Precedence-Inclusion Patterns and Relational Learning

Frank J. Oles

IBM T.J. Watson Research Center
Yorktown Heights, NY

July 9, 2003

Research Goal:

Create a new

Mathematical Theory of Pattern Generalization

containing a

Precise Definition of Best Generalization

that will be

Computable in a Practical Sense

and that will apply to

Learning How to Find Relation Instances in Text.

Illustration of Relational Learning from Text

Raw Examples

Symptoms of erythromycin ovedose may include nausea.
Other symptoms of Prozac overdose included agitation and restlessness.

Processed Examples

Symptoms of erythromycin overdose may include nausea.
Other symptoms of Prozac overdose included agitation and restlessness.

Relation Instances Exemplified -- There are 3, not 2.

overdose_symptom (erythromycin, nausea)
overdose_symptom (Prozac, agitation)
overdose_symptom (Prozac, restlessness)

Induced Rule

if pattern is
symptoms ; of ; <drug> ; overdose ; <include> ; <symptom>
then
overdose_symptom (<drug>, <symptom>)

Rule has enough complexity so it does not apply to
The most frequent side effect of erythromycin is abdominal cramping.

Relational Learning:

Pattern generalization with the aim of picking out elements of a pattern.

The induction from examples of some number of assertions that certain elements $x_1, x_2, \cdots$ of a structure S are in some particular relation $R(x_1, x_2, \cdots)$ to one another when the structure S is a specific instance of a more general pattern.

Classification:

Just tell if an instance is in a general class, so the issue of identifying particular elements becomes moot.

The special case of relational learning where the R has no arguments.

Principled Approaches to Relational Learning

Statistical Machine Learning
Requires vector representation of data.
Better-suited to classification.

Precedence-Inclusion Patterns
Computable solution to best generalization.
Applicable to text, video, images, taxonomies, etc.
Maybe concurrent processes, patterns in DNA, XML files, too.

Inductive Logic Programming
Very general.

Logical Approaches to Inductive Learning	Category-Theoretic Inductive Learning
Background Theory for a Problem	Category Suitable to a Problem
Logical Formulae Describing Objects	The Objects Themselves
Construction of Proofs that One Formula Implies Another	Construction of Morphisms from One Object to Another
Least General Generalizations: Not Unique	Minimal Most Specific Generalizations: Unique up to Isomorphism

Ideas Present in the Theory of Precedence-Inclusion Patterns

Consider

1. Mary went to the store.

The same pattern appears in

2. Last night, Mary went to the store.
3. Mary went quickly to the store.
4. Mary went to the new store.

But not in

5. Mary went to the movies after the store closed.

Conclusion: Patterns in text should involve more than identifying sequences of tokens -- even tokens with types assigned.

In defining text-based patterns, use two interrelated strict partial orders:

strictly precedes: $x < y$

Mary < went < to the store

strictly includes: $x \supset y$

to the store ⊃ store

A pattern P generalizes a pattern Q when there is a pattern-preserving map from P to Q.

Ideas Present in the Theory of Precedence-Inclusion Patterns

There are pattern-preserving maps from

1. (Mary went (to the store))

to each of

2. (Last night Mary went (to the store))
3. (Mary went quickly (to the store))
4. (Mary went (to the new store))

but not to

5. (Mary went (to the movies) (after the store closed.))

Input sentence: Carlson acquired Ask Mr Foster in 1979.

Named Entities Identified:

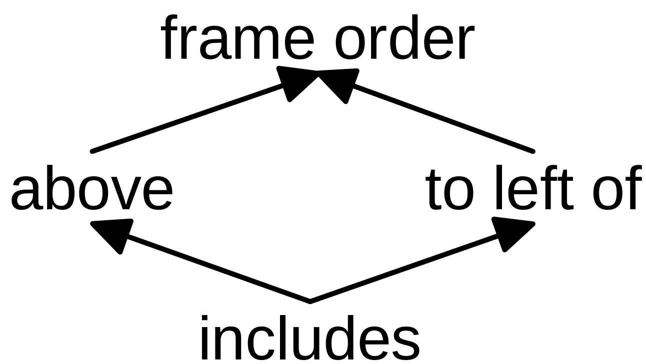
<ne lex="propn ORGANIZATION"> Carlson </ne> acquired <ne lex="propn ORGANIZATION"> Ask Mr Foster </ne> in <ne lex="n tm DATE"> 1979 </ne> .

