	b_0	c_{01}	...	c_{0m}
a_1	b_0	c_{11}	...	c_{1m}	a_1	b_0	c_{11}	...	c_{1m}
a_1	b_1	c_{21}	...	c_{2m}	a_0	b_1	c_{21}	...	c_{2m}

Table 2
Semantics of ttgds

	$(h(a_0), h(b_0), h(c_{01}), \ldots, h(c_{0m})) \in I$
And	$(h(a_1), h(b_0), h(c_{11}), \ldots, h(c_{1m})) \in I$
And	$(h(a_1), h(b_1), h(c_{21}), \ldots, h(c_{2m})) \in I$
Jointly imply	$(h(a_0), h(b_1), h(c_{21}), \ldots, h(c_{2m})) \in I$

Given a conjunction H of U-dependencies, let H^*_U denote the conjunction arising from H if each fd $B \to A$ is replaced by $(B \to A)^*_U$. As a special case of Theorem 7 in Beeri and Vardi [2], one obtains the following.

Theorem 4. *Let F be an emvd and H a conjunction of emvds and fds of the form $B \to A$. Assume that $H \Rightarrow F$ is a U-implication. Then $H \Rightarrow F$ is $U - U$-similar to $H^*_U \Rightarrow F$.*

The idea of proof is now as follows. In Lemma 17 of [3] it was shown that fds can be replaced by mvds if one uses a second copy of the attribute set. The equivalence of an attribute A and its copy $\hat{A}$ can, of course, be captured by the fds $A \to \hat{A}$ and $\hat{A} \to A$. In the context of similarity, the latter can be replaced by ttgds according to Theorem 4. Moreover, the fds imply the ttgds. The key is to interpolate such implications with mvds. In order to do so, we use a third copy of U. A similar technique is used in the coordinatization of lattices and relation algebras and in commutator theory of algebraic structures.

The following auxiliary definitions and results are formulated for all U-databases and pairwise distinct $A, B, C \in U$. We write

$$A \leftrightarrow B := A \to B \wedge B \to A$$
$$A \leftrightarrow B \leftrightarrow C := A \leftrightarrow B \wedge A \leftrightarrow C \wedge B \leftrightarrow C$$

and consider the following conjunctions $\eta_U(A, B, C)$ of U-mvds:

$$[ACD, AB] \wedge [BCD, AB] \wedge [BCD, AC] \quad \text{where } D = U - ABC.$$

Lemma 5. *$A \leftrightarrow B \leftrightarrow C$ U-implies $\eta_U(A, B, C)$.*

Proof. $A \to B$ implies $[AB, ACD]$. Indeed, if $t_1[A] = t_2[A]$, then let $t = t_2$. Then remaining cases follow by symmetry. □

Lemma 6. *$\eta_U(A, B, C)$ U-implies $(A \to B)^*_U$, $(A \to B)^*_U$, $(B \to A)^*_U$, $(A \to C)^*_U$, $(C \to A)^*_U$, $(B \to C)^*_U$, and $(C \to B)^*_U$.*

Proof of $(B \to A)^*_U$ by the tuple generating chases in Table 3. D stands for $U - ABC$, the d_i for corresponding parts of tuples. We name the mvds used and the tuples involved, first the one which is kept except changing the value of a single attribute. The remaining cases follow by symmetry. □

Preparing for the proof of Theorem 3, fix a countably infinite set U_∞ (think of its members as possible original attributes). Let $\hat{U}_\infty$ and $\tilde{U}_\infty$ be disjoint copies of U_∞ (providing the copy attributes) and $A \mapsto \hat{A}$ and $A \mapsto \tilde{A}$ bijections from U_∞ onto $\hat{U}_\infty$ and $\tilde{U}_\infty$, respectively. For $U \subseteq U_\infty$ let

Table 3
Tuple generating chases

	A	B	C	D	
1	a_0	b_0	c_0	d_0	
2	a_1	b_0	c_1	d_1	
3	a_1	b_1	c_2	d_2	
4	a_1	b_0	c_2	d_2	3,2, $[ACD, AB]$
5	a_0	b_0	c_2	d_2	4,1, $[BCD, AB]$
	a_0	b_1	c_2	d_2	3,5, $[BCD, AC]$
	A	B	C	D	
1	a_0	b_0	c_0	d_0	
2	a_1	b_0	c_1	d_1	
3	a_0	b_1	c_2	d_2	
4	a_0	b_0	c_2	d_2	3,1, $[ACD, AB]$
5	a_1	b_0	c_2	d_2	4,2, $[BCD, AB]$
	a_1	b_1	c_2	d_2	5,3, $[BCD, AC]$

$\hat{U}$ and $\tilde{U}$ denote the images under these maps and $U' = U \cup \hat{U} \cup \tilde{U}$. Let I_U be the conjunction of all $A \leftrightarrow \hat{A} \leftrightarrow \tilde{A}, A \in U$. For an fd $X \to A$, let $(X \to A)'_U$ be defined as the U'-mvd $[U' - A, XA]$.

