Appropriate Reasoning about Data Dependencies in Fixed and Undetermined Universes

— On Leaning Theories of Functional and Multivalued Dependencies —

Joachim Biskup and **Sebastian Link**

University of Dortmund
Germany

Massey University
New Zealand

This research is supported by Marsden Funding, Royal Society of New Zealand.

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Outline

Motivation: Efficient Updates and Querying

Criticism: Desirable Properties of Inference Systems

Some Leaning Inference Systems for FDs and MVDs

Appropriate Reasoning in Fixed Universes

An Axiomatisation in Undetermined Universes

Extensions to Full Hierarchical Dependencies

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Motivation: A Design Problem

- manufacturers supply articles from a location at a certain cost
- *Article*, *Manufacturer*, *Location*, and *Costs*
- set Σ of semantic constraints:
  - $Article \longrightarrow Manufacturer$
  - $Article,\ Location \longrightarrow Costs$
  - $Manufacturer \twoheadrightarrow Location$
- target database is supposed to
  - process efficiently updates on *Costs* based on the *Article*, e.g.,

$$\pi_{\text{Costs}}(\sigma_{\text{Article=MP3-Player}}(\{Article, Costs\}))$$

  - efficiently process queries about *Article*,*Location*-information of manufacturers such as

$$\pi_{\text{Article, Location}}(\sigma_{\text{Manufacturer=Sony}}(\{Article, Manufacturer, Location\}))$$

- good design that meets these criteria?

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Motivation: A Design Solution

- the FD $Article \to Costs$ is implied by Σ

- as reasonable semantic constraint it results in decomposition

$$\{Article, Costs\} \text{ and } \{Article, Manufacturer, Location\}$$

- efficient updates on first schema (no data redundancy)

- efficient query processing on second schema (no joining necessary)

- efficient solutions to implication problem unlock good design choices

- inference systems to represent properties of constraints appropriately

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Reminder of Definitions

- FD $X \rightarrow Y$ on relation schema R satisfied by R-relation r iff for all $t_1, t_2 \in r$: if $t_1[X] = t_2[X]$, then $t_1[Y] = t_2[Y]$
- MVD $X \twoheadrightarrow Y$ on relation schema R satisfied by R-relation r iff for all $t_1, t_2 \in r$: if $t_1[X] = t_2[X]$, then there is some $t \in r$ such that $t[XY] = t_1[XY]$ and $t[X(R - XY)] = t_2[X(R - XY)]$
- if r satisfies $X \rightarrow Y$, then $r = r[XY] \bowtie r[X(R - XY)]$
- r satisfies $X \twoheadrightarrow Y$ if and only if $r = r[XY] \bowtie r[X(R - XY)]$
- $\Sigma \cup \{\varphi\}$ set of FDs and MVDs on R: Σ *R-implies* φ iff every R-relation r that satisfies all $\sigma \in \Sigma$ also satisfies φ

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Example

- consider relation r over {Article, Manufacturer, Location, Costs}:

Article	Manufacturer	Location	Costs
MP3-Player	Sony	Singapore	500
Camera	Sony	Hong Kong	800
Camera	Sony	Singapore	800
MP3-Player	Sony	Hong Kong	500

- r violates $Manufacturer \rightarrow Location$
- r satisfies $Manufacturer \twoheadrightarrow Location$
- r is the lossless join of the following two projections:

Manufacturer	Location
Sony	Singapore
Sony	Hong Kong

Article	Manufacturer	Costs
MP3-Player	Sony	500
Camera	Sony	800

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Sound Inference Rules under Discussion

$$\frac{}{X \to Y}\, Y \subseteq X$$
(reflexivity, $\mathcal{R}_{\mathrm{F}}$)

$$\frac{X \to Y}{X \to XY}$$
(extension, $\mathcal{E}_{\mathrm{F}}$)

$$\frac{X \to Y, Y \to Z}{X \to Z}$$
(transitivity, $\mathcal{T}_{\mathrm{F}}$)