English Slot Grammar (ESG) Parse:

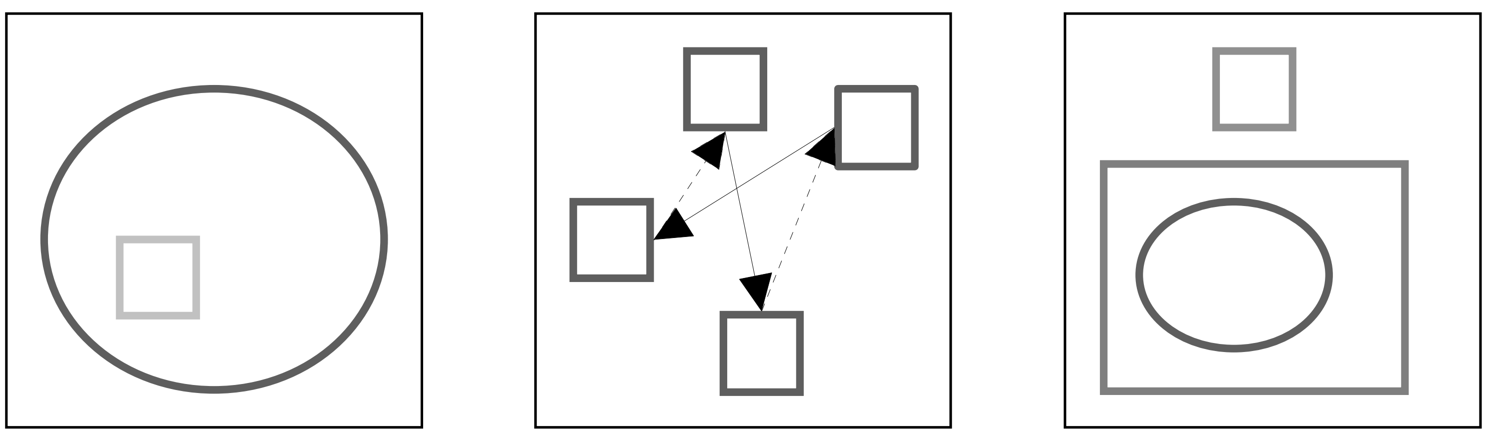
```
<ph id="2" slot="top" f="verb vfin vpast sg vsubj">
  <ph id="1" slot="subj(n)" f="noun propn sg ORGANIZATION">
    <hd w=" Carlson " c=" Carlson " s=" Carlson " a=""/>
  </ph>
  <hd w="acquired" c="acquire" s="acquire1" a="1,3"/>
  <ph id="3" slot="obj(n)" f="noun propn sg ORGANIZATION">
    <hd w=" Ask Mr Foster " c=" Ask Mr Foster " s=" Ask Mr Foster " a=""/>
  </ph>
  <ph id="4" slot="vprep" f="prep staticp timepp">
    <hd w="in" c="in" s="in1" a="5"/>
    <ph id="5" slot="objprep(n)" f="noun cn sg advnoun tm DATE">
      <hd w="  1979 " c="  1979 " s="  1979 " a=""/>
    </ph>
  </ph>
</ph>
```

Precedence-Inclusion Pattern:

- A set with a strictly partially ordered set of strict partial orders **(+ add'l structure)**
- **picture elements in video:**

