Decomposition

Object Representation

Property family	Car body	Motor	Radio	Doors	Seat cover	Makeup mirrow	...
Property	Hatch-back	2.8 L 150 kW Otto	Type alpha	4	Leather, Type L3	yes	...

About 200 variables

Typically 4 to 8, but up to 150 possible instances per variable

More than 2^{200} possible combinations available

Example 1: Planning in car manufacturing

Available information: 10000 technical rules, 200 attributes

> "If Motor $= m_4$ and Heating $= h_1$ then Generator $\in \{g_1, g_2, g_3\}$"
>
> "Engine type e_1 can only be combined with transmission t_2 or t_5."
>
> "Transmission t_5 requires crankshaft c_2."
>
> "Convertibles have the same set of radio options as SUVs."

Each peace of information corresponds to a constraint in a high dimensional subspace, possible questions/inferences:

> "Can a station wagon with engine e_4 be equipped with tire set y_6?"
>
> "Supplier S_8 failed to deliver on time. What production line has to be modified and how?"
>
> "Are there any peculiarities within the set of cars that suffered an aircondition failure?"

Idea to Solve the Problems

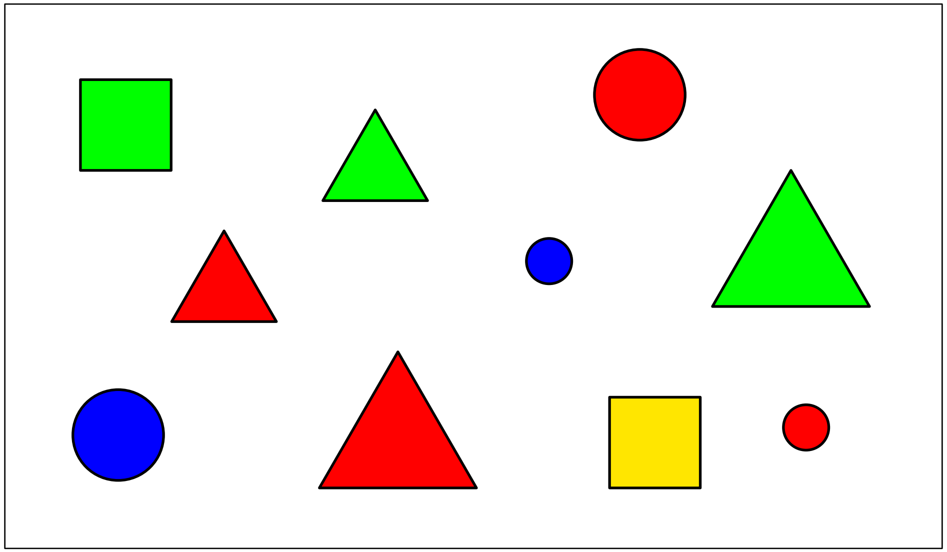
Given: A large (high-dimensional) δ representing the domain knowledge.

Desired: A set of smaller (lower-dimensional) $\{\delta_1, \ldots, \delta_s\}$ (maybe overlapping) from which the original δ *could* be reconstructed with no (or as few as possible) errors.

With such a decomposition we can draw any conclusions from $\{\delta_1, \ldots, \delta_s\}$ that could be inferred from δ — without, however, actually reconstructing it.

Example

Example World

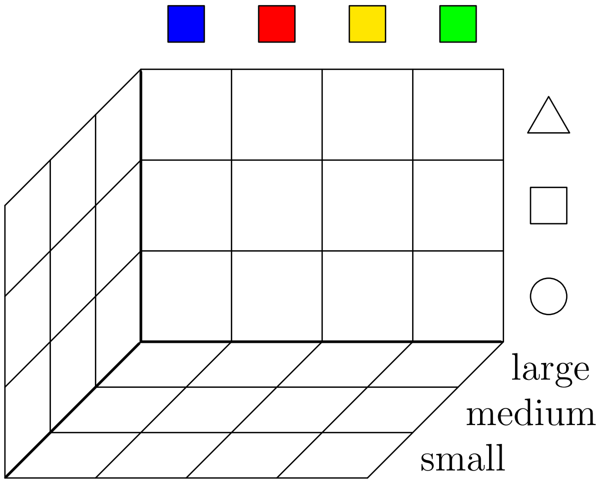


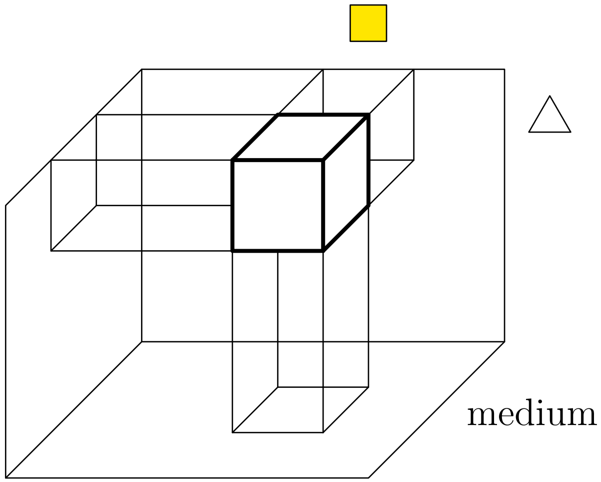
Relation

color	shape	size
blue	○	small
blue	○	medium
red	○	small
red	○	medium
red	△	medium
red	△	large
yellow	□	medium
green	□	medium
green	△	medium
green	△	large

- 10 simple geometric objects, 3 attributes
- One object is chosen at random and examined
- Inferences are drawn about the unobserved attributes

The Reasoning Space





The reasoning space consists of a finite set Ω of states.

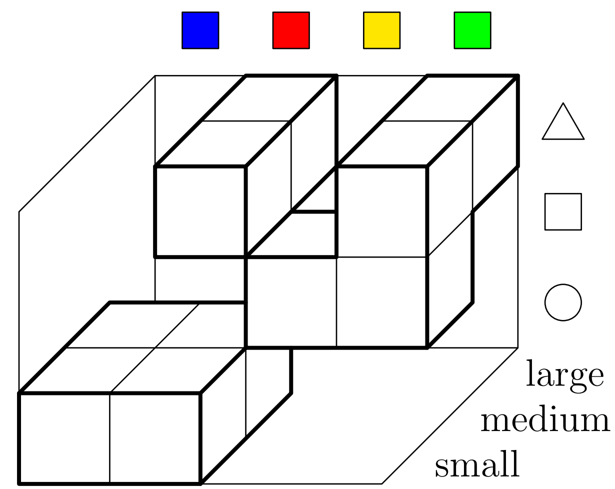
The states are described by a set of n attributes A_i, $i = 1, \dots, n$, whose domains $\{a_1^{(i)}, \dots, a_{n_i}^{(i)}\}$ can be seen as sets of propositions or events.

The events in a domain are mutually exclusive and exhaustive.

Relation

color	shape	size
blue	○	small
blue	○	medium
red	○	small
red	○	medium
red	△	medium
red	△	large
yellow	□	medium
green	□	medium
green	△	medium
green	△	large

Relation in the Reasoning Space



Each cube represents one tuple.

The spatial representation helps to understand the decomposition mechanism.

Possibility-Based Formalization

Definition: Let Ω be a (finite) sample space.
A **discrete possibility measure** R on Ω is a function $R : 2^\Omega \to \{0, 1\}$ satisfying

1. $R(\emptyset) = 0$ and
2. $\forall E_1, E_2 \subseteq \Omega : R(E_1 \cup E_2) = \max\{R(E_1), R(E_2)\}$.

Similar to Kolmogorov's axioms of probability theory.

If an event E can occur (if it is possible), then $R(E) = 1$,
otherwise (if E cannot occur/is impossible) then $R(E) = 0$.

$R(\Omega) = 1$ is not required, because this would exclude the empty relation.

From the axioms it follows $R(E_1 \cap E_2) \leq \min\{R(E_1), R(E_2)\}$.

Attributes are introduced as random variables (as in probability theory).

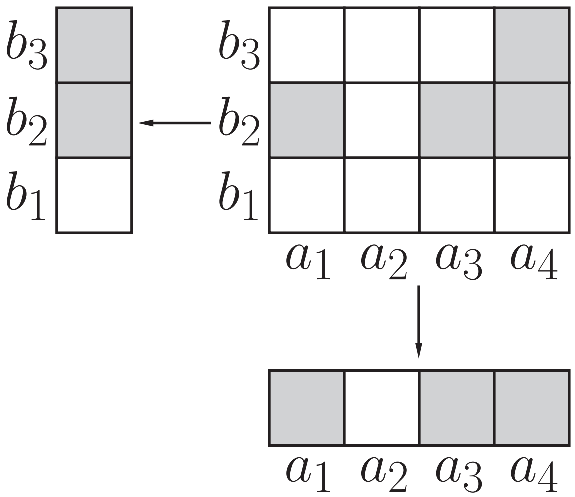
$R(A = a)$ and $R(a)$ are abbreviations of $R(\{\omega \mid A(\omega) = a\})$.

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.

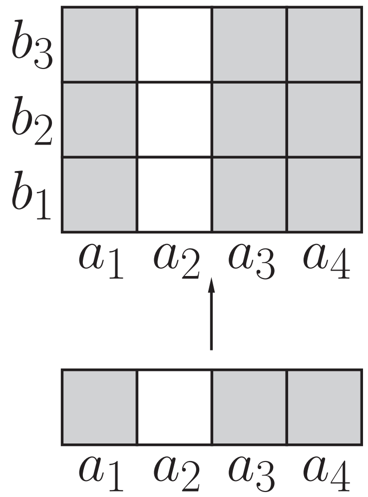


Cylindrical Extention

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.

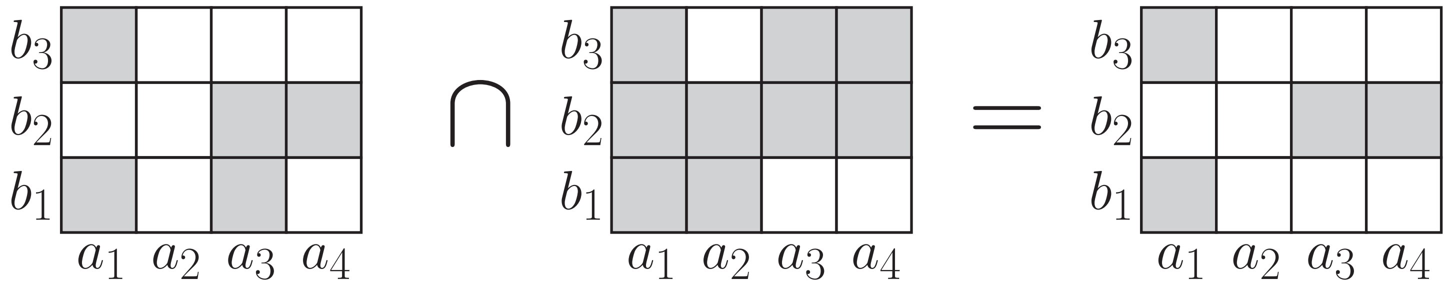


Intersection

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:

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

This principle is easily generalized to sets of attributes.

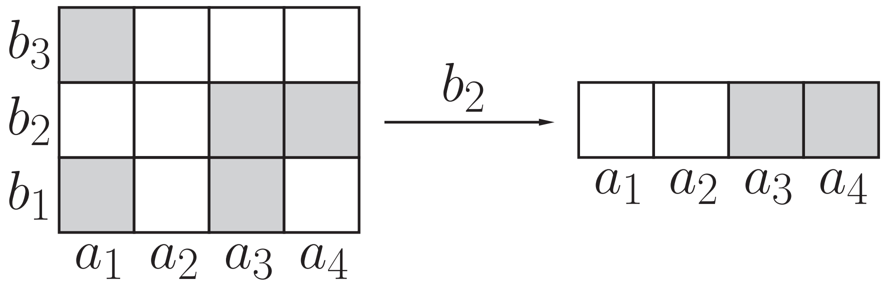


Conditional Relation

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

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

This principle is easily generalized to sets of attributes.