$$\frac{X \twoheadrightarrow Y}{XU \twoheadrightarrow YV}\, V \subseteq U$$
(augmentation, $\mathcal{A}_{\mathrm{M}}$)

$$\frac{X \twoheadrightarrow Y}{X \twoheadrightarrow R - Y}$$
(R-complementation, $\mathcal{C}^{R}_{\mathrm{M}}$)

$$\frac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow Z - Y}$$
(pseudo-transitivity, $\mathcal{T}_{\mathrm{M}}$)

$$\frac{X \twoheadrightarrow Y, W \twoheadrightarrow Z}{X \twoheadrightarrow Y \cap Z}\, Y \cap W = \emptyset$$
(subset, $\mathcal{S}_{\mathrm{M}}$)

$$\frac{X \twoheadrightarrow Y, Y \twoheadrightarrow Z}{X \twoheadrightarrow YZ}$$
(additive transitivity, $\mathcal{T}^{*}_{\mathrm{M}}$)

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

$$\frac{X \twoheadrightarrow Y, Y \to Z}{X \to Z - Y}$$
(mixed pseudo-transitivity, $\mathcal{T}_{\mathrm{FM}}$)

$$\frac{X \twoheadrightarrow Y, W \to Z}{X \to Y \cap Z}\, Y \cap W = \emptyset$$
(mixed subset, $\mathcal{S}_{\mathrm{FM}}$)

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Complete, but not Complementary

- $\mathfrak{F}_{\mathcal{C}} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{T}_{\mathrm{FM}}, \mathcal{C}^{\mathrm{R}}_{\mathrm{M}}\}$ is R-complete for all R
- $R = ABCD$ with A = Movie, B = Actor, C = Title, D = YearBorn
- inference of $A \twoheadrightarrow B, C$ from $\Sigma = \{A \twoheadrightarrow B, A \twoheadrightarrow C\}$ using $\mathfrak{F}_{\mathcal{C}}$

$$\cfrac{\cfrac{\cfrac{\cfrac{A \twoheadrightarrow B}{A \twoheadrightarrow A, B}\,\mathcal{A}_{\mathrm{M}} \qquad \cfrac{\cfrac{A \twoheadrightarrow C}{A \twoheadrightarrow A, B, D}\,\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}}{A, B \twoheadrightarrow A, B, D}\,\mathcal{A}_{\mathrm{M}}}{A \twoheadrightarrow D}\,\mathcal{T}_{\mathrm{M}}}{A \twoheadrightarrow A, D}\,\mathcal{A}_{\mathrm{M}}}{A \twoheadrightarrow B, C}\,\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}$$

- some MVD-inferences by $\mathfrak{F}_{\mathcal{C}}$ require $\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}$ not just in last step

Complementarity

- Biskup'80: $\mathfrak{C}_{\mathcal{C}} = \mathfrak{F}_{\mathcal{C}} \cup \{\mathcal{S}_{\mathrm{M}}, \mathcal{T}^*_{\mathrm{M}}\}$ is R-complementary for all R
- $R = ABCD$ with $A =$ Movie, $B =$ Actor, $C =$ Title, $D =$ YearBorn
- inference of $A \twoheadrightarrow B, C$ from $\Sigma = \{A \twoheadrightarrow B, A \twoheadrightarrow C\}$ using $\mathfrak{C}_{\mathcal{C}}$

$$\dfrac{\dfrac{\dfrac{}{A \to A}\mathcal{R}_{\mathrm{F}}}{A \twoheadrightarrow A}\mathcal{I}_{\mathrm{FM}} \qquad \dfrac{\dfrac{A \twoheadrightarrow B}{A \twoheadrightarrow A, B}\mathcal{A}_{\mathrm{M}} \qquad \dfrac{A \twoheadrightarrow C}{A, B \twoheadrightarrow C}\mathcal{A}_{\mathrm{M}}}{A \twoheadrightarrow A, B, C}\mathcal{T}^*_{\mathrm{M}}}{A \twoheadrightarrow B, C}\mathcal{T}_{\mathrm{M}}$$