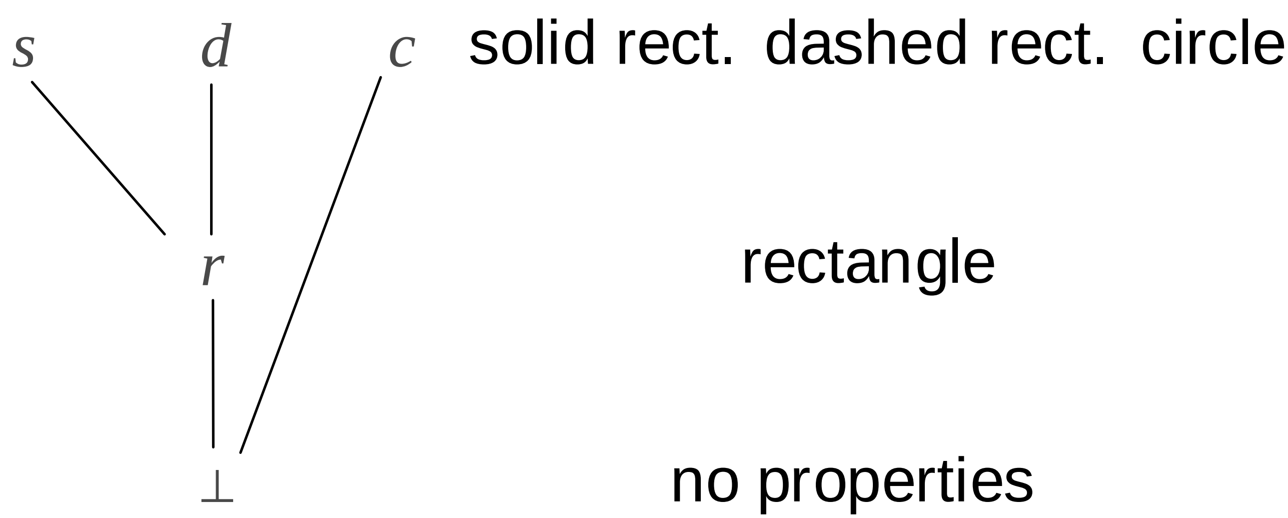


Axioms of
Interactive Transitivity
Interactive Irreflexivity



Precedence-Inclusion Pattern Prereq's:

L - a nonempty, bounded complete poset *(the property poset)*.



A - *the set of argument names* for R, *the relation of interest*.
e.g, for a binary relation $A = \{1, 2\}$.

Definition:
$\Sigma = \langle O, L, A \rangle$ is a pattern signature if
O is a strictly partially ordered set (of relation symbols)
L is a bounded complete poset (of labels or attributes),
A is a set (of argument names).

P is a Σ-pattern if
1. every $\sigma \in O$ is interpreted as a strict partial order $\prec_\sigma$ on P,
2. $\Lambda_P : P \to L$ (labeling function)
3. $\alpha_P : A \to P$ -a partial function-(argument naming)
4. Interactive transitivity: if $\tau < \sigma$ then
   $x \prec_\sigma y$ and $y \prec_\tau z$ implies $x \prec_\sigma z$, and
   $y \prec_\tau x$ and $y \prec_\sigma z$ implies $x \prec_\sigma z$.
5. Interactive irreflexivity: if $\sigma_1 < \sigma_2 < ... < \sigma_n$, then there is no $x \in P$ such that $x \prec_{\sigma_1} x' \prec_{\sigma_2} ... \prec_{\sigma_n} x$.

How common are patterns?

Example: $S = \langle a, b, c, d, e \rangle$ - a string

$W = \{\langle a, b, c\rangle, \langle a\rangle, \langle b, c\rangle, \langle b\rangle, \langle d, e\rangle\}$ - a pattern, e.g.,

$$\langle a\rangle \prec \langle b, c\rangle \prec \langle d, e\rangle,$$
$$\langle a, b, c\rangle \supset \langle b, c\rangle \supset \langle b\rangle.$$

In the example the subsequences of S are substrings of S, and they obey the usual parenthesization constraints, so we can describe W by putting balanced parentheses into S:

$$W = ((a)((b)c))(d\,e).$$

Theorem: Every set of occurrences of subsequences of a string is a purely positional classification pattern.

Constituent Structure Trees

No argument naming function
1 Labeling Function (but no bounded complete poset)
2 Relations: Precedence, Dominance (a partial order)
3 Conditions: Single Root Condition (omitted)
Nontangling Condition (omitted)
(close to interactive transitivity)
Exclusivity Condition:

$$(\forall x)(\forall y)((x \prec y \text{ or } y \prec x) \leftrightarrow (x \not\supseteq y \text{ and } y \not\supseteq x)$$

Theorem: Every constituent structure tree gives rise to a classification pattern.