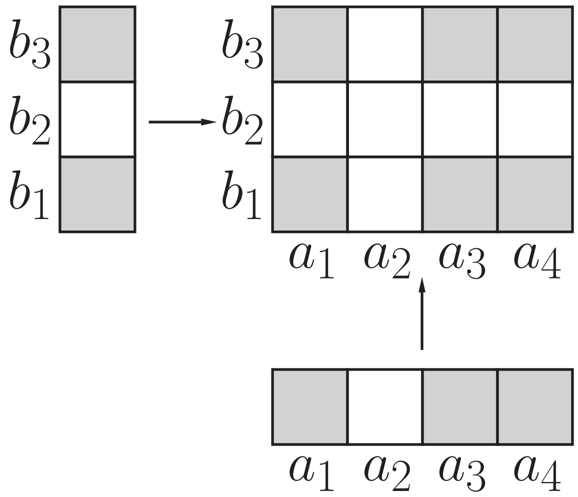


(Unconditional) Independence

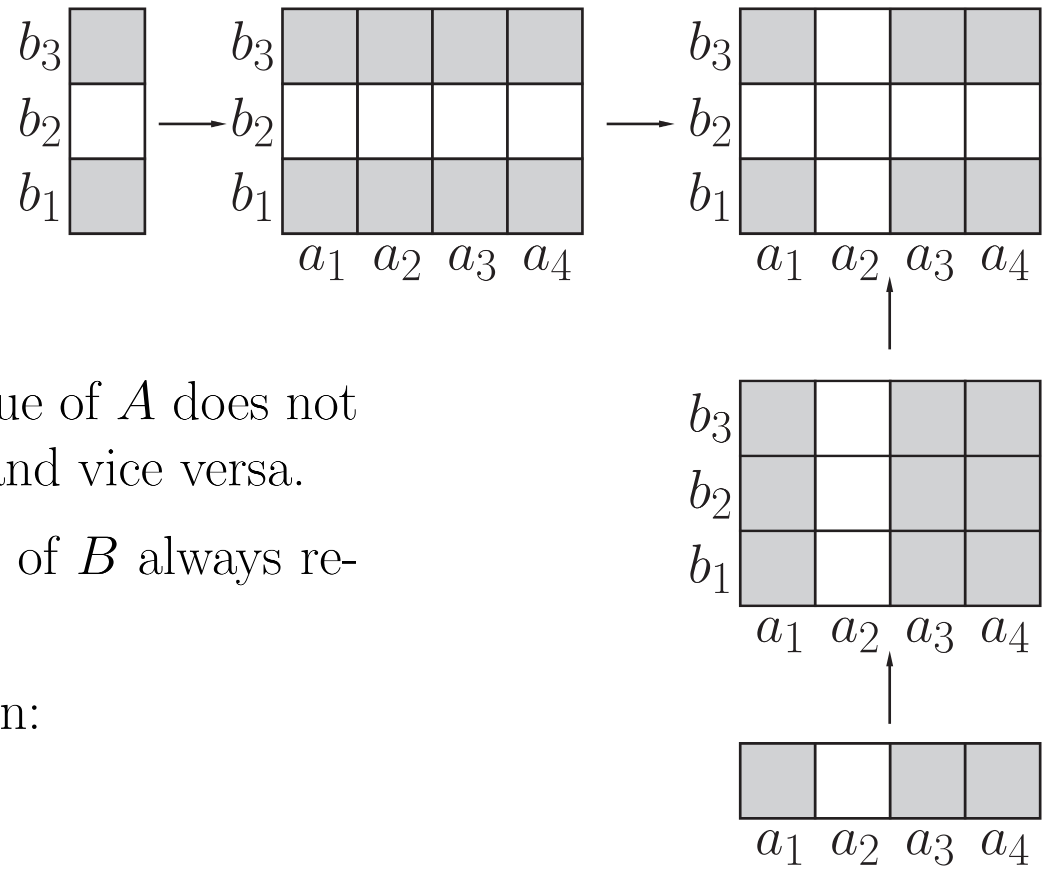
Let R_{AB} be a relation over the attribute schema $\{A, B\}$. We call A and B relationally independent (w. r. t. R_{AB}) if the following condition holds:

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

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.

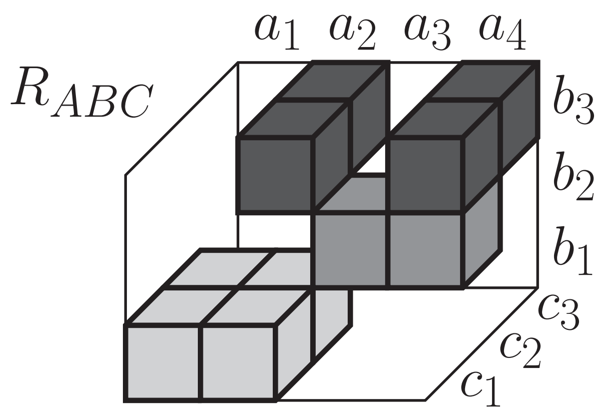
The definition for (unconditional) independence already told us how to do so:

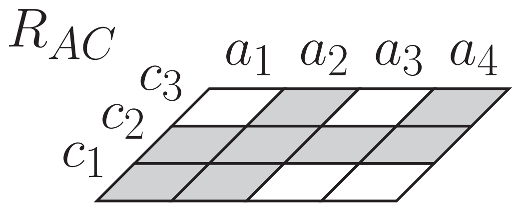
$$R_{AB}(A = a, B = b) \;=\; \min\{R_A(A = a), R_B(B = b)\}$$

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?

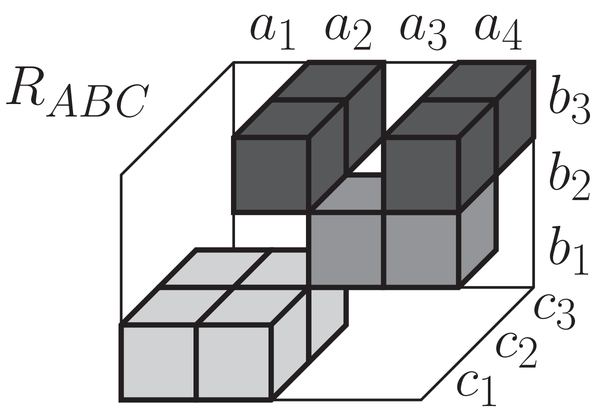
Conditional Relational Independence

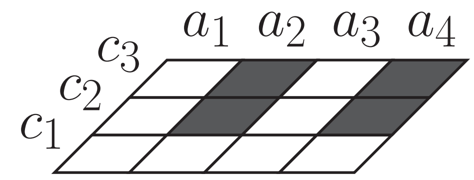




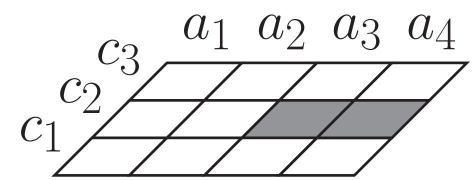
Clearly, A and C are unconditionally dependent, i. e. the relation R_{AC} cannot be reconstructed from R_A and R_C.

Conditional Relational Independence

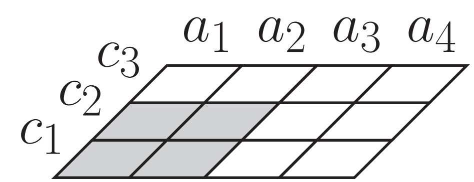




$R_{AC}(\,\cdot, \cdot \mid B = b_3)$



$R_{AC}(\,\cdot, \cdot \mid B = b_2)$



$R_{AC}(\,\cdot, \cdot \mid B = b_1)$

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

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

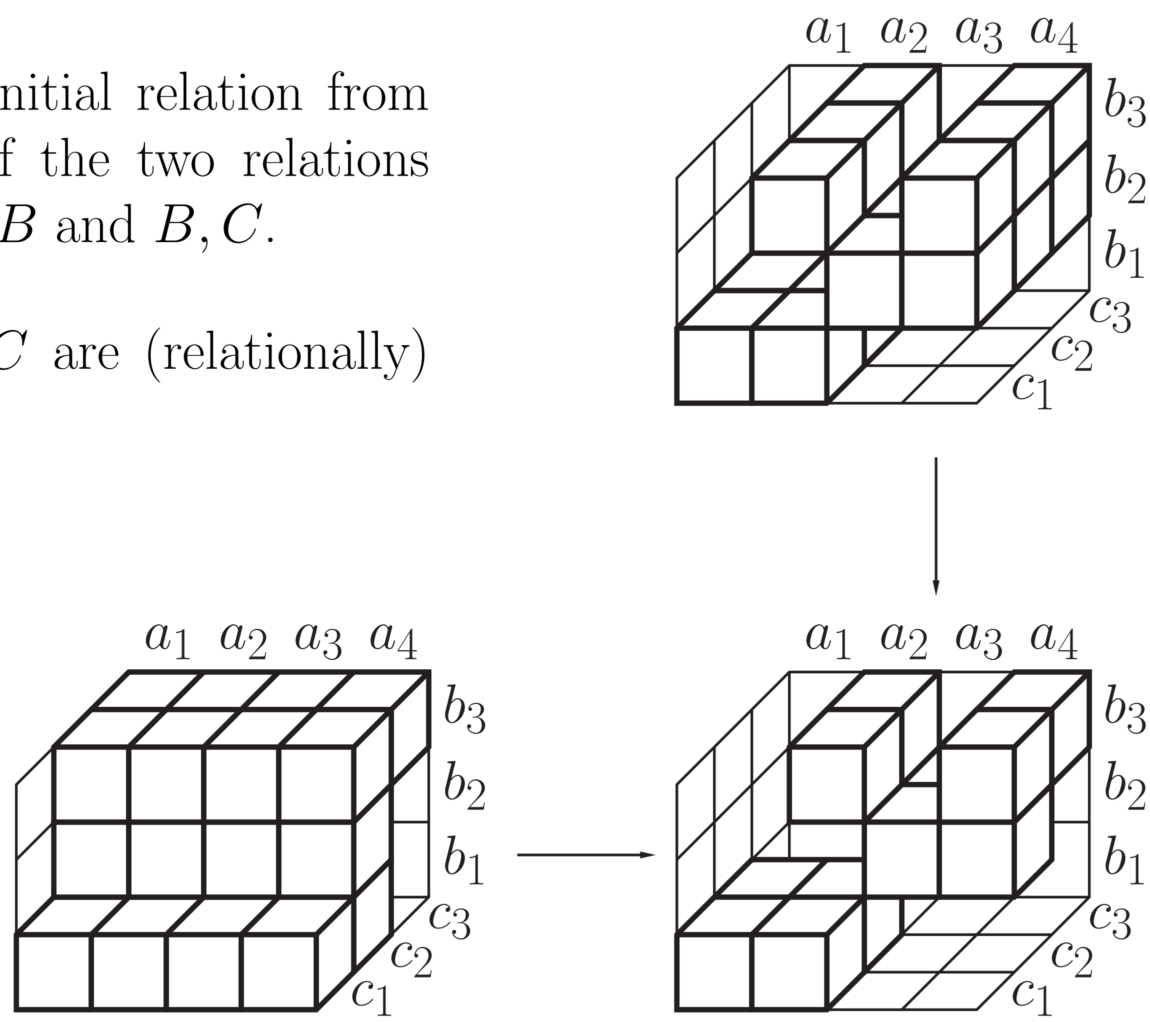
With the definition of a conditional relation, the decomposition description for R_{ABC} reads:

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

Conditional Relational Independence

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.



