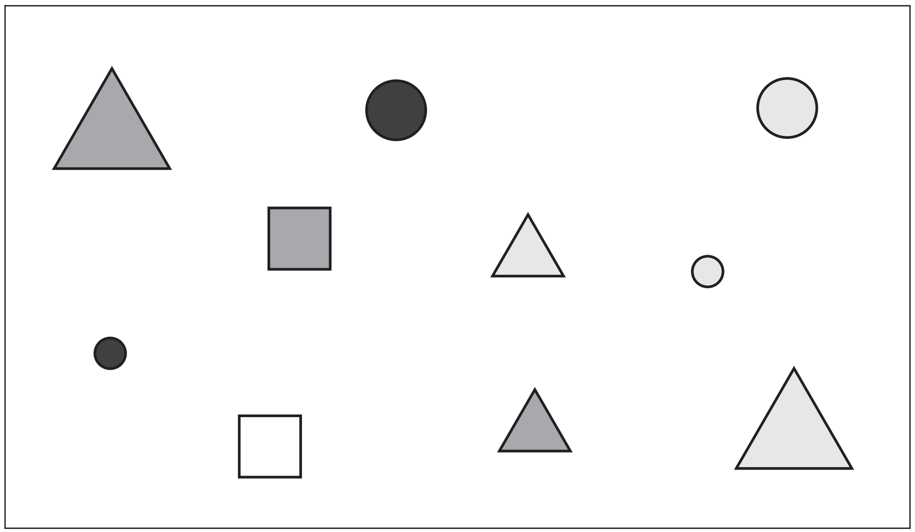
Decomposition

Example World



- 10 simple geometric objects
- 3 attributes

Relation

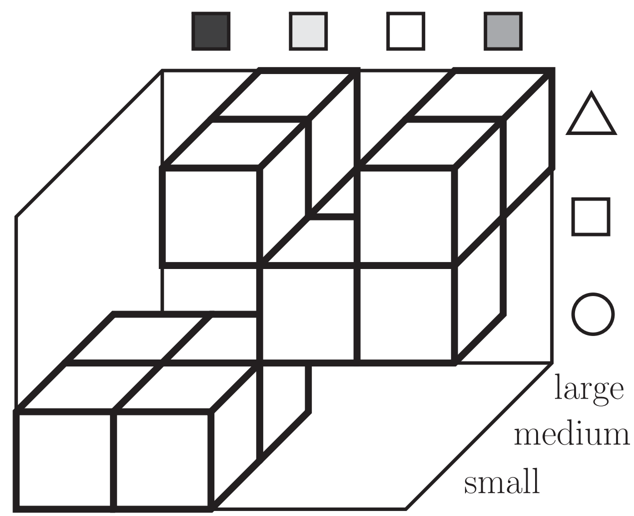
color	shape	size
■	○	small
■	○	medium
▢	○	small
▢	○	medium
▢	△	medium
▢	△	large
□	□	medium
▨	□	medium
▨	△	medium
▨	△	large

Example

Relation

color	shape	size
■	○	small
■	○	medium
▢	○	small
▢	○	medium
▢	△	medium
▢	△	large
□	□	medium
▩	□	medium
▩	△	medium
▩	△	large

Geometric Representation



Object Representation

- **Universe of Discourse**: Ω
- $\omega \in \Omega$ represents a single abstract object.
- A subset $E \subseteq \Omega$ is called an **event**.
- For every event we use the function R to determine whether E is possible or not.

$$R: \; 2^{\Omega} \to \{0, 1\}$$

- We claim the following properties of R:
  1. $R(\emptyset) = 0$
  2. $\forall E_1, E_2 \subseteq \Omega: \; R(E_1 \cup E_2) \; = \; \max\{R(E_1), R(E_2)\}$
- For example:

$$R(E) \; = \; \begin{cases} 0 & \text{if } \; E = \emptyset \\ 1 & \text{otherwise} \end{cases}$$

Object Representation

- Attributes or Properties of these objects are introduced by functions: (later referred to as **random variables**)

$$A: \ \Omega \to \mathrm{dom}(A)$$

where $\mathrm{dom}(A)$ is the domain (i. e., set of all possible values) of A.