Theorem: Only the left-to-right implication in the Exclusivity Condition (properly translated) holds for precedence-inclusion patterns.

Pattern-Preserving Maps

Recall def. of precedence-inclusion pattern:

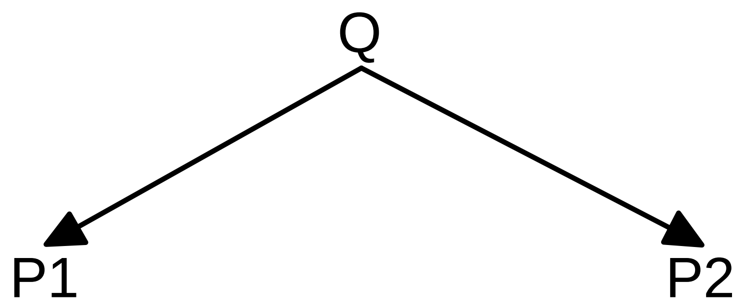
$\prec_\sigma$, total function $\Lambda : P \to L$, partial function $a : A \to P$

Let P and Q be patterns.

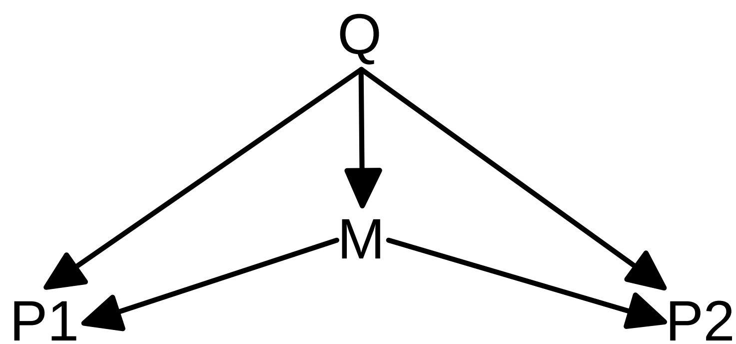
If $\mathrm{dom}\ a_P \not\subseteq \mathrm{dom}\ a_Q$, then there are no pattern-preserving maps from P to Q. Otherwise, a function $f : P \to Q$ is *a pattern-preserving map* if, for all $x, y \in P$, for all relation symbols σ, and for all argument names $a \in \mathrm{dom}\ a_P$,

1. $x \prec_{\sigma,P} y$ implies $f(x) \prec_{\sigma,Q} f(y)$,
2. $\Lambda_P(x) \leq \Lambda_Q(f(x))$, and <-- delicate point: not equality
3. $f(a_P(a)) = a_Q(a)$.

Q Generalizes P1 and P2:



A Most Specific Generalization M of P1 and P2:



Minimal Most Specific Generalization:

No proper subpattern of M is a most specific generalization of P1 and P2.

Main Theorem & Retracts

Theorem: Every nonempty finite set of finite patterns has a minimal most specific generalization, which is unique up to isomorphism.

How to compute minimal most specific generalizations?

A pattern-preserving map $g : P \to P$ is a *retraction of* P if it is *idempotent*, i.e., for all $x \in P$, $g(g(x)) = g(x)$.

A subpattern of P is a *retract of* P if it is the image of a retraction of P.

A retract of P is a *proper retract of* P if it is proper subset.

Analagous concepts exist in topology and domain theory.

How to compute the MMSG

Let $\{P_1, P_2, \cdots, P_n\}$ be a finite set of finite patterns.