- every MVD-inference by $\mathfrak{C}_{\mathcal{C}}$ requires $\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}$ in at most last step

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Complete and Complementary, but not Adequate

- $\mathfrak{C}_{\mathcal{C}} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{S}_{\mathrm{M}}, \mathcal{T}^{*}_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{T}_{\mathrm{FM}}, \mathcal{C}^{\mathrm{R}}_{\mathrm{M}}\}$
- $R = ABC$:
  inference of $A \to B$ from $\Sigma = \{A \twoheadrightarrow B, C \to B\}$ using $\mathfrak{C}_{\mathcal{C}}$

$$\dfrac{\dfrac{\dfrac{A \twoheadrightarrow B}{A \twoheadrightarrow A, B}\,\mathcal{A}_{\mathrm{M}}}{A \twoheadrightarrow C}\,\mathcal{C}^{\mathrm{R}}_{\mathrm{M}} \qquad C \to B}{A \to B}\,\mathcal{T}_{\mathrm{FM}}$$

- some FD-inferences by $\mathfrak{C}_{\mathcal{C}}$ require $\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}$

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Appropriate: Complementary and Adequate

- $\mathfrak{A}\mathfrak{C}_{\mathcal{C}} = (\mathfrak{C}_{\mathcal{C}} - \{\mathcal{T}_{\mathrm{FM}}\}) \cup \{\mathcal{S}_{\mathrm{FM}}\}$

- $R = ABC$:
  inference of $A \to B$ from $\Sigma = \{A \twoheadrightarrow B, C \to B\}$ using $\mathfrak{A}\mathfrak{C}_{\mathcal{C}}$

$$\frac{A \twoheadrightarrow B \qquad C \to B}{A \to C}$$

- no FD-inference by $\mathfrak{A}\mathfrak{C}_{\mathcal{C}}$ requires $\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}$

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Definitions of Complementarity and Adequacy

- let $\mathfrak{S}$ contain some inference rules under discussion and $\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}$
- Biskup'80:
  R-complete set $\mathfrak{S}$ is said to be *R-complementary* if and only if for every set Σ of FDs and MVDs on R and every MVD φ on R such that φ is R-implied by Σ there is an inference of φ from Σ by $\mathfrak{S}$ in which the R-complementation rule $\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}$ is applied at most once, and if it is applied, then it is applied in the last step of the inference
- Biskup/Link'08:
  R-complete set $\mathfrak{S}$ is said to be *R-adequate* if and only if if for every set Σ of FDs and MVDs on R and every FD φ on R such that φ is R-implied by Σ there is an inference of φ from Σ by $\mathfrak{S}$, i.e., an inference in which the R-complementation rule $\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}$ is not utilised at all

Main Results for FDs and MVDs

$$\mathfrak{A}\mathfrak{C}_{\mathcal{C}} = \mathfrak{A}_{\mathcal{C}} \cup \{\mathcal{S}_{\mathrm{M}}, \mathcal{T}^{*}_{\mathrm{M}}\} = (\mathfrak{C}_{\mathcal{C}} - \{\mathcal{T}_{\mathrm{FM}}\}) \cup \{\mathcal{S}_{\mathrm{FM}}\}$$

appropriate: adequate and complementary

$$\mathfrak{A}_{\mathcal{C}} = (\mathfrak{F}_{\mathcal{C}} - \{\mathcal{T}_{\mathrm{FM}}\}) \cup \{\mathcal{S}_{\mathrm{FM}}\}$$

adequate

$$\mathfrak{C}_{\mathcal{C}} = \mathfrak{F}_{\mathcal{C}} \cup \{\mathcal{S}_{\mathrm{M}}, \mathcal{T}^{*}_{\mathrm{M}}\}$$

complementary

$$\mathfrak{F}_{\mathcal{C}} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{T}_{\mathrm{FM}}, \mathcal{C}^{\mathrm{R}}_{\mathrm{M}}\}$$

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Inadequacy of $\mathfrak{C}_{\mathcal{C}}$ and $\mathfrak{F}_{\mathcal{C}}$