- A set of attibutes $U = \{A_1, \ldots, A_n\}$ is called an **attribute schema**.
- The **preimage** of an attribute defines an **event**:

$$\forall a \in \mathrm{dom}(A): \ A^{-1}(a) = \{\omega \in \Omega \mid A(\omega) = a\} \subseteq \Omega$$

- Abbreviation: $A^{-1}(a) \quad = \quad \{\omega \in \Omega \mid A(\omega) = a\} \quad = \quad \{A = a\}$
- We will index the function R to stress on which events it is defined. R_{AB} will be short for $R_{\{A,B\}}$.

$$R_{AB}: \bigcup_{a \in \mathrm{dom}(A)} \bigcup_{b \in \mathrm{dom}(B)} \Big\{\{A = a, B = b\}\Big\} \to \{0, 1\}$$

Formal Representation

A = color	B = shape	C = size
a_1 = ■	b_1 = ○	c_1 = small
a_1 = ■	b_1 = ○	c_2 = medium
a_2 = □	b_1 = ○	c_1 = small
a_2 = □	b_1 = ○	c_2 = medium
a_2 = □	b_3 = △	c_2 = medium
a_2 = □	b_3 = △	c_3 = large
a_3 = □	b_2 = □	c_2 = medium
a_4 = ■	b_2 = □	c_2 = medium
a_4 = ■	b_3 = △	c_2 = medium
a_4 = ■	b_3 = △	c_3 = large

$$
\begin{aligned}
&R_{ABC}(A = a, B = b, C = c) \\
&\quad = R_{ABC}(\{A = a, B = b, C = c\}) \\
&\quad = R_{ABC}(\{\omega \in \Omega \mid A(\omega) = a \wedge \\
&\qquad\qquad B(\omega) = b \wedge \\
&\qquad\qquad C(\omega) = c\}) \\
&\quad = \begin{cases} 0 & \text{if there is no tuple } (a, b, c) \\ 1 & \text{else} \end{cases}
\end{aligned}
$$

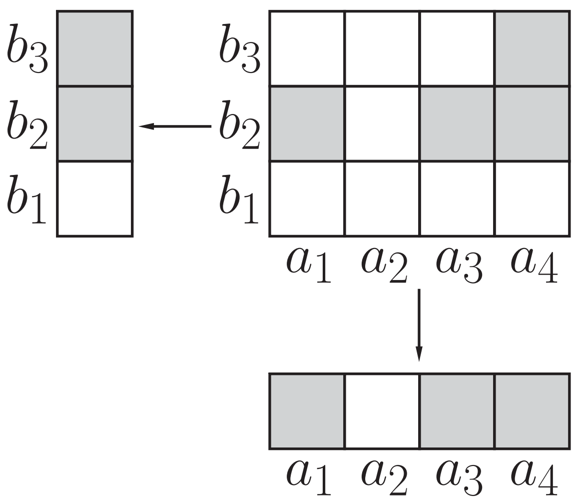
R serves as an indicator function.

Projection / Marginalization

Let R_{AB} be a relation over two attributes A and B. The projection (or marginalization) from schema $\{A, B\}$ to schema $\{A\}$ is defined as:

$$\forall a \in \mathrm{dom}(A): \; R_A(A = a) \; = \max_{\forall b \in \mathrm{dom}(B)} \{R_{AB}(A = a, B = b)\}$$

This principle is easily generalized to sets of attributes.

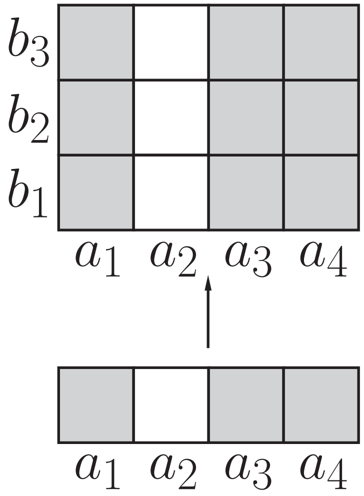


Cylindrical Extention

Let R_A be a relation over an attribute A. The cylindrical extention R_{AB} from $\{A\}$ to $\{A, B\}$ is defined as:

$$\forall a \in \mathrm{dom}(A) : \forall b \in \mathrm{dom}(B) : \; R_{AB}(A = a, B = b) \; = \; R_A(A = a)$$

This principle is easily generalized to sets of attributes.