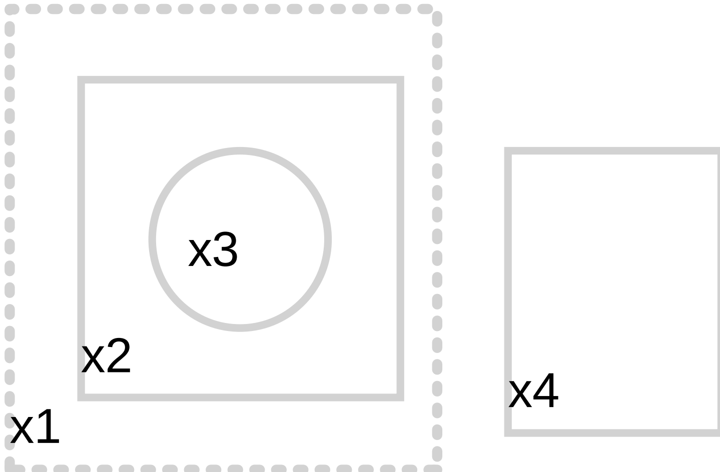
Minimal Most Specific Generalization Procedure:

$M := P_1 \times P_2 \times \cdots \times P_n$;
while there exists a proper retract Q of M
do $M := Q$;
return M;

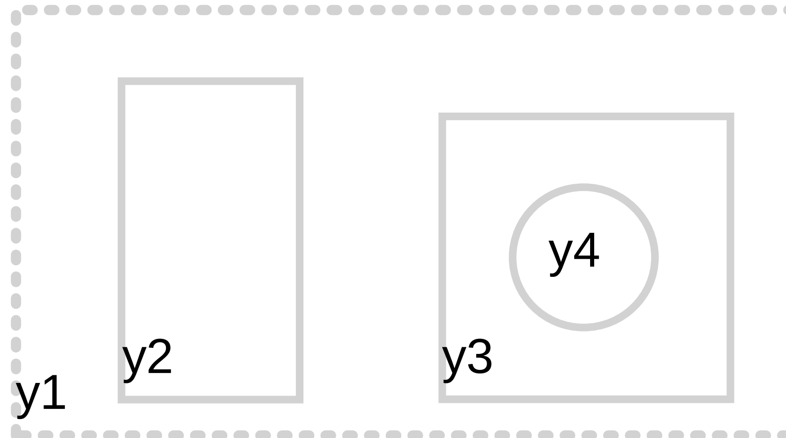
Finding proper retracts appears to be easy in practice.

So does determining if one pattern is a generalization of another because there are so many constraints on where an element can go under a pattern-preserving map.