Possibility-Based Formalization (continued)

Definition: Let $U = \{A_1, \ldots, A_n\}$ be a set of attributes defined on a (finite) sample space Ω with respective domains $\mathrm{dom}(A_i)$, $i = 1, \ldots, n$. A **relation** r_U over U is the restriction of a discrete possibility measure R on Ω to the set of all events that can be defined by stating values for all attributes in U. That is, $r_U = R|_{\mathcal{E}_U}$, where

$$\begin{aligned}
\mathcal{E}_U &= \left\{ E \in 2^\Omega \,\middle|\, \exists a_1 \in \mathrm{dom}(A_1) : \ldots \exists a_n \in \mathrm{dom}(A_n) : E \,\widehat{=} \bigwedge_{A_j \in U} A_j = a_j \right\} \\
&= \left\{ E \in 2^\Omega \,\middle|\, \exists a_1 \in \mathrm{dom}(A_1) : \ldots \exists a_n \in \mathrm{dom}(A_n) : \right. \\
&\qquad\qquad \left. E = \left\{ \omega \in \Omega \,\middle|\, \bigwedge_{A_j \in U} A_j(\omega) = a_j \right\} \right\}.
\end{aligned}$$

A relation corresponds to the notion of a probability distribution.

Advantage of this formalization: No index transformation functions are needed for projections, there are just fewer terms in the conjunctions.

Definition: Let $U = \{A_1, \ldots, A_n\}$ be a set of attributes and r_U a relation over U. Furthermore, let $\mathcal{M} = \{M_1, \ldots, M_m\} \subseteq 2^U$ be a set of nonempty (but not necessarily disjoint) subsets of U satisfying

$$\bigcup_{M \in \mathcal{M}} M = U.$$

r_U is called **decomposable** w.r.t. $\mathcal{M}$ iff

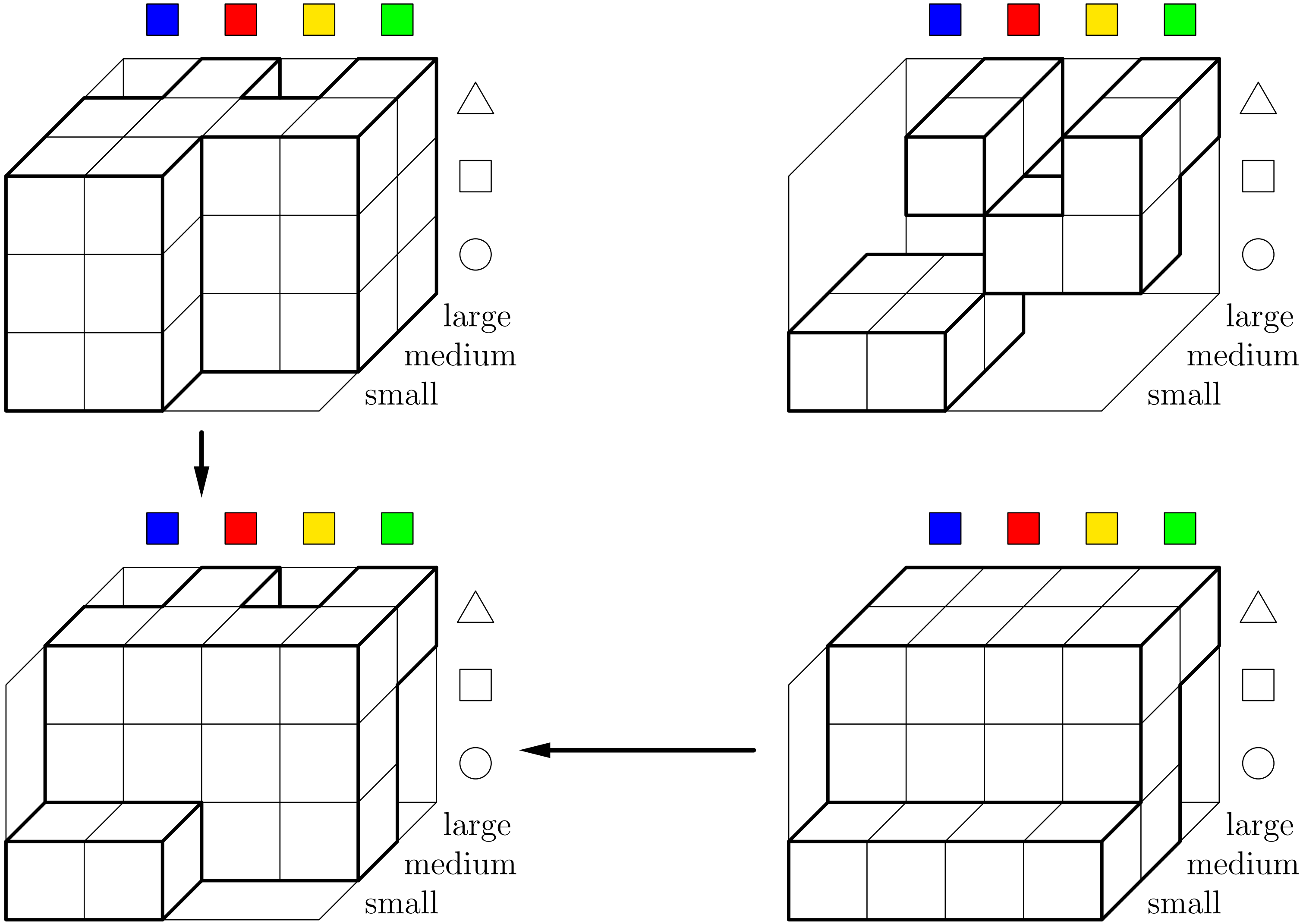
$$\forall a_1 \in \mathrm{dom}(A_1) : \ldots \forall a_n \in \mathrm{dom}(A_n) :$$
$$r_U\left(\bigwedge_{A_i \in U} A_i = a_i\right) = \min_{M \in \mathcal{M}} \left\{ r_M\left(\bigwedge_{A_i \in M} A_i = a_i\right)\right\}.$$

If r_U is decomposable w.r.t. $\mathcal{M}$, the set of relations

$$\mathcal{R}_{\mathcal{M}} = \{r_{M_1}, \ldots, r_{M_m}\} = \{r_M \mid M \in \mathcal{M}\}$$

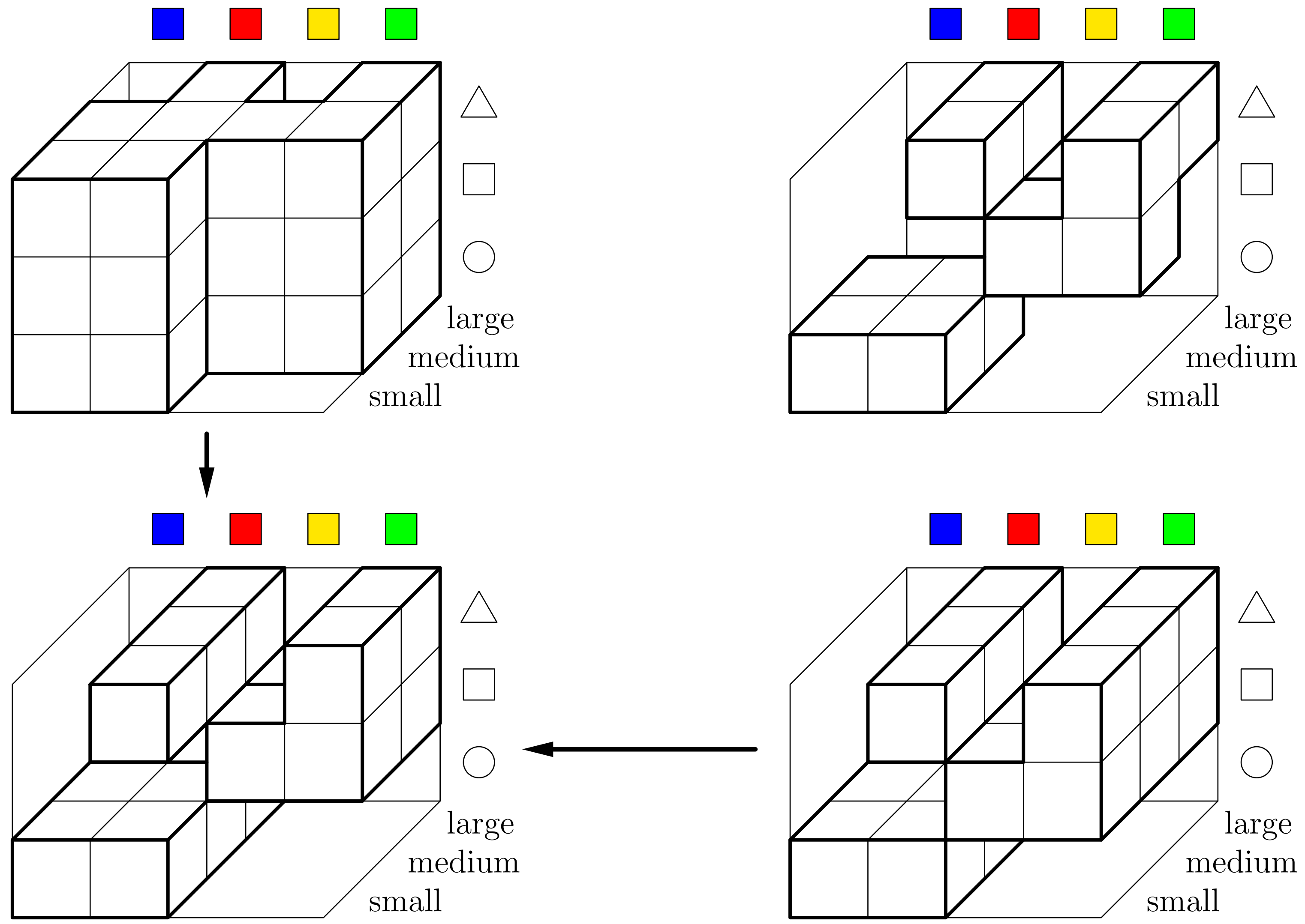
is called the **decomposition** of r_U.

Equivalent to **join decomposability** in database theory (natural join).



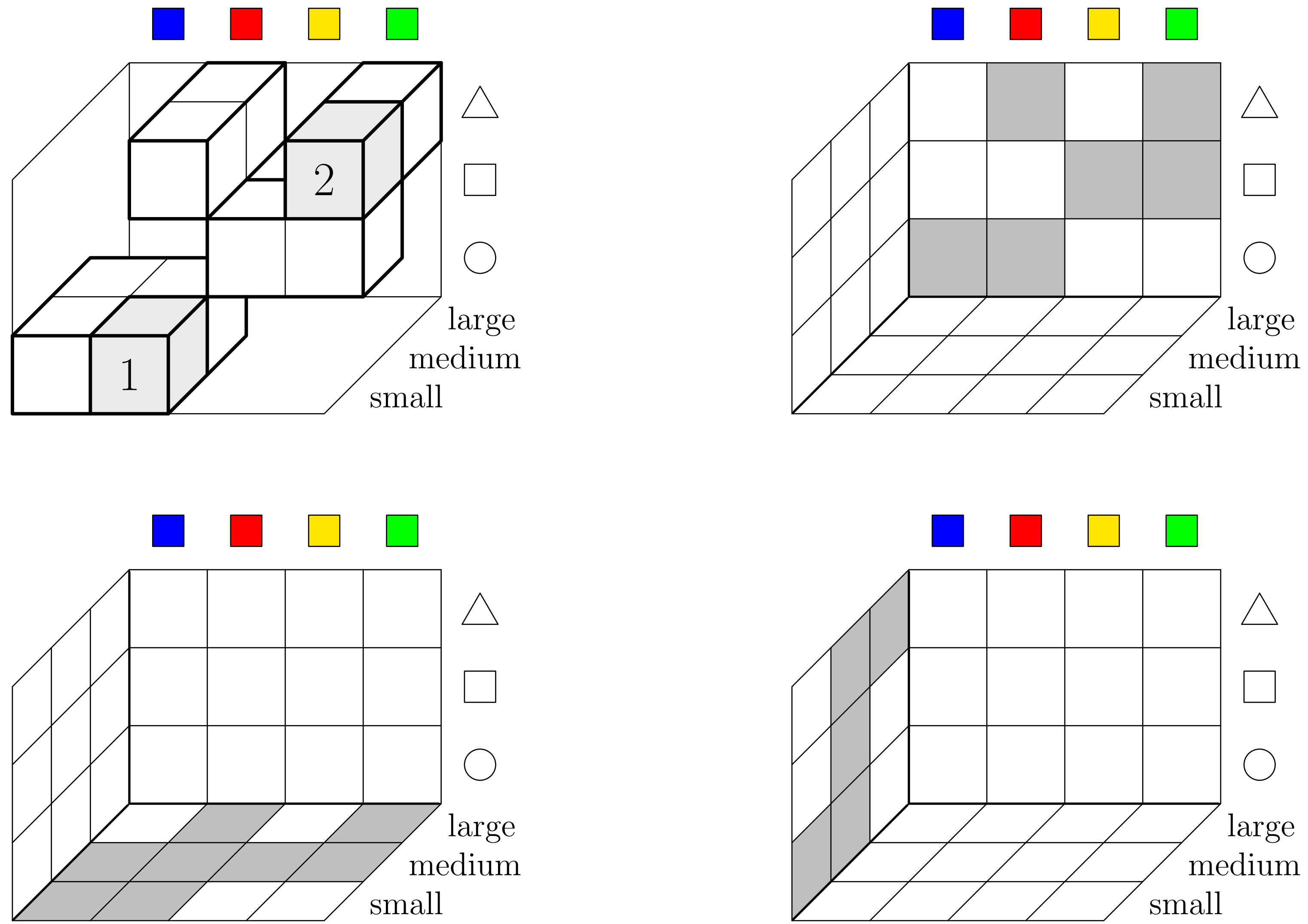
This choice of subspaces does not yield a decomposition.