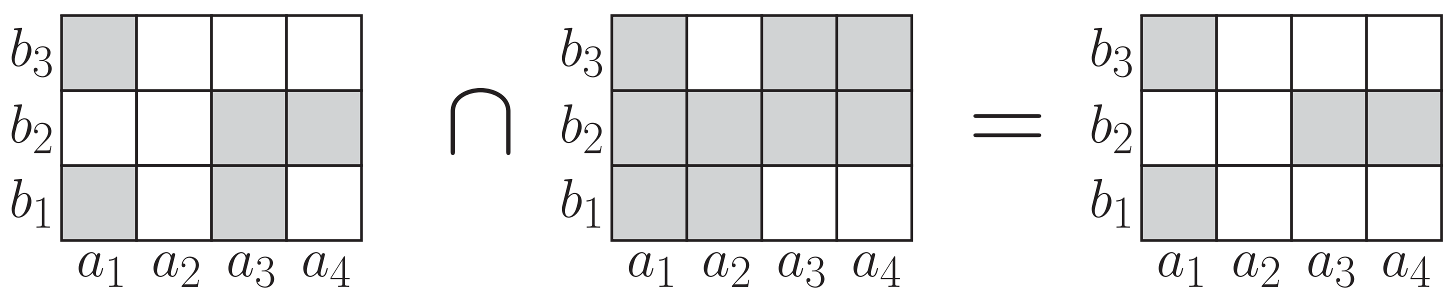


Intersection

Let $R_{AB}^{(1)}$ and $R_{AB}^{(2)}$ be two relations with attribute schema $\{A, B\}$. The intersection R_{AB} of both is defined in the natural way:

$$\forall a \in \text{dom}(A) : \forall b \in \text{dom}(B) :$$
$$R_{AB}(A = a, B = b) \;=\; \min\{R_{AB}^{(1)}(A = a, B = b), R_{AB}^{(2)}(A = a, B = b)\}$$

This principle is easily generalized to sets of attributes.

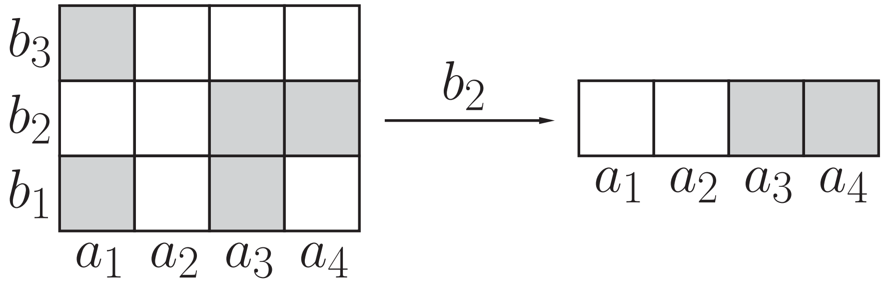


Conditional Relation

Let R_{AB} be a relation over the attribute schema $\{A, B\}$. The conditional relation of A given B is defined as follows:

$$\forall a \in \text{dom}(A) : \forall b \in \text{dom}(B) : R_A(A = a \mid B = b) \;=\; R_{AB}(A = a, B = b)$$

This principle is easily generalized to sets of attributes.