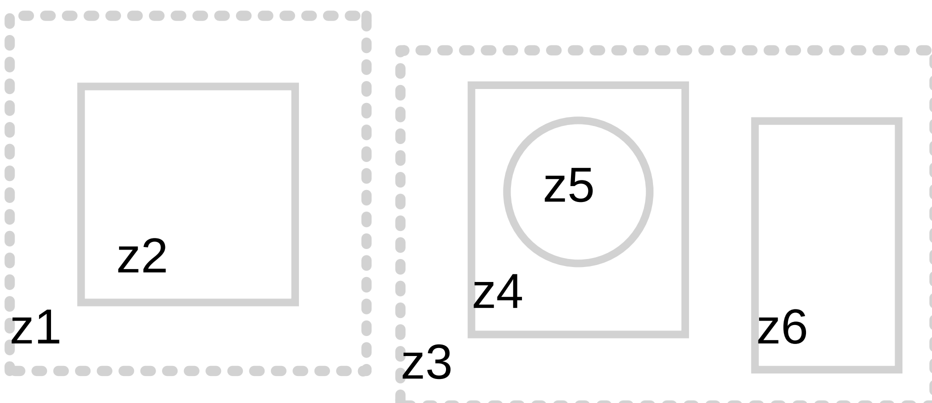
Geometric Patterns



X, in which R(x2,x4) holds



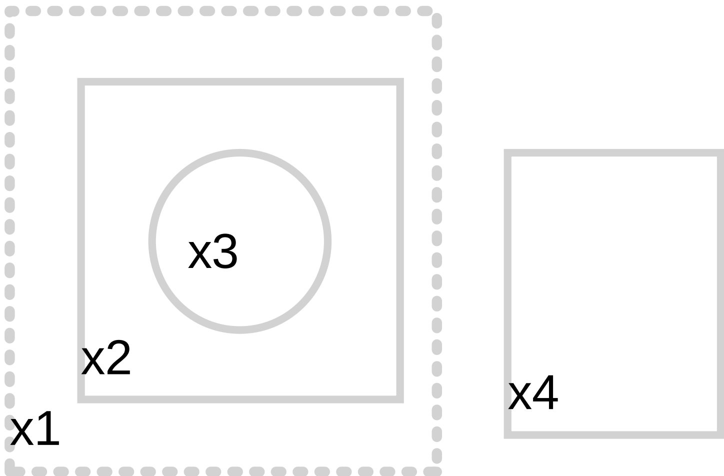
Y, in which R(y2,y3) holds



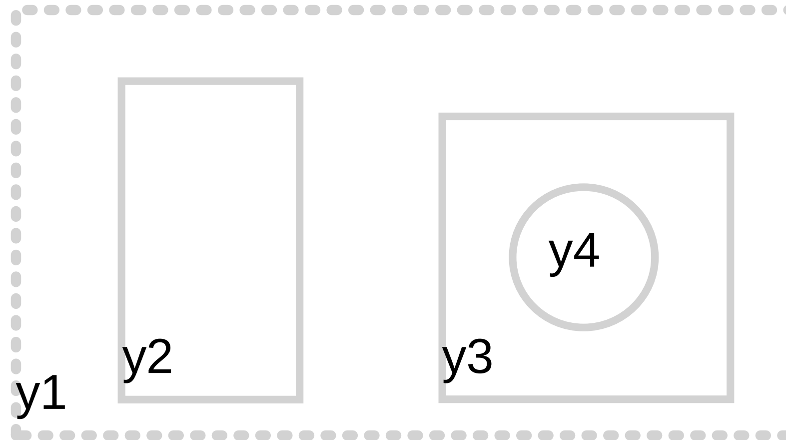
Z. Does R(a,b) hold for some a,b?

On the basis of the evidence provided by the patterns X and Y, can we claim in a principled way that another instance of the relation R occurs in the pattern Z?

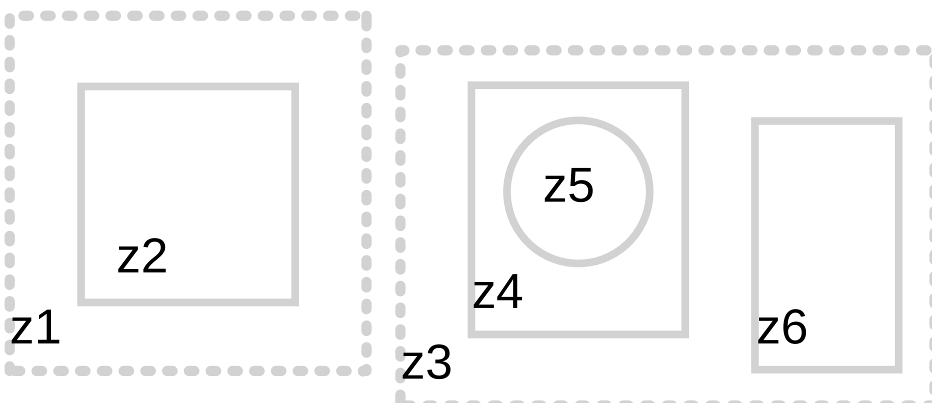
Geometric Patterns



X, in which R(x2,x4) holds



Y, in which R(y2,y3) holds



Z. Does R(a,b) hold for some a,b?

Consider some possibities:

R(z2,z5)

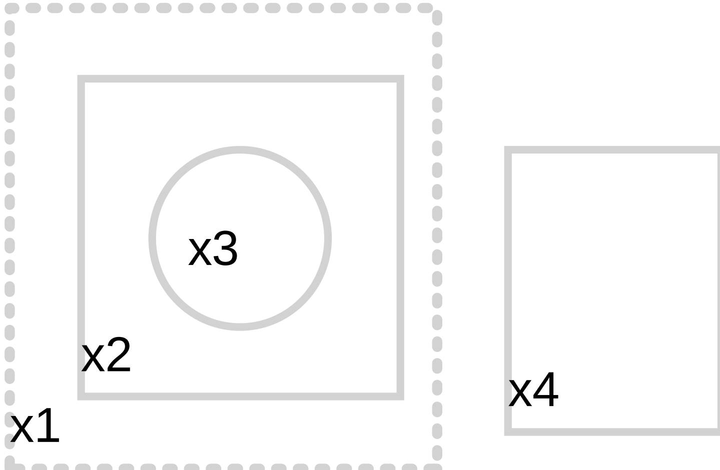
R(z1,z3)