Lemma 7. *Let F be an emvd and H a conjunction of emvds and fds of the form $X \to A$ with A not in X. Assume that $H \Rightarrow F$ is a U-implication and let H' arise from H by replacing $X \to A$ with $(X \to A)'_U$ and adding the conjunct I_U. Then $H \Rightarrow F$ and $H' \Rightarrow F$ are $U - U'$-similar.*

This is basically Lemma 17 in [3]. **Proof.** The dependencies $X \to A$ and $(X \to A)'_U$ are equivalent for all U'-databases which satisfy I_U. Namely, consider a U'-model J' of I_U and $[U' - A, XA]$ and $t, u \in J'$ such that $t[X] = u[X]$. By the mvd one has $w \in J'$ such that

$$w[XA] = u[XA], \quad w[U' - A] = t[U' - A].$$

In particular, $w(A) = u(A)$ and $w(\hat{A}) = t(\hat{A})$, whence $w(A) = t(A)$ by I_U and $t(A) = u(A)$. The converse (that the fd implies the mvd) is trivial.

Now, given a U-model J of H, choose the domains for the new attributes such that for each $A \in U$ there are bijections

$$\phi_A : \mathrm{DOM}(A) \to \mathrm{DOM}(\hat{A}), \quad \psi_A : \mathrm{DOM}(A) \to \mathrm{DOM}(\hat{A})$$

and define the U' database J' to consist of all t such that

$$t[U] \in J, \quad t(\hat{A}) = \phi_A(t(A)), \quad t(\tilde{A}) = \psi_A(t(A)) \text{ for all } A \in U.$$

Then J' is a model of H'. Conversely, from a U'-model J' of H' pass to J just by restricting J' to U to obtain a model of H. In both directions, the status of the emvd F remains unchanged. $\square$

Proof of Theorem 3. We may assume that the fds in H are of the form $X \to A$ with A not in X, omitting the trivial ones. Form U' and H' according to Lemma 7. Let $U^{\#} = U'$ and $H^{\#}$ be the conjunction of emvds which arises from H' replacing $A \leftrightarrow \hat{A} \leftrightarrow \tilde{A}$ by $\eta_{U'}(A, \hat{A}, \tilde{A})$. In view of Lemma 7 it suffices to show that $H' \Rightarrow F$ and $H^{\#} \Rightarrow F$ are $U' - U'$-similar.

Now, applying Lemma 5 to the attribute set U', we have that $H' U'$-implies $H^{\#}$. Hence, $H^{\#} \Rightarrow FU'$-implies $H' \Rightarrow F$. In particular, if $H^{\#} \Rightarrow F$ holds for all (finite) U'-databases, then so does $H' \Rightarrow F$.

To prove the converse, assume there is a U'-model $J^{\#}$ of $H^{\#}$ which is not a model of F. Since $H^{\#}$ arises from $H'^{*}_{U'}$, replacing $(A \leftrightarrow \hat{A} \leftrightarrow \tilde{A})^{*}_{U'}$, with $\eta_{U'}(A, \hat{A}, \tilde{A})$, from Lemma 6 we have that $H^{\#} U'$-implies $H'^{*}_{U'}$. It follows that $J^{\#}$ is also a model of $H'^{*}_{U'}$. Now, by Theorem 4 there is a U'-model J' of H' which is not a model of F. And J' can be chosen finite if $J^{\#}$ is finite. $\square$

The author must thank Luc Segoufin and Moshe Vardi for pointing out the error and for support in the repair effort.

References

[1] C. Beeri, M.Y. Vardi, The implication problem for data dependencies, in: S. Even, O. Kariv (Eds.), ICALP 81, LNCS, vol. 115, Springer, Berlin, 1981, pp. 73–85.
[2] C. Beeri, M.Y. Vardi, A proof procedure for data dependencies, JACM 31 (1984) 718–741.
[3] C. Herrmann, On the undecidability of implications between embedded multivalued database dependencies, Inform. and Comput. 122 (1995) 221–235.