CORRIGENDUM: ON THE UNDECIDABILITY OF IMPLICATIONS BETWEEN EMBEDDED MULTIVALUED DATABASE DEPENDENCIES

CHRISTIAN HERRMANN

ABSTRACT. A gap is filled which has occurred in the proof of the result mentioned in the title.

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 has been 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 [2], Lemma 4. However, this reference does not supply the elimination claimed in Thm.16 of [3]. On the other hand, the arguments given in the paper do not provide a proof - those for Lemma 18 are vacuous. Both facts have been observed by Luc Segoufin. The purpose of the present note 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 unsolavble.*

The proof of Thm.2 will be self contained but relying on Thm.1 and on Thm.4, below, which recalls the result of Beeri and Vardi [2], crucial for the elimination of fds. In the final paragraphs we will indicate how to verify Thm.1 from [3].

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]). Such, we call a U-database. By convention, U will always denote a finite set so that it makes sense to consider U as 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 $[X, Y]$, resp.. 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 resp. implication we mean one with all attributes in U. A U-mvd is a U-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 explicite 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 Thm.2. Any decision procedure for the emvd implication problem could be converted into one solving the problem referred to in Thm.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 Thm.3, and apply the decision procedure. □.

It remains to prove Thm.3. In [2], Beeri and Vardi associate with an fd $B \to A$ over the universe U the conjunction $(B \to A)^*_U$ of two total tuple generating dependencies (ttgds). Let $C_1, \ldots, C_m$ be a listing of $U - AB$ with not 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 the following tableaux - the first tuple stands for the conclusion, the others for the premise

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

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

By definition, the first ttgd is valid in a relation I over U if for each map h form the set of variables into the set of values of I the following holds:

$$\begin{array}{rl} & (h(a_0), h(b_0), h(c_{01}), \dots, h(c_{0m})) \in I \\ \text{and} & (h(a_1), h(b_0), h(c_{11}), \dots, h(c_{1m})) \in I \\ \text{and} & (h(a_1), h(b_1), h(c_{21}), \dots, h(c_{2m})) \in I \\ \text{jointly imply} & (h(a_0), h(b_1), h(c_{21}), \dots, h(c_{2m})) \in I \end{array}$$

Similarly, for the second ttgd. 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 Thm.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 $B \to A$. Asume 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, by 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 Thm.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 coordianatization 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 following chases. D stands for $U - ABC$, the d_i for correspondig 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.

	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	$5, 3, [BCD, AC]$

The remaining cases follow by symmetry. □

Preparing for the proof of Thm.3, fix a countably infinite set U_∞ (think of its members as possible original attributes). Let $\hat{U}_\infty$ and $\bar{U}_\infty$ be disjoint copies of U_∞ (providing the copy attributes) and $A \mapsto \hat{A}$ and $A \mapsto \bar{A}$ bijections from U_∞ onto $\hat{U}_\infty$ and $\bar{U}_\infty$, respectively. For $U \subseteq U_\infty$ let $\hat{U}$ and $\bar{U}$ denote the images under these maps and $U' = U \cup \hat{U} \cup \bar{U}$. Let I_U be the conjunction of all

$$A \leftrightarrow \hat{A} \leftrightarrow \bar{A}, \quad A \in U.$$

For an fd $X \to A$ let $(X \to A)'_U$ 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

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

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

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

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. □

Proof of Thm.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 $A \leftrightarrow \hat{A} \leftrightarrow \bar{A}$ by $\eta_{U'}(A.\hat{A}, \bar{A})$. 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, $H^\# \Rightarrow F$ U'-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 $(A \leftrightarrow \hat{A} \leftrightarrow \bar{A})^*_{U'}$ with $\eta_{U'}(A, \hat{A}, \bar{A})$, from Lemma 6 we have that $H^\#$ U'-implies $H'^*_{U'}$. It follows, that $J^\#$ is also a model of $H'^*_{U'}$. Now, by Thm.4 there is a U'-model J' of H' which is not a model of F. And J' can be choosen finite if $J^\#$ is finite. □

In the sequel, we assume the definitions of [3] and all results from Prop.4 up to Thm.12 and Prop.19 up to Thm.33. Then one can read the proof of Thm.1 in [3] as a proof of Thm.1 of the present note: one just has to omit the sentence refering to the 'crucial Thm.16'.

Moreover, we obtain a proof of Thm.16 in [3] which stated that, fixing the class of all (finite) databases, there is an interpretation of implications $H \Rightarrow F$ with F an emvd and H a conjunction of fds and emvds into emvd implications and vice versa. The nontrivial interpretation is given by Thm.3, the converse by containement.

Corollary 8. *The implication problem and the finite implication problem for implications $H \Rightarrow F$, where F is an fd and H a conjunction of fds and emvds, are unsolvable.*

Proof. Let $L_{\Pi f}$ be defined the same way as $L_{\Pi s}$ with the only difference that F_1 has the form of an inclusion $\alpha \subseteq \beta$. Of course, Thm.12 is valid with $L_{\Pi f}$ in place of $L_{\Pi s}$. The same applies to Thm.22 and Cor.24. In the proof one has to choose F' as $\rho_{x_0 12} \subseteq \rho_{y_0 12}$ to obtain an equivalent for $r(x_0) = r(y_0)$. This is immediate from Lemma 20. Again, Thm.33 is valid with $L_{\Pi f}$ in place of $L_{\Pi s}$ (and the very same proof). Now, as above, the proof of Thm.1 in [3] yields the variant of Thm.1 where F is an fd. □

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

References

[1] Beeri,C and Vardi,M.Y. (1981), The implication problem for data dependencies, in "ICALP 81" (S.Even and O.Kariv, Eds.), LNCS **115**, pp. 73-85, Springer, Berlin

[2] Beeri,C and Vardi,M.Y. (1984), A proof procedure for data dependencies, JACM **31**, 718-741

[3] Herrmann,C. (1995), On the undecidability of implications between embedded multivalued database dependencies, Information and Computation **122**, 221-235

FB4 AG1 TU Darmstadt
D64289 Darmstadt, Germany