R(z2,z4)

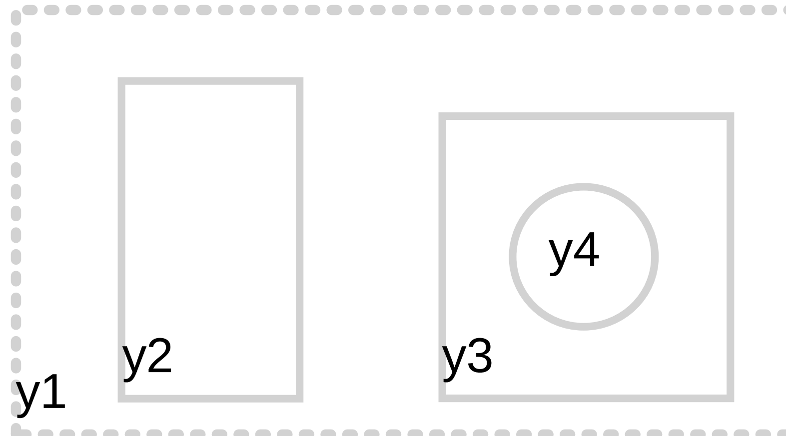
R(z2,z6)

R(z4,z6)

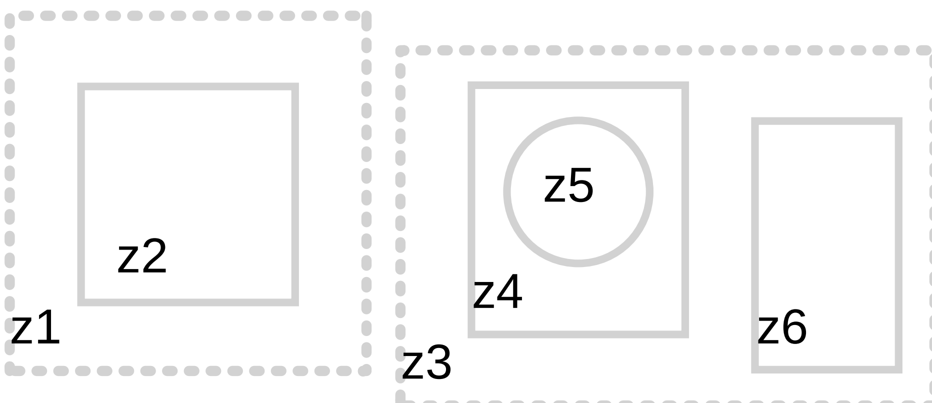
Geometric Patterns



X, in which R(x2,x4) holds



Y, in which R(y2,y3) holds



Z. Does R(a,b) hold for some a,b?

Consider some possibilities:

R(z2,z5)

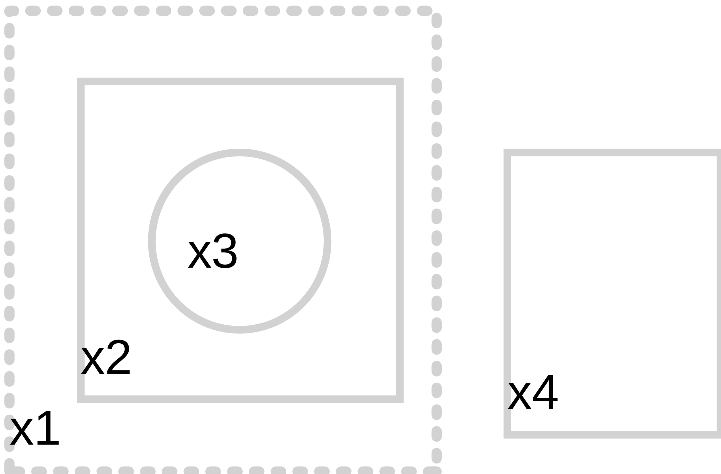
R(z1,z3)