- $\mathfrak{C} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{S}_{\mathrm{M}}, \mathcal{T}^*_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{T}_{\mathrm{FM}}\}$
- $\Sigma = \{\emptyset \twoheadrightarrow A, B \rightarrow A\}$ and $\varphi = \emptyset \rightarrow A$
- closure of Σ under inferences with $\mathfrak{C}$ (no trivial FDs/MVDs):

$\rightarrow$	$\emptyset$	A	B	AB
$\emptyset$	×			
A	×	×		
B	×	×	×	×
AB	×	×	×	×

$\twoheadrightarrow$	$\emptyset$	A	B	AB
$\emptyset$	×	×		
A	×	×		
B	×	×	×	×
AB	×	×	×	×

- $\varphi \notin \Sigma^+_{\mathfrak{C}}$ but $\varphi \in \Sigma^+_{\mathfrak{C}\cup\{\mathcal{S}_{\mathrm{FM}}\}}$
- there is a relation schema R, a set Σ of FDs and MVDs on R and an FD φ on R such that $\varphi \in \Sigma^+_{\mathfrak{C}_{\mathcal{C}}}$ but $\varphi \notin \Sigma^+_{\mathfrak{C}}$
- systems $\mathfrak{F}_{\mathcal{C}}$ and $\mathfrak{C}_{\mathcal{C}}$ are not R-adequate for some R

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Appropriateness of $\mathfrak{AC}_{\mathcal{C}}$

- $\mathfrak{A}_{\mathcal{C}} = \{\mathcal{R}_{\text{F}}, \mathcal{E}_{\text{F}}, \mathcal{T}_{\text{F}}, \mathcal{A}_{\text{M}}, \mathcal{T}_{\text{M}}, \mathcal{I}_{\text{FM}}, \mathcal{S}_{\text{FM}}, \mathcal{C}^{\text{R}}_{\text{M}}\}$ is R-complete for all R
- Let R be some relation schema, and let Σ be a set of FDs and MVDs on R. For every inference γ from Σ by the system $\mathfrak{A}_{\mathcal{C}}$ there is an inference ξ from Σ by the system $\mathfrak{AC}_{\mathcal{C}} = \mathfrak{A}_{\mathcal{C}} \cup \{\mathcal{S}_{\text{M}}, \mathcal{T}^{*}_{\text{M}}\}$ with:

(i) if γ infers an MVD, then

- γ and ξ infer the same MVD,
- in ξ the R-complementation rule $\mathcal{C}^{\text{R}}_{\text{M}}$ is applied at most once, and
- if $\mathcal{C}^{\text{R}}_{\text{M}}$ is applied in ξ, then $\mathcal{C}^{\text{R}}_{\text{M}}$ is applied as the last rule.

(ii) if γ infers an FD, then

- γ and ξ infer the same FD, and
- in ξ the R-complementation rule $\mathcal{C}^{\text{R}}_{\text{M}}$ is not applied at all.

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Example Transformation into an Appropriate Inference

- inappropriate inference step:

$$\dfrac{\dfrac{X \twoheadrightarrow Y}{\mathcal{C}^{\mathrm{R}}_{\mathrm{M}}:\ X \twoheadrightarrow R-Y} \qquad W \rightarrow Z}{\mathcal{S}_{\mathrm{FM}}:\ X \rightarrow \underbrace{(R-Y)\cap Z}_{=Z-Y}}\ (R-Y)\cap W=\emptyset.$$