Using other Projections 2



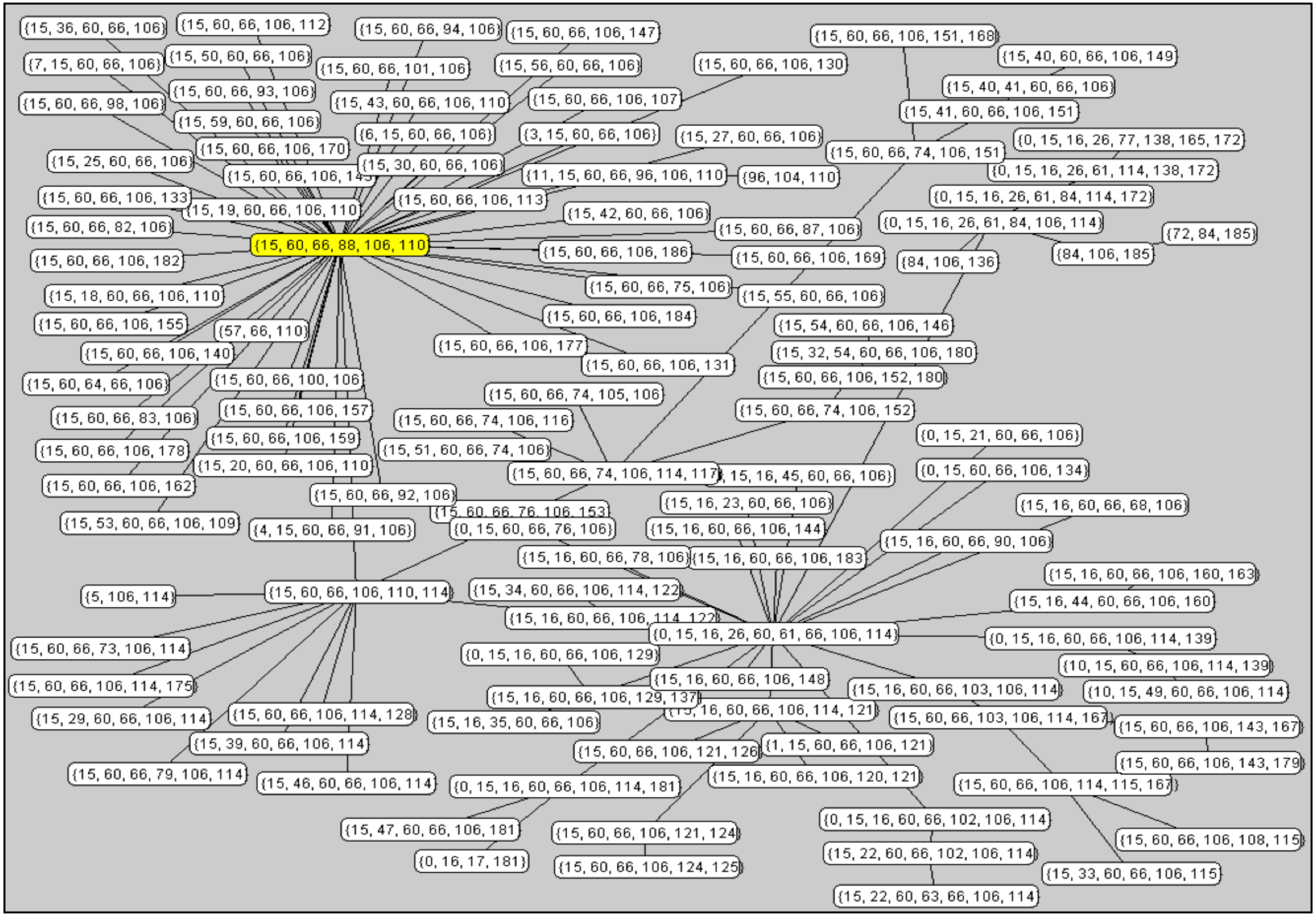
This choice of subspaces does not yield a decomposition.

Is Decomposition Always Possible?



A modified relation (without tuples 1 or 2) may not possess a decomposition.

Example: VW Bora

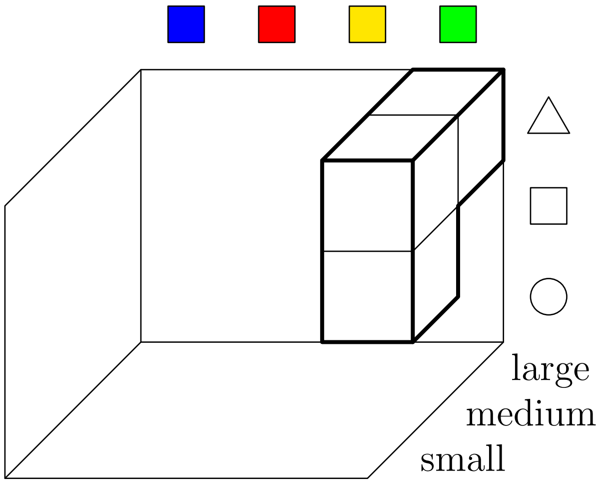


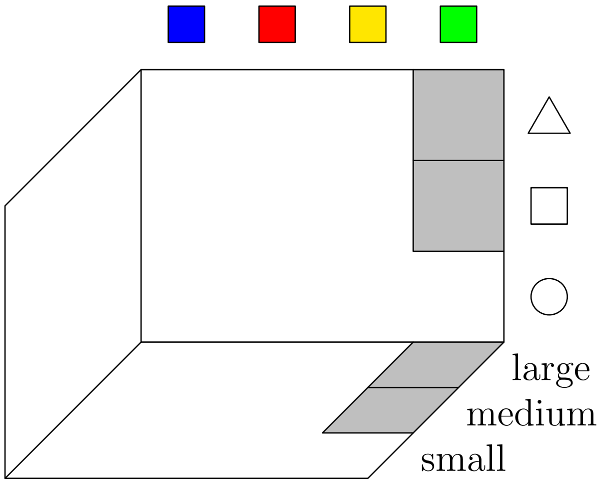
10000 rules
186 variables
174 subspaces

Reasoning

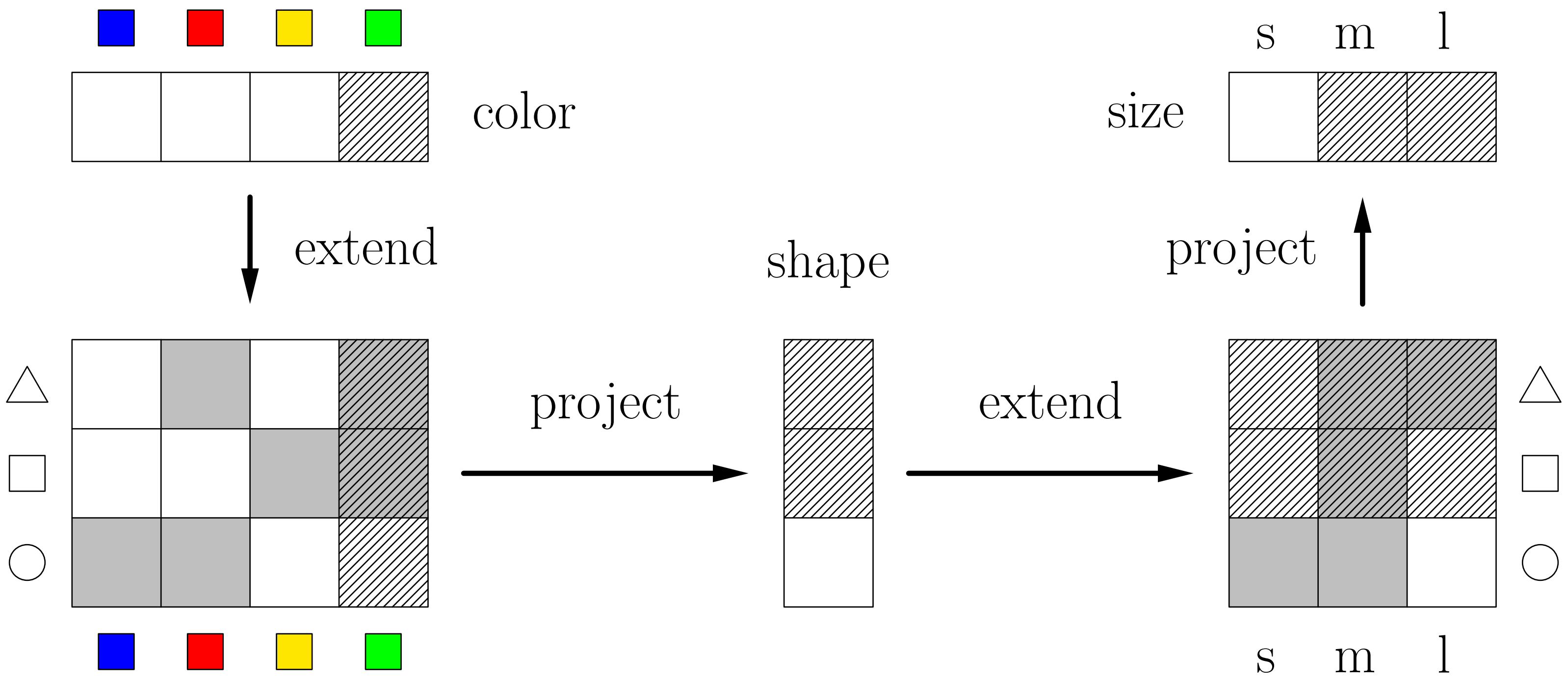
Let it be known (e.g. from an observation) that the given object is green.
This information considerably reduces the space of possible value combinations.
From the prior knowledge it follows that the given object must be

- either a triangle or a square and
- either medium or large.