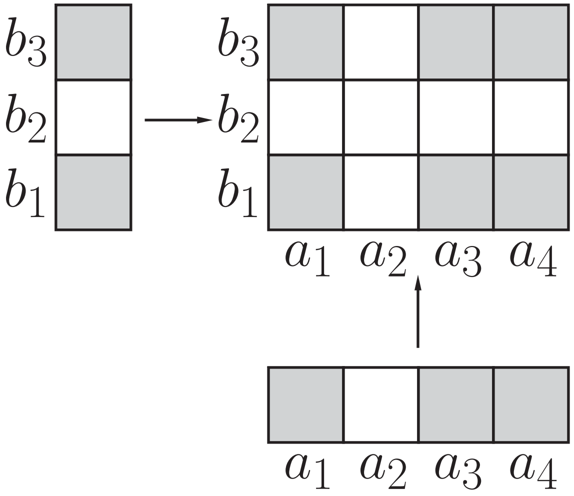


(Unconditional) Independence

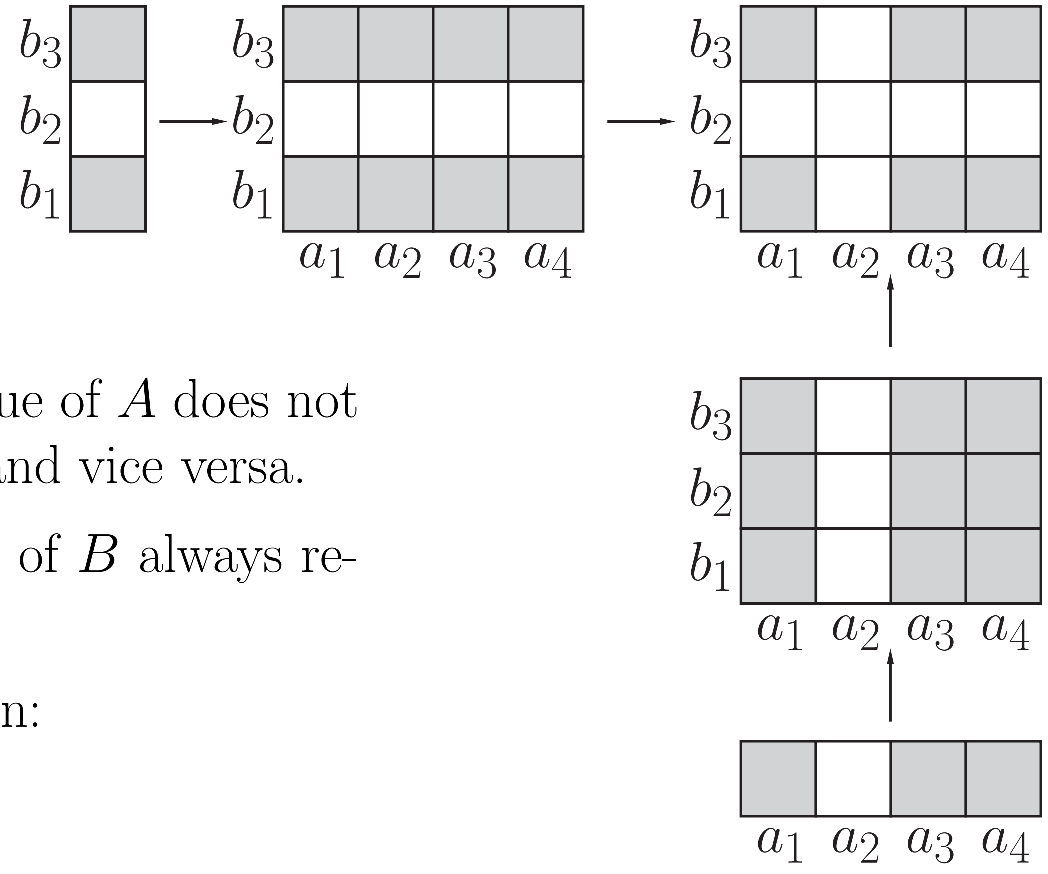
Let R_{AB} be a relation over the attribute schema $\{A, B\}$. We call A and B relationally independent (w. r. t. R_{AB}) if the following condition holds:

$$\forall a \in \mathrm{dom}(A) : \forall b \in \mathrm{dom}(B) : R_{AB}(A = a, B = b) \; = \; \min\{R_A(A = a), R_B(B = b)\}$$

This principle is easily generalized to sets of attributes.



(Unconditional) Independence



Intuition: Fixing one (possible) value of A does not restrict the (possible) values of B and vice versa.

Conditioning on any possible value of B always results in the same relation R_A.

Alternative independence expression:

$$\forall b \in \mathrm{dom}(B): \; R_B(B = b) = 1 : \\ R_A(A = a \mid B = b) \; = \; R_A(A = a)$$

Decomposition

- Obviously, the original two-dimensional relation can be reconstructed from the two one-dimensional ones, if we have (unconditional) independence.