- since $(R-Y)\cap W=\emptyset$ holds we have $W \subseteq Y$; replacement:

$$\dfrac{\dfrac{X \twoheadrightarrow Y \qquad \dfrac{\dfrac{\dfrac{}{\mathcal{R}_{\mathrm{F}}:\ Y \rightarrow W}\ W\subseteq Y \qquad W \rightarrow Z}{\mathcal{T}_{\mathrm{F}}:\ Y \rightarrow Z}}{\mathcal{I}_{\mathrm{FM}}:\ Y \twoheadrightarrow Z}}{\mathcal{T}_{\mathrm{M}}:\ X \twoheadrightarrow Z-Y} \qquad \dfrac{\dfrac{}{\mathcal{R}_{\mathrm{F}}:\ Y \rightarrow W}\ W\subseteq Y \qquad W \rightarrow Z}{\mathcal{T}_{\mathrm{F}}:\ Y \rightarrow Z}}{\mathcal{T}_{\mathrm{FM}}:\ X \rightarrow \underbrace{(Z-Y)\cap Z}_{=Z-Y}}$$

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Nearly Complete Reasoning in Fixed Universes

- $\mathfrak{AC} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{S}_{\mathrm{M}}, \mathcal{T}^{*}_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{S}_{\mathrm{FM}}\}$

- Let $\Sigma \cup \{\varphi\}$ be a finite set of FDs and MVDs with $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \subseteq R$. Then

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

  - If φ denotes the MVD $X \twoheadrightarrow Y$, then: $X \twoheadrightarrow Y \in \Sigma^{+}_{\mathfrak{AC}_{\mathcal{C}}}$ if and only if $X \twoheadrightarrow Y \in \Sigma^{+}_{\mathfrak{AC}}$ or $X \twoheadrightarrow (R - Y) \in \Sigma^{+}_{\mathfrak{AC}}$.

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Implication in Undetermined Universes

- Let $\Sigma \cup \{\varphi\}$ be a set of FDs and MVDs. We say that Σ *implies* φ if and only if every relation r satisfies the following condition: if $\cup_{\sigma\in\Sigma} Attr(\sigma) \cup Attr(\varphi) \subseteq Dom(r)$ and r satisfies all $\sigma \in \Sigma$, then r also satisfies φ.
- FILM={Movie,Director,Actor} and

$$\Sigma = \{\text{Movie} \twoheadrightarrow \text{Director}, \text{Actor} \rightarrow \text{Director}\}$$

- Σ FILM-implies Movie $\twoheadrightarrow$ Actor but does not imply Movie $\twoheadrightarrow$ Actor

Movie	Director	Actor	Year_born
La Dolce Vita	Frederico Fellini	M. Mastroianni	1924
La Dolce Vita	Frederico Fellini	A. Ekberg	1931

- Let Σ be a set of functional dependencies. Then Σ implies the MVD $X \twoheadrightarrow Y$ if and only if Σ implies the FD $X \rightarrow Y$.

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

First ever combined Axiomatisation of FDs and MVDs in Undetermined Universes

- $\mathfrak{AC} = \{\mathcal{R}_{\mathrm{F}}, \mathcal{E}_{\mathrm{F}}, \mathcal{T}_{\mathrm{F}}, \mathcal{A}_{\mathrm{M}}, \mathcal{T}_{\mathrm{M}}, \mathcal{S}_{\mathrm{M}}, \mathcal{T}^*_{\mathrm{M}}, \mathcal{I}_{\mathrm{FM}}, \mathcal{S}_{\mathrm{FM}}\}$ is sound and complete for FD and MVD implication in undetermined universes
- completeness of $\mathfrak{AC}$: assume $\varphi \notin \Sigma^+_{\mathfrak{AC}}$
  - $T := \cup_{\sigma \in \Sigma} Attr(\sigma) \cup Attr(\varphi) \subseteq Dom(r)$
  - let $R \subseteq \mathcal{A}$ be finite proper superset of R ($R - rhs(\varphi) \not\subseteq T$)
  - if φ FD, then $\varphi \notin \Sigma^+_{\mathfrak{AC}_{\mathcal{C}}}$, but R-completeness of $\mathfrak{AC}_{\mathcal{C}}$ shows that Σ does not R-imply φ; hence, Σ does not imply φ
  - if $\varphi = X \twoheadrightarrow Y$, then $X \twoheadrightarrow R - Y \notin \Sigma^+_{\mathfrak{AC}}$ since $R - Y$ is not a subset of T; from $X \twoheadrightarrow Y \notin \Sigma^+_{\mathfrak{AC}}$ and $X \twoheadrightarrow R - Y \notin \Sigma^+_{\mathfrak{AC}}$ we conclude $X \twoheadrightarrow Y \notin \Sigma^+_{\mathfrak{AC}_{\mathcal{C}}}$; R-completeness of $\mathfrak{AC}_{\mathcal{C}}$ shows that Σ does not R-imply φ; hence, Σ does not imply φ