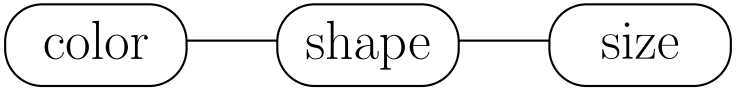




The reasoning result can be obtained using only the projections to the subspaces without reconstructing the original three-dimensional relation:

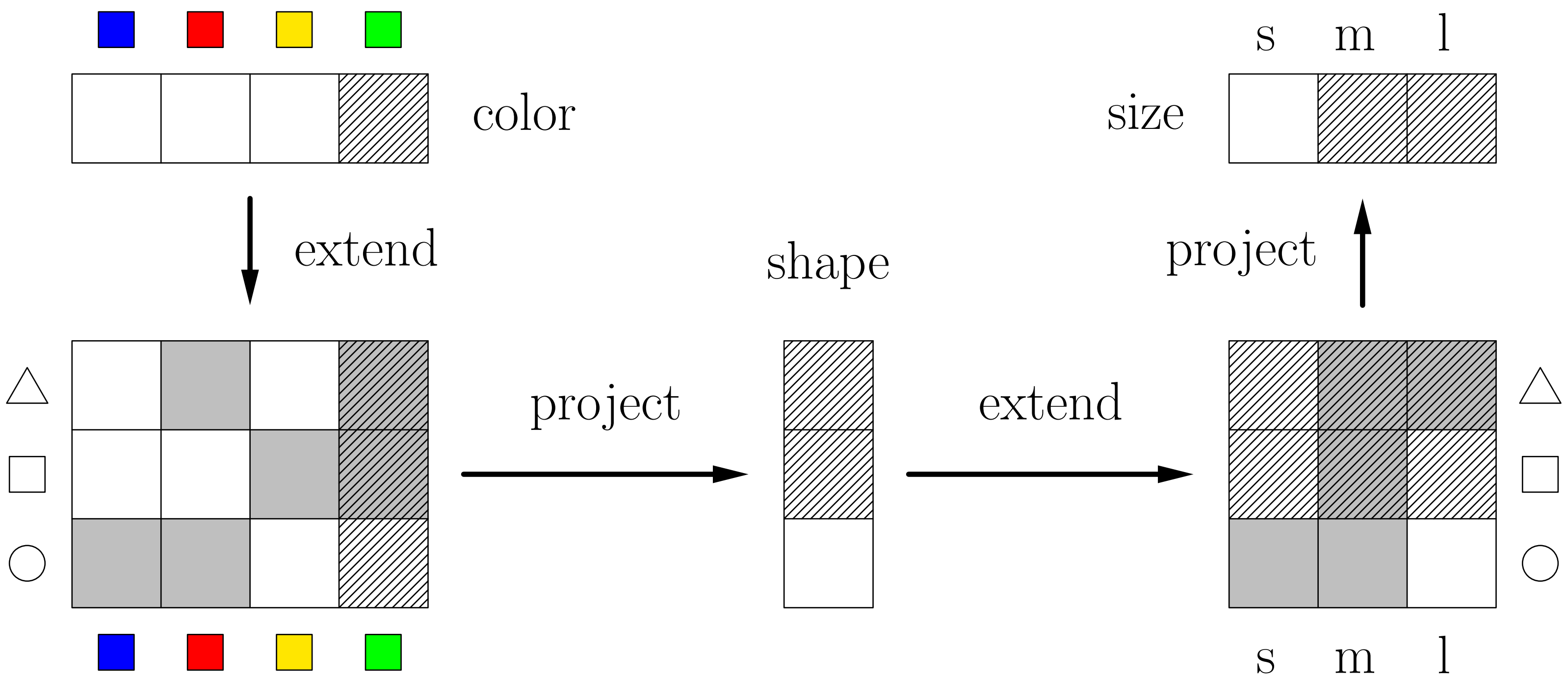


This justifies a graph representation:



Relational Evidence Propagation

Due to the fact that color and size are conditionally independent given the shape, the reasoning result can be obtained using only the projections to the subspaces:



This reasoning scheme can be formally justified with discrete possibility measures.

Relational Evidence Propagation, Step 1

A:	color
B:	shape
C:	size

$$
\begin{aligned}
&R(B = b \mid A = a_{\text{obs}}) \\
&= \quad R\left(\bigvee_{a \in \text{dom}(A)} A = a, B = b, \bigvee_{c \in \text{dom}(C)} C = c \,\middle|\, A = a_{\text{obs}}\right) \\
&\overset{(1)}{=} \quad \max_{a \in \text{dom}(A)} \{ \max_{c \in \text{dom}(C)} \{ R(A = a, B = b, C = c \mid A = a_{\text{obs}}) \} \} \\
&\overset{(2)}{=} \quad \max_{a \in \text{dom}(A)} \{ \max_{c \in \text{dom}(C)} \{ \min\{ R(A = a, B = b, C = c), R(A = a \mid A = a_{\text{obs}}) \} \} \} \\
&\overset{(3)}{=} \quad \max_{a \in \text{dom}(A)} \{ \max_{c \in \text{dom}(C)} \{ \min\{ R(A = a, B = b), R(B = b, C = c), \\
&\qquad\qquad\qquad\qquad\qquad\qquad R(A = a \mid A = a_{\text{obs}}) \} \} \} \\
&= \quad \max_{a \in \text{dom}(A)} \{ \min\{ R(A = a, B = b), R(A = a \mid A = a_{\text{obs}}), \\
&\qquad\qquad\qquad \underbrace{\max_{c \in \text{dom}(C)} \{ R(B = b, C = c) \} }_{= R(B=b) \geq R(A=a, B=b)} \} \} \\
&= \quad \max_{a \in \text{dom}(A)} \{ \min\{ R(A = a, B = b), R(A = a \mid A = a_{\text{obs}}) \} \}.
\end{aligned}
$$

Relational Evidence Propagation, Step 1 (continued)

(1) holds because of the second axiom a discrete possibility measure has to satisfy.

(3) holds because of the fact that the relation R_{ABC} can be decomposed w.r.t. the set $\mathcal{M} = \{\{A, B\}, \{B, C\}\}$. ($A$: color, B: shape, C: size)

(2) holds, since in the first place

$$
\begin{aligned}
R(A = a, B = b, C = c \,|\, A = a_{obs}) &= R(A = a, B = b, C = c, A = a_{obs}) \\
&= \begin{cases} R(A = a, B = b, C = c), & \text{if } a = a_{\text{obs}}, \\ 0, & \text{otherwise}, \end{cases}
\end{aligned}
$$

and secondly

$$
\begin{aligned}
R(A = a \mid A = a_{\text{obs}}) &= R(A = a, A = a_{\text{obs}}) \\
&= \begin{cases} R(A = a), & \text{if } a = a_{\text{obs}}, \\ 0, & \text{otherwise}, \end{cases}
\end{aligned}
$$

and therefore, since trivially $R(A = a) \geq R(A = a, B = b, C = c)$,

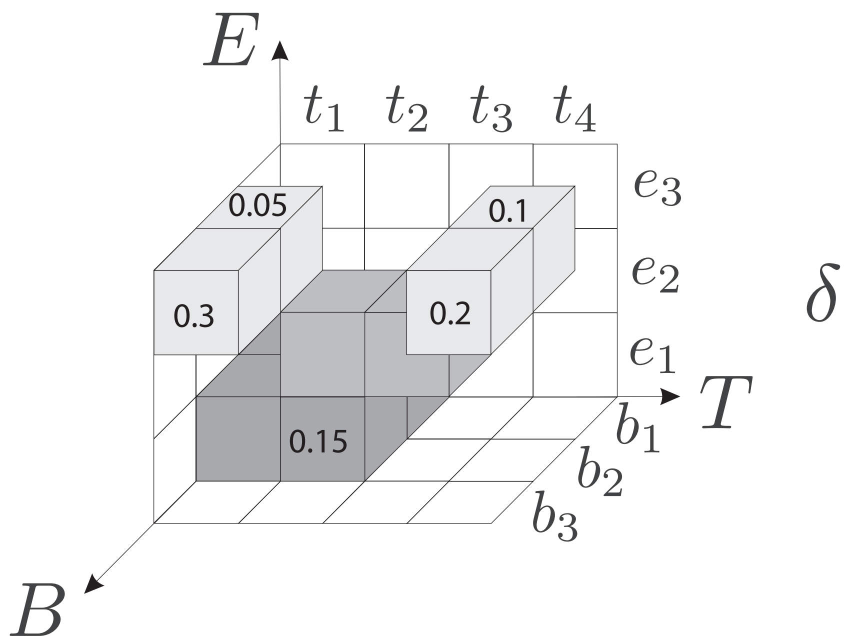
$$
\begin{aligned}
&R(A = a, B = b, C = c \mid A = a_{obs}) \\
&\quad = \min\{R(A = a, B = b, C = c), R(A = a \mid A = a_{\text{obs}})\}.
\end{aligned}
$$

Relational Evidence Propagation, Step 2

A:	color
B:	shape
C:	size

$$
\begin{aligned}
& R(C = c \mid A = a_{\text{obs}}) \\
= \quad & R\left(\bigvee_{a \in \text{dom}(A)} A = a, \bigvee_{b \in \text{dom}(B)} B = b, C = c \,\middle|\, A = a_{\text{obs}} \right) \\
\stackrel{(1)}{=} \quad & \max_{a \in \text{dom}(A)} \{ \max_{b \in \text{dom}(B)} \{ R(A = a, B = b, C = c \mid A = a_{\text{obs}}) \} \} \\
\stackrel{(2)}{=} \quad & \max_{a \in \text{dom}(A)} \{ \max_{b \in \text{dom}(B)} \{ \min\{ R(A = a, B = b, C = c), R(A = a \mid A = a_{\text{obs}}) \} \} \} \\
\stackrel{(3)}{=} \quad & \max_{a \in \text{dom}(A)} \{ \max_{b \in \text{dom}(B)} \{ \min\{ R(A = a, B = b), R(B = b, C = c), \\
& \qquad\qquad R(A = a \mid A = a_{\text{obs}}) \} \} \} \\
= \quad & \max_{b \in \text{dom}(B)} \{ \min\{ R(B = b, C = c), \\
& \qquad \underbrace{\max_{a \in \text{dom}(A)} \{ \min\{ R(A = a, B = b), R(A = a \mid A = a_{\text{obs}}) \} \} \}}_{= R(B = b \mid A = a_{\text{obs}})} \\
= \quad & \max_{b \in \text{dom}(B)} \{ \min\{ R(B = b, C = c), R(B = b \mid A = a_{\text{obs}}) \} \}.
\end{aligned}
$$

Probable car configurations



Every cube designates a value combination with its probability.