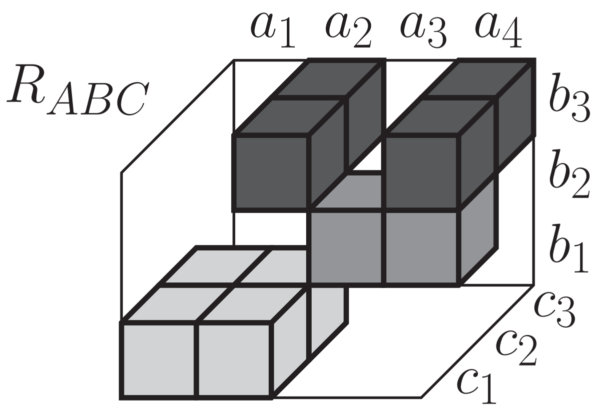
- The definition for (unconditional) independence already told us how to do so:

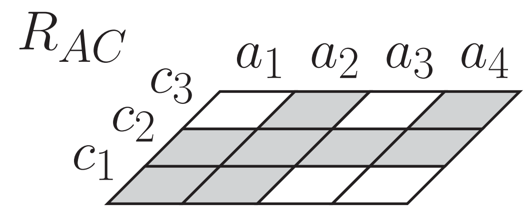
$$R_{AB}(A = a, B = b) \;=\; \min\{R_A(A = a), R_B(B = b)\}$$

- Storing R_A and R_B is sufficient to represent the information of R_{AB}.

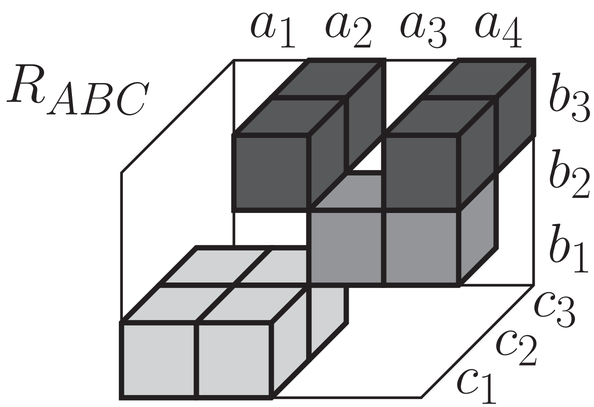
- **Question:** The (unconditional) independence is a rather strong restriction. Are there other types of independence that allow for a decomposition as well?

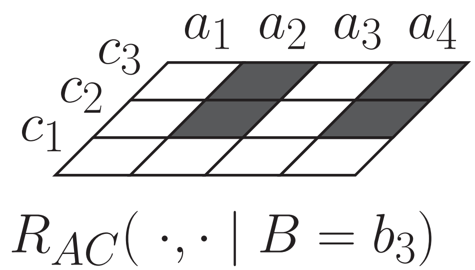
Conditional Relational Independence

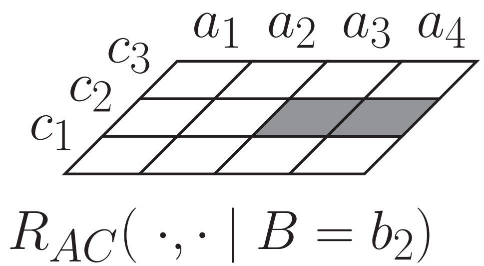


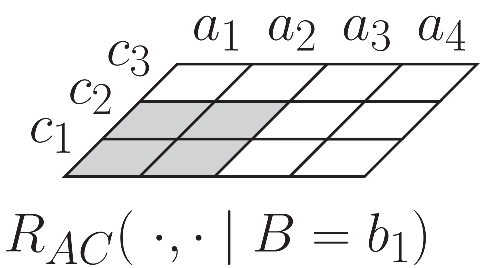


Clearly, A and C are unconditionally dependent, i. e. the relation R_{AC} cannot be reconstructed from R_A and R_C.









However, given all possible values of B, all respective conditional relations R_{AC} show the independence of A and C.

$$R_{AC}(a, c \mid b) \;=\; \min\{R_A(a \mid b), R_C(c \mid b)\}$$

With the definition of a conditional relation, the decomposition description for R_{ABC} reads:

$$R_{ABC}(a, b, c) \;=\; \min\{R_{AB}(a, b), R_{BC}(b, c)\}$$

Again, we reconstruct the initial relation from the cylindrical extentions of the two relations formed by the attributes A, B and B, C.

It is possible since A and C are (relationally) independent given B.