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Full Hierarchical Dependencies

- *full hierarchical dependency* (FHD) on relation schema R is expression $X : S$ where $X \subseteq R$ and S is a non-empty set of pairwise disjoint subsets of R that are also disjoint from X, i.e., $S \neq \emptyset$, for all $Y \in S$ we have $Y \subseteq R$ and for all $Y, Z \in S \cup \{X\}$ we have $Y \cap Z = \emptyset$

- R-relation r *satisfies* $X : \{Y_1, \ldots, Y_k\}$ on R iff for all $t_1, \ldots, t_{k+1} \in r$ the following condition is satisfied: if $t_i[X] = t_j[X]$ for all $1 \leq i, j \leq k+1$, then there is some $t \in r$ such that $t[XY_i] = t_i[XY_i]$ for $i = 1, \ldots, k$ and $t[X(R - XY_1 \cdots Y_k)] = t_{k+1}[X(R - XY_1 \cdots Y_k)]$

- Let $X, Y_1, \ldots, Y_k \subseteq R$ be pairwise disjoint and $k \geq 1$.
  An R-relation r satisfies the FHD $X : \{Y_1, \ldots, Y_k\}$ on R if and only if $r = r[XY_1] \bowtie \cdots \bowtie r[XY_k] \bowtie r[X(R - XY_1 \cdots Y_k)]$.

Sound Inference Rules for FDs and FHDs

$$\frac{}{X \to Y}\, Y \subseteq X$$
(reflexivity, $\mathcal{R}_{\mathrm{F}}$)

$$\frac{X \to Y}{X \to XY}$$
(extension, $\mathcal{E}_{\mathrm{F}}$)

$$\frac{X \to Y, Y \to Z}{X \to Z}$$
(transitivity, $\mathcal{T}_{\mathrm{F}}$)

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

$$\frac{XY : \{Y_1, \ldots, Y_k\}, X : \{Y\}}{X : \{Y_1, \ldots, Y_k, Y\}}$$
(transitivity, $\mathcal{T}_{\mathrm{H}}$)

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

$$\frac{X : \{Y_1, \ldots, Y_k, Y_{k+1}\}}{X : \{Y_1, \ldots, Y_k Y_{k+1}\}}$$
(merging, $\mathcal{M}_{\mathrm{H}}$)

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

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

$$\frac{X : \{Y\}, W \to Z}{X \to Y \cap Z}\, Y \cap W = \emptyset$$
(mixed subset, $\mathcal{S}_{\mathrm{FH}}$)

$$\frac{X : \{Y\}, W : \{Y_1, \ldots, Y_k\}}{X : \{Y \cap Y_1, \ldots, Y \cap Y_k, Y - Y_1 \cdots Y_k\}}\, Y \cap W = \emptyset$$
(subset, $\mathcal{S}_{\mathrm{H}}$)

$$\frac{X : \{Y\}, XY \to Z}{X \to Z - Y}$$
(mixed transitivity, $\mathcal{T}_{\mathrm{FH}}$)

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Main Results for FDs and FHDs

$$\mathfrak{HAE}_{\mathcal{C}} = \mathfrak{HA}_{\mathcal{C}} \cup \{\mathcal{S}_{\mathrm{H}}, \mathcal{M}_{\mathrm{H}}\} = (\mathfrak{HE}_{\mathcal{C}} - \{\mathcal{T}_{\mathrm{FH}}\}) \cup \{\mathcal{S}_{\mathrm{FH}}\}$$

appropriate: adequate and complementary

$$\mathfrak{HA}_{\mathcal{C}} = (\mathfrak{HF}_{\mathcal{C}} - \{\mathcal{T}_{\mathrm{FH}}\}) \cup \{\mathcal{S}_{\mathrm{FH}}\}$$

adequate

$$\mathfrak{HE}_{\mathcal{C}} = \mathfrak{HF}_{\mathcal{C}} \cup \{\mathcal{S}_{\mathrm{H}}, \mathcal{M}_{\mathrm{H}}\}$$

complementary

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

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Example Transformation - Shifting Complementation