The installation rate of a value combinations is a good estimate for the probability all numbers in parts per 1000

	■	▨	□	▒
	220	330	170	280

large (300)

	■	▨	□	▒	
△	20	90	10	80	400
□	2	1	20	17	240
○	28	24	5	3	360

medium (460)

	■	▨	□	▒
△	18	81	9	72
□	8	4	80	68
○	56	48	10	6

small (240)

	■	▨	□	▒
△	2	9	1	8
□	2	1	20	17
○	84	72	15	9

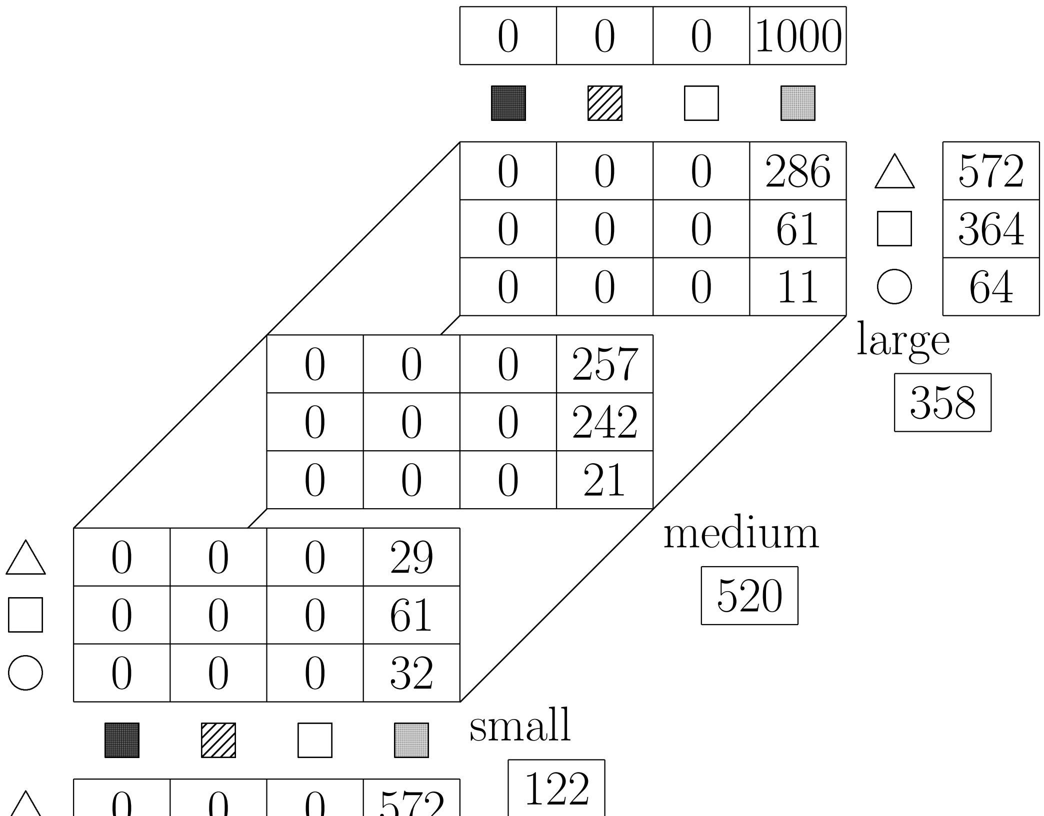
	■	▨	□	▒
△	40	180	20	160
□	12	6	120	102
○	168	144	30	18

	s	m	l
△	20	180	200
□	40	160	40
○	180	120	60

	■	▨	□	▒
large	50	115	35	100
medium	82	133	99	146
small	88	82	36	34

- The numbers state the probability of the corresponding value combination.

Reasoning: Computiong Conditional Probabilities



	■	▨	□	▒
△	0	0	0	572
□	0	0	0	364
○	0	0	0	64

all numbers in parts per 1000

	s	m	l
△	29	257	286
□	61	242	61
○	32	21	11

	■	▨	□	▒
large	0	0	0	358
medium	0	0	0	520
small	0	0	0	122

- Using the information that the given object is green.

Probabilistic Decomposition: Simple Example

- As for relational networks, the three-dimensional probability distribution can be decomposed into projections to subspaces, namely the marginal distribution on the subspace formed by color and shape and the marginal distribution on the subspace formed by shape and size.

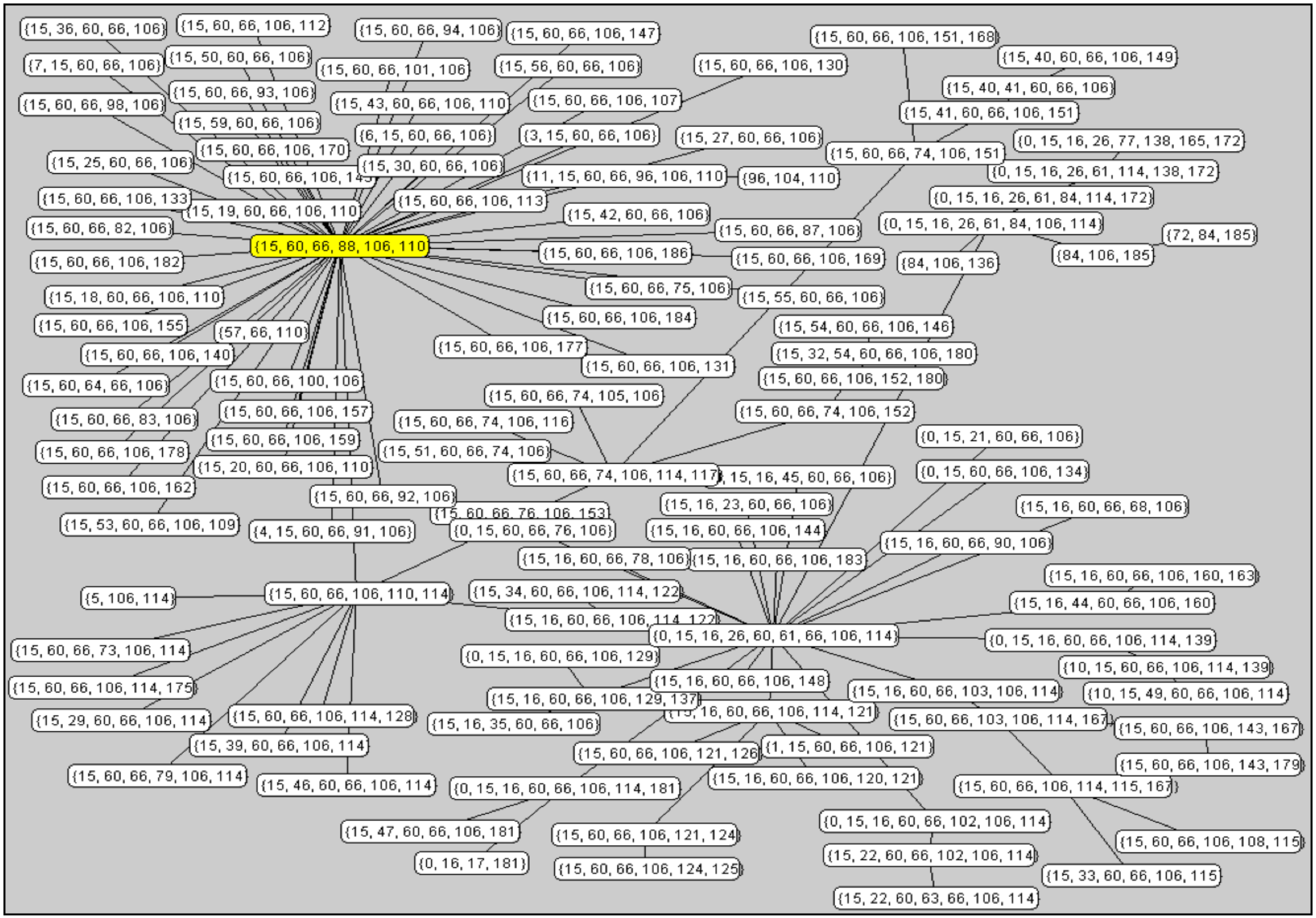
- The original probability distribution can be reconstructed from the marginal distributions using the following formulae $\forall i, j, k$:

$$\begin{aligned} P(\omega_i^{(\text{color})}, \omega_j^{(\text{shape})}, \omega_k^{(\text{size})}) &= P(\omega_i^{(\text{color})}, \omega_j^{(\text{shape})}) \cdot P(\omega_k^{(\text{size})} \mid \omega_j^{(\text{shape})}) \\ &= P(\omega_i^{(\text{color})}, \omega_j^{(\text{shape})}) \cdot \frac{P(\omega_j^{(\text{shape})}, \omega_k^{(\text{size})})}{P(\omega_j^{(\text{shape})})} \end{aligned}$$

- These equations express the *conditional independence* of attributes *color* and *size* given the attribute *shape*, since they only hold if $\forall i, j, k$:

$$P(\omega_k^{(\text{size})} \mid \omega_j^{(\text{shape})}) = P(\omega_k^{(\text{size})} \mid \omega_i^{(\text{color})}, \omega_j^{(\text{shape})})$$

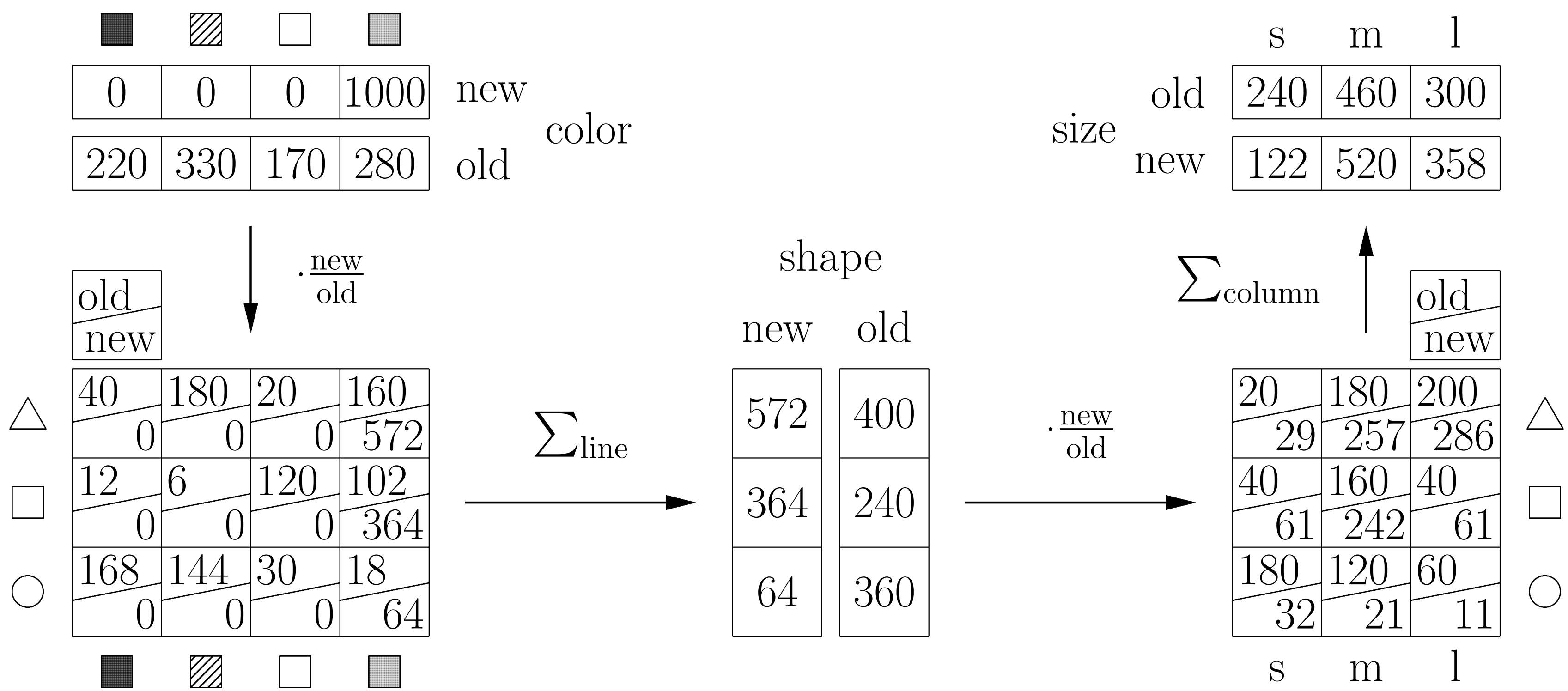
Example: VW Bora



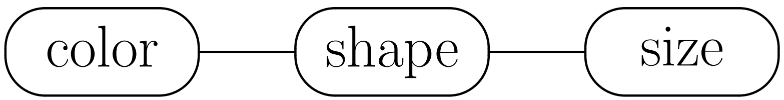
186 dim Probability space
174 Marginal Probability spaces

Reasoning with Projections

Again the same result can be obtained using only projections to subspaces (marginal distributions):



This justifies a network representation:



Probabilistic Decomposition

Definition: Let $U = \{A_1, \ldots, A_n\}$ be a set of attributes and p_U a probability distribution over U. Furthermore, let $\mathcal{M} = \{M_1, \ldots, M_m\} \subseteq 2^U$ be a set of nonempty (but not necessarily disjoint) subsets of U satisfying

$$\bigcup_{M \in \mathcal{M}} M = U.$$

p_U is called **decomposable** or **factorizable** w.r.t. $\mathcal{M}$ iff it can be written as a product of m nonnegative functions $\phi_M : \mathcal{E}_M \to \mathbb{R}_0^+$, $M \in \mathcal{M}$, i.e., iff

$$\forall a_1 \in \mathrm{dom}(A_1) : \ldots \forall a_n \in \mathrm{dom}(A_n) :$$
$$p_U\left(\bigwedge_{A_i \in U} A_i = a_i\right) = \prod_{M \in \mathcal{M}} \phi_M\left(\bigwedge_{A_i \in M} A_i = a_i\right).$$

If p_U is decomposable w.r.t. $\mathcal{M}$ the set of functions

$$\Phi_{\mathcal{M}} = \{\phi_{M_1}, \ldots, \phi_{M_m}\} = \{\phi_M \mid M \in \mathcal{M}\}$$

is called the **decomposition** or the **factorization** of p_U.
The functions in $\Phi_{\mathcal{M}}$ are called the **factor potentials** of p_U.

Conditional Independence

Definition: Let Ω be a (finite) sample space, P a probability measure on Ω, and A, B, and C attributes with respective domains $\mathrm{dom}(A)$, $\mathrm{dom}(B)$, and $\mathrm{dom}(C)$. A and B are called **conditionally probabilistically independent** given C, written $A \perp\!\!\!\perp_P B \mid C$, iff

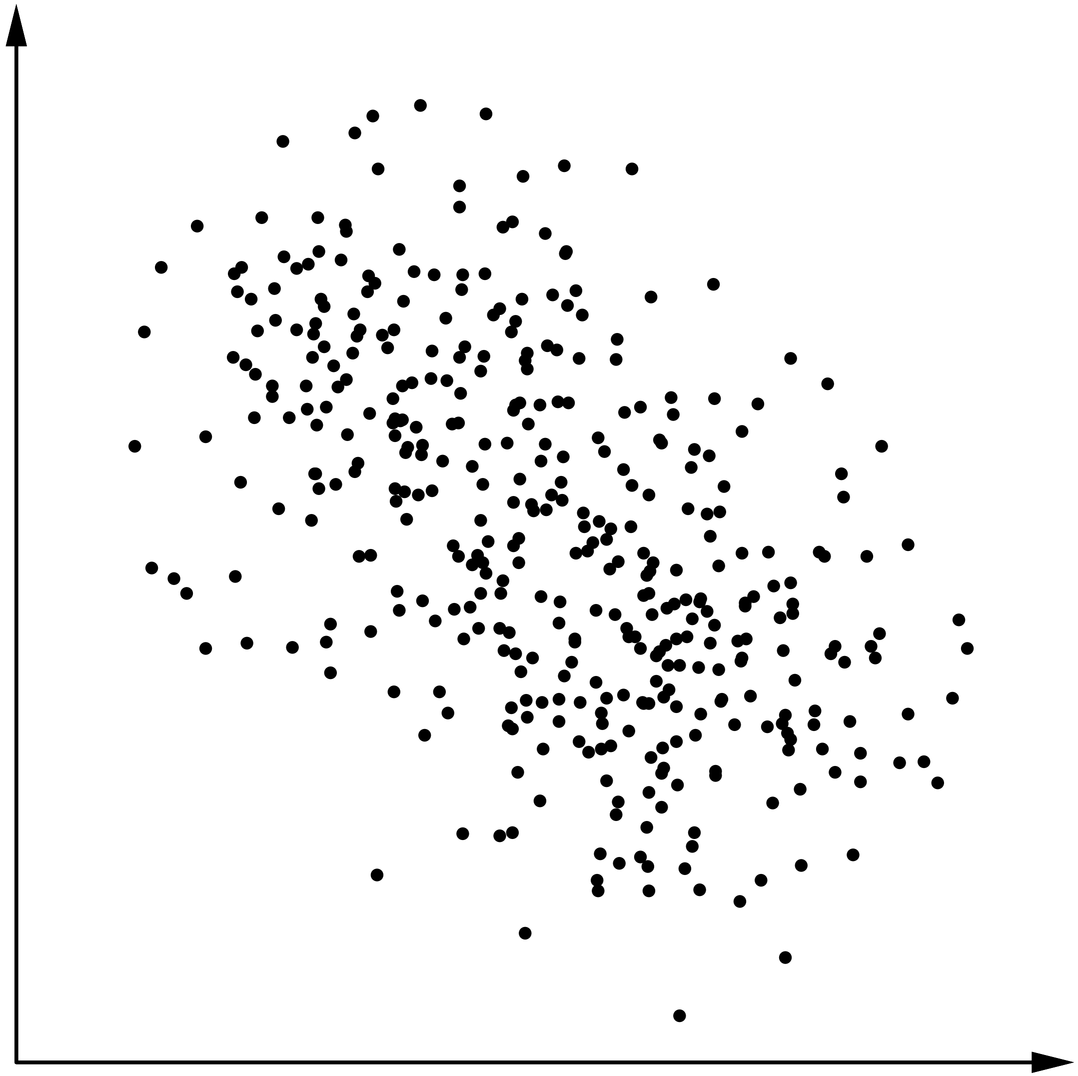
$$\forall a \in \mathrm{dom}(A) : \forall b \in \mathrm{dom}(B) : \forall c \in \mathrm{dom}(C) :$$
$$P(A = a, B = b \mid C = c) = P(A = a \mid C = c) \cdot P(B = b \mid C = c)$$

Equivalent formula (sometimes more convenient):

$$\forall a \in \mathrm{dom}(A) : \forall b \in \mathrm{dom}(B) : \forall c \in \mathrm{dom}(C) :$$
$$P(A = a \mid B = b, C = c) = P(A = a \mid C = c)$$

Conditional independences make it possible to consider parts of a probability distribution independent of others.
Therefore it is plausible that a set of conditional independences may enable a decomposition of a joint probability distribution.

Conditional Independence: An Example



Dependence (fictitious) between smoking and life expectancy.

Each dot represents one person.

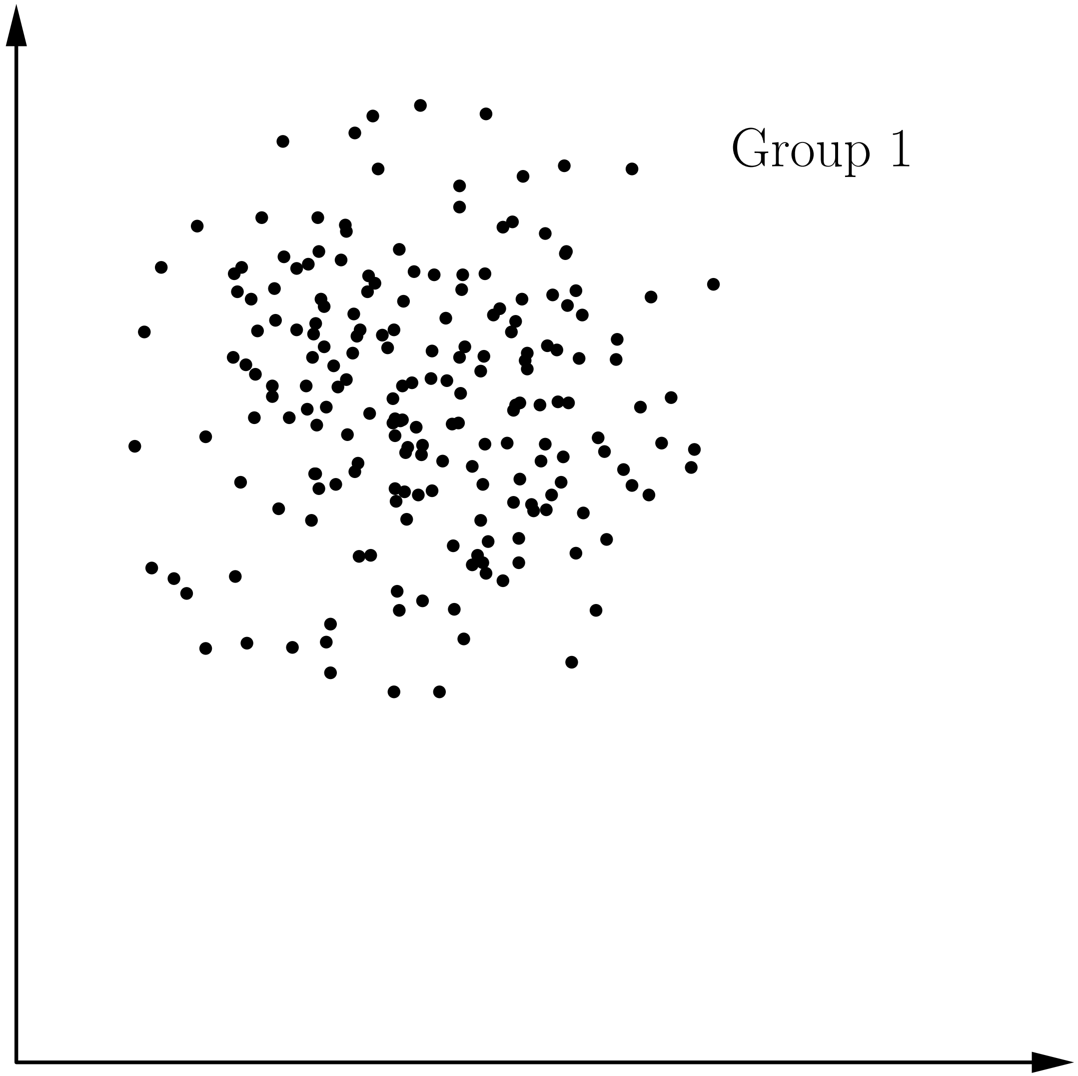
x-axis: age at death
y-axis: average number of cigarettes per day

Weak, but clear dependence:

The more cigarettes are smoked, the lower the life expectancy.

(Note that this data is artificial and thus should not be seen as revealing an actual dependence.)

Conditional Independence: An Example



Conjectured explanation:

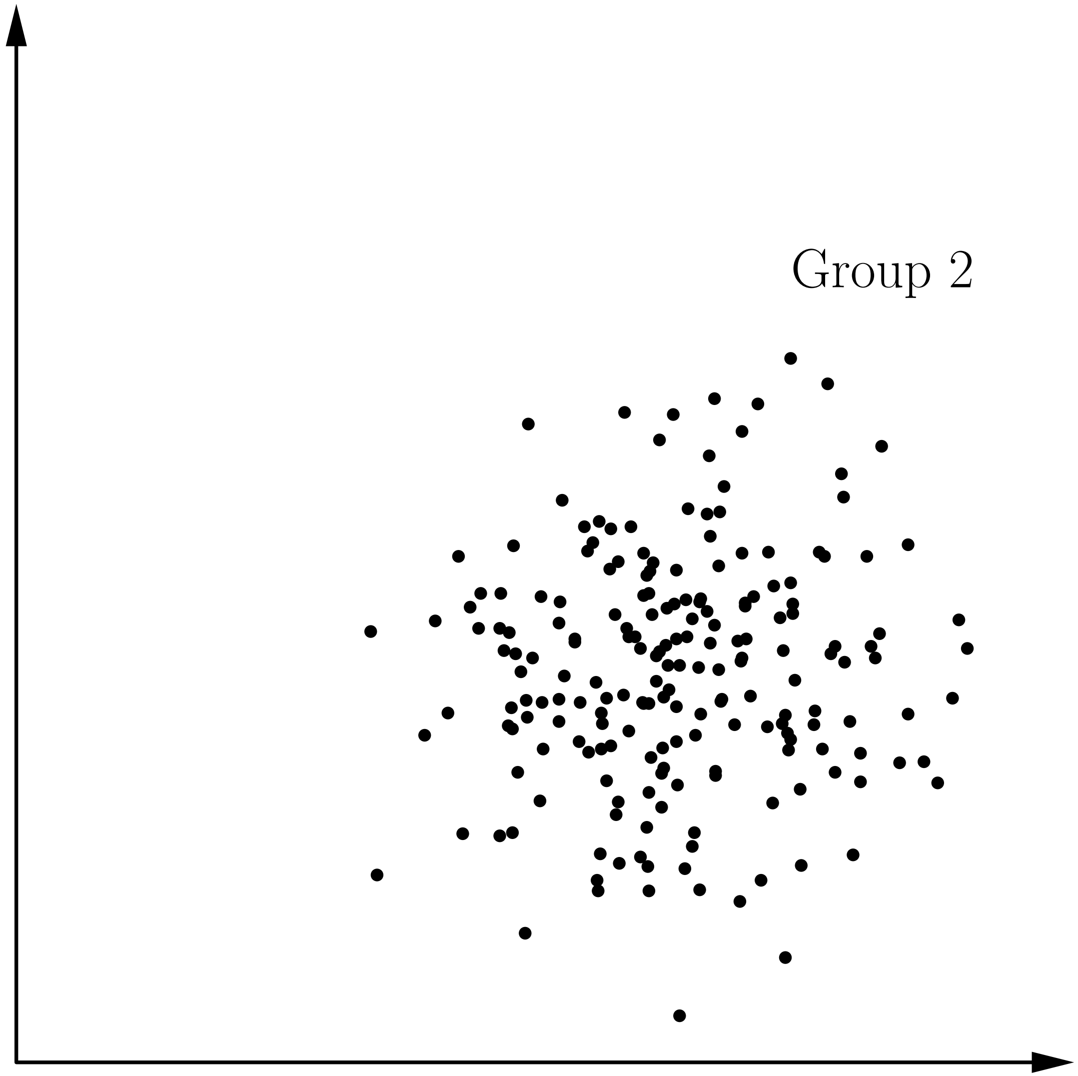
There is a common cause, namely whether the person is exposed to stress at work.