- inappropriate inference:

$$\dfrac{\dfrac{X:\{Y\}}{\mathcal{C}^{\mathrm{R}}_{\mathrm{H}}: \quad X:\{R-XY\}} \quad X(R-XY):\{Y_1,\dots,Y_k\}}{\mathcal{T}_{\mathrm{H}}: \quad X:\{Y_1,\dots,Y_k,R-XY\}}$$

- appropriate inference: applying subset rule $\mathcal{S}_{\mathrm{H}}$ instead

$$\dfrac{\dfrac{X:\{Y\} \qquad X(R-XY):\{Y_1,\dots,Y_k\}}{\mathcal{S}_{\mathrm{H}}: \quad X:\{Y_1\cap Y,\dots,Y_k\cap Y, Y-Y_1\cdots Y_k\}}}{\mathcal{C}^{\mathrm{R}}_{\mathrm{H}}: \quad X:\{Y_1,\dots,Y_k,\underbrace{R-XY_1\cdots Y_k(Y-Y_1\cdots Y_k)}_{=R-XYY_1\cdots Y_k=R-XY}\}}$$

- Y and $X(R-XY)$ are disjoint, and $Y_1\cdots Y_k$ and $X(R-XY)$ are disjoint; hence, $Y_1\cdots Y_k\subseteq Y$, and thus $Y_i\cap Y=Y_i$ for all $i=1,\dots,k$

Example Tranformation - Eliminating Complementation

- inappropriate inference:

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

- notice that X and Y are disjoint, Hence, $X(R - XY) = R - Y$ and, consequently, $Z - (R - Y) = Y \cap Z$
- approprite inference:

$$\dfrac{X : \{Y\} X(R - XY) \to Z}{\mathcal{S}_{\mathrm{FH}}: \quad X \to Y \cap Z}$$

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Literature

- *Beeri, Fagin, Howard: A Complete Axiomatization for Functional and Multivalued Dependencies in Database Relations, SIGMOD, pp. 47-61, 1977*
- *Biskup: On the complementation rule for multivalued dependencies in database relations, Act.Inf.:10(3), pp. 297-305, 1978.*
- *Mendelzon: On axiomatising multivalued dependencies in relational databases, J. ACM:26(1), pp. 37-44, 1979.*
- *Biskup: Inferences of multivalued dependencies in fixed and undetermined universes, Theor.Comp.Sci:10(1), pp. 93-106, 1980.*
- *Link: On multivalued dependencies in fixed and undetermined universes, LNCS 3861, pp. 258-277, 2006.*
- *Hartmann, Link: On a problem of Fagin concerning multivalued dependencies in relational databases, Theor.Comp.Sci.:353(1-3), pp. 53-62, 2006*
- *Hartmann, Köhler, Link: Full hierarchical dependencies in fixed and undetermined universes, Ann. Math. Artif. Intell. 50(1-2): 195-226, 2007*
- *Biskup, Link: Appropriate reasoning about data dependencies in fixed and undetermined universes. LNCS 4932: pp. 58-77, 2008*

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

Conclusion and Outlook

- Complementation rule is mere means of database normalisation:
  - FHD-inference requires at most one application in last step;
  - FD-inference does not require an application at all
- Inappropriate inferences can be converted into appropriate inferences.
- Study the significance of undetermined universes for views!
- Find appropriate axiomatisations of F(H)Ds in ER Models.
- Find appropriate axiomatisations of F(H)Ds in Nested Models.
- Find appropriate axiomatisations of fuzzy and approximate F(H)Ds.