If this were correct, splitting the data should remove the dependence.

Group 1:
exposed to stress at work

(Note that this data is artificial and therefore should not be seen as an argument against health hazards caused by smoking.)

Conditional Independence: An Example



Conjectured explanation:

There is a common cause, namely whether the person is exposed to stress at work.

If this were correct, splitting the data should remove the dependence.

Group 2:
not exposed to stress at work

(Note that this data is artificial and therefore should not be seen as an argument against health hazards caused by smoking.)

Probabilistic Decomposition (continued)

Chain Rule of Probability:

$$\forall a_1 \in \mathrm{dom}(A_1) : \ldots \forall a_n \in \mathrm{dom}(A_n) :$$
$$P\left(\bigwedge\nolimits_{i=1}^{n} A_i = a_i\right) = \prod_{i=1}^{n} P\left(A_i = a_i \,\middle|\, \bigwedge\nolimits_{j=1}^{i-1} A_j = a_j\right)$$

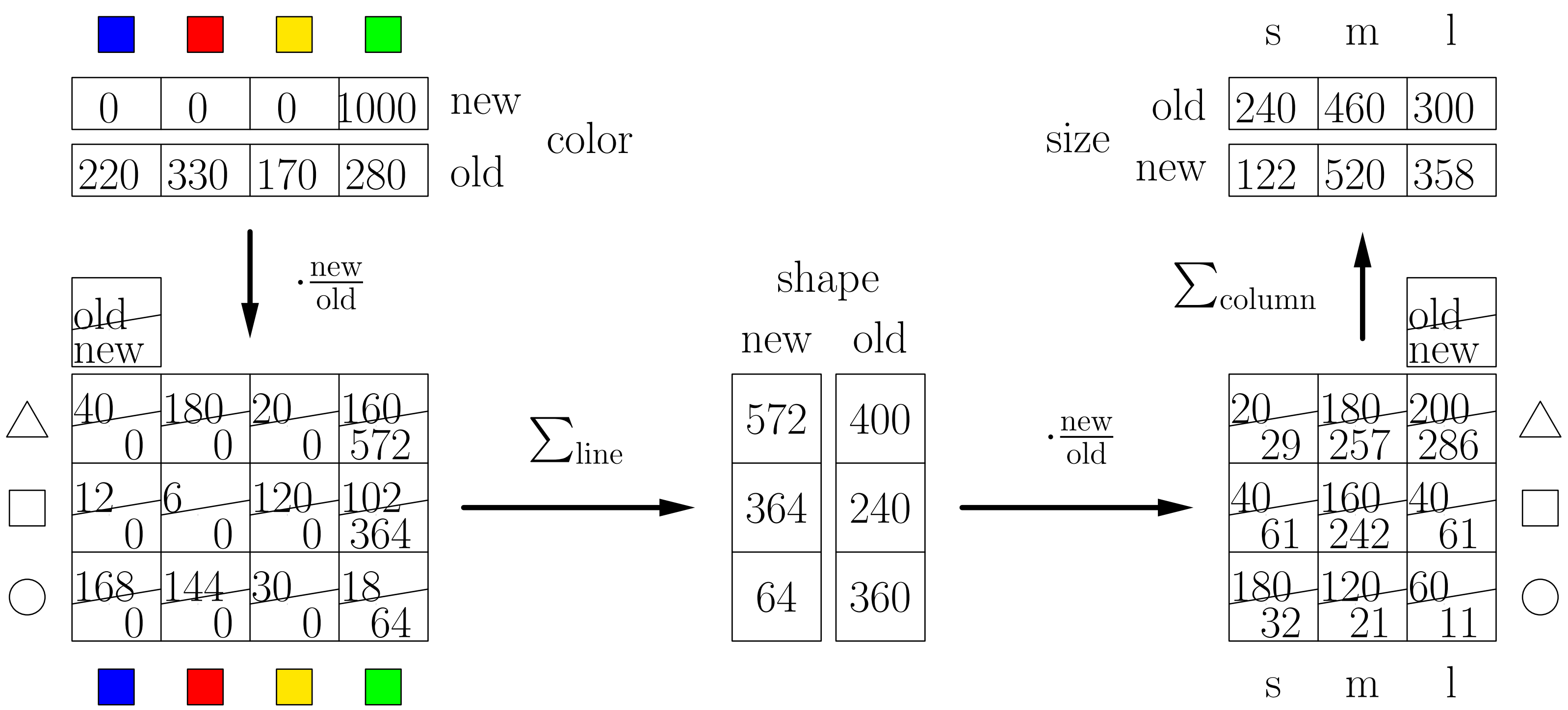
The chain rule of probability is valid in general
(or at least for strictly positive distributions).

Chain Rule Factorization:

$$\forall a_1 \in \mathrm{dom}(A_1) : \ldots \forall a_n \in \mathrm{dom}(A_n) :$$
$$P\left(\bigwedge\nolimits_{i=1}^{n} A_i = a_i\right) = \prod_{i=1}^{n} P\left(A_i = a_i \,\middle|\, \bigwedge\nolimits_{A_j \in \mathrm{parents}(A_i)} A_j = a_j\right)$$

Conditional independence statements are used to “cancel” conditions.

Due to the fact that color and size are conditionally independent given the shape, the reasoning result can be obtained using only the projections to the subspaces:



This reasoning scheme can be formally justified with probability measures.

Probabilistic Evidence Propagation, Step 1

$$
\begin{aligned}
&P(B = b \mid A = a_{\text{obs}}) \\
&= P\left(\bigvee_{a \in \text{dom}(A)} A = a, B = b, \bigvee_{c \in \text{dom}(C)} C = c \,\middle|\, A = a_{\text{obs}} \right) \\
&\overset{(1)}{=} \sum_{a \in \text{dom}(A)} \sum_{c \in \text{dom}(C)} P(A = a, B = b, C = c \mid A = a_{\text{obs}}) \\
&\overset{(2)}{=} \sum_{a \in \text{dom}(A)} \sum_{c \in \text{dom}(C)} P(A = a, B = b, C = c) \cdot \frac{P(A = a \mid A = a_{\text{obs}})}{P(A = a)} \\
&\overset{(3)}{=} \sum_{a \in \text{dom}(A)} \sum_{c \in \text{dom}(C)} \frac{P(A = a, B = b) P(B = b, C = c)}{P(B = b)} \cdot \frac{P(A = a \mid A = a_{\text{obs}})}{P(A = a)} \\
&= \sum_{a \in \text{dom}(A)} P(A = a, B = b) \cdot \frac{P(A = a \mid A = a_{\text{obs}})}{P(A = a)} \underbrace{\sum_{c \in \text{dom}(C)} P(C = c \mid B = b)}_{=1} \\
&= \sum_{a \in \text{dom}(A)} P(A = a, B = b) \cdot \frac{P(A = a \mid A = a_{\text{obs}})}{P(A = a)}.
\end{aligned}
$$

A:	color
B:	shape
C:	size

Probabilistic Evidence Propagation, Step 1 (continued)

(1) holds because of Kolmogorov's axioms.

(3) holds because of the fact that the distribution p_{ABC} can be decomposed w.r.t. the set $\mathcal{M} = \{\{A, B\}, \{B, C\}\}$. ($A$: color, B: shape, C: size)

(2) holds, since in the first place

$$P(A = a, B = b, C = c \,|\, A = a_{obs}) = \frac{P(A = a, B = b, C = c, A = a_{obs})}{P(A = a_{\text{obs}})}$$

$$= \begin{cases} \dfrac{P(A = a, B = b, C = c)}{P(A = a_{\text{obs}})}, & \text{if } a = a_{\text{obs}}, \\ 0, & \text{otherwise}, \end{cases}$$

and secondly

$$P(A = a, A = a_{\text{obs}}) = \begin{cases} P(A = a), & \text{if } a = a_{\text{obs}}, \\ 0, & \text{otherwise}, \end{cases}$$

and therefore

$$P(A = a, B = b, C = c \mid A = a_{obs})$$
$$= P(A = a, B = b, C = c) \cdot \frac{P(A = a \mid A = a_{\text{obs}})}{P(A = a)}.$$

Probabilistic Evidence Propagation, Step 2

A:	color
B:	shape
C:	size

$$
\begin{aligned}
P(C = c \mid A = a_{\text{obs}}) \\
&= P\left(\bigvee_{a \in \text{dom}(A)} A = a, \bigvee_{b \in \text{dom}(B)} B = b, C = c \,\middle|\, A = a_{\text{obs}} \right) \\
&\overset{(1)}{=} \sum_{a \in \text{dom}(A)} \sum_{b \in \text{dom}(B)} P(A = a, B = b, C = c \mid A = a_{\text{obs}}) \\
&\overset{(2)}{=} \sum_{a \in \text{dom}(A)} \sum_{b \in \text{dom}(B)} P(A = a, B = b, C = c) \cdot \frac{P(A = a \mid A = a_{\text{obs}})}{P(A = a)} \\
&\overset{(3)}{=} \sum_{a \in \text{dom}(A)} \sum_{b \in \text{dom}(B)} \frac{P(A = a, B = b) P(B = b, C = c)}{P(B = b)} \cdot \frac{P(A = a \mid A = a_{\text{obs}})}{P(A = a)} \\
&= \sum_{b \in \text{dom}(B)} \frac{P(B = b, C = c)}{P(B = b)} \underbrace{\sum_{a \in \text{dom}(A)} P(A = a, B = b) \cdot \frac{R(A = a \mid A = a_{\text{obs}})}{P(A = a)}}_{= P(B = b \mid A = a_{\text{obs}})} \\
&= \sum_{b \in \text{dom}(B)} P(B = b, C = c) \cdot \frac{P(B = b \mid A = a_{\text{obs}})}{P(B = b)}.
\end{aligned}
$$

Objective

It is often possible to exploit local constraints (wherever they may come from — both structural and expert knowledge-based) in a way that allows for a decomposition of the large (intractable) distribution $P(X_1, \ldots, X_n)$ into several sub-structures $\{C_1, \ldots, C_m\}$ such that:

The collective size of those sub-structures is much smaller than that of the original distribution P.

The original distribution P is recomposable (with no or at least as few as possible errors) from these sub-structures in the following way:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{m} \Psi_i(c_i)$$

where c_i is an instantiation of C_i and $\Psi_i(c_i) \in \mathbb{R}^+$ a *factor potential.*