R(z2,z4)

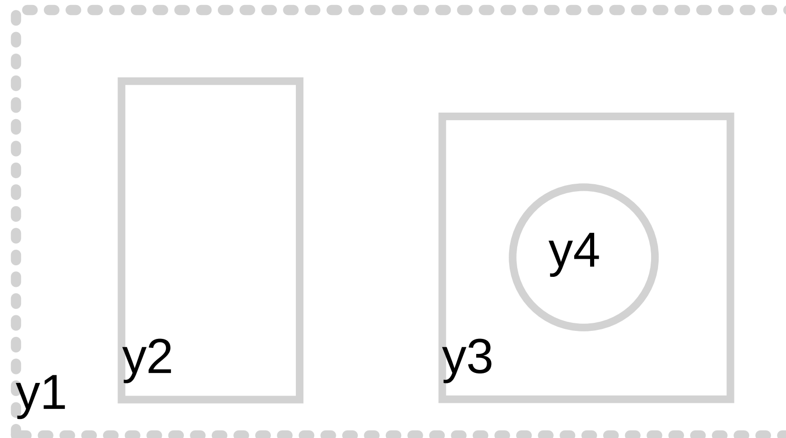
R(z2,z6)

R(z4,z6)

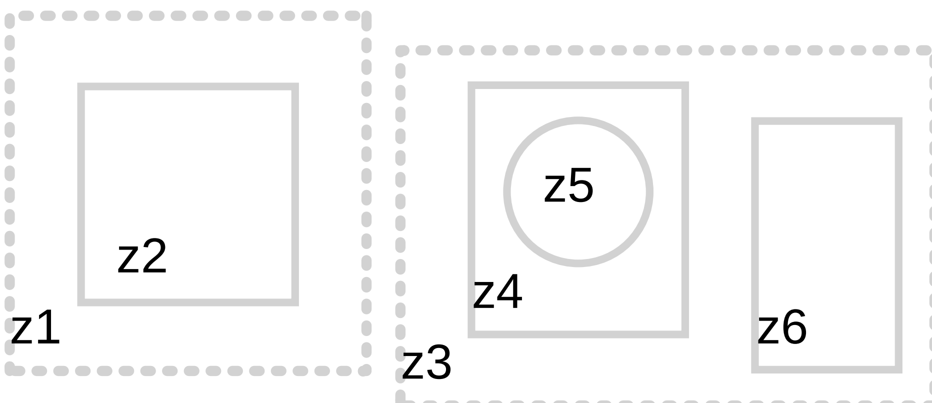
Geometric Patterns



X, in which R(x2,x4) holds



Y, in which R(y2,y3) holds



Z. Does R(a,b) hold for some a,b?

Consider some possibities:

R(z2,z5) 

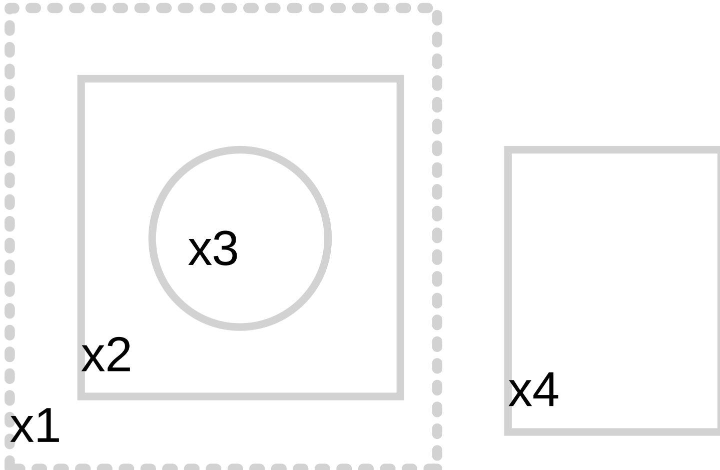
R(z1,z3) 

R(z2,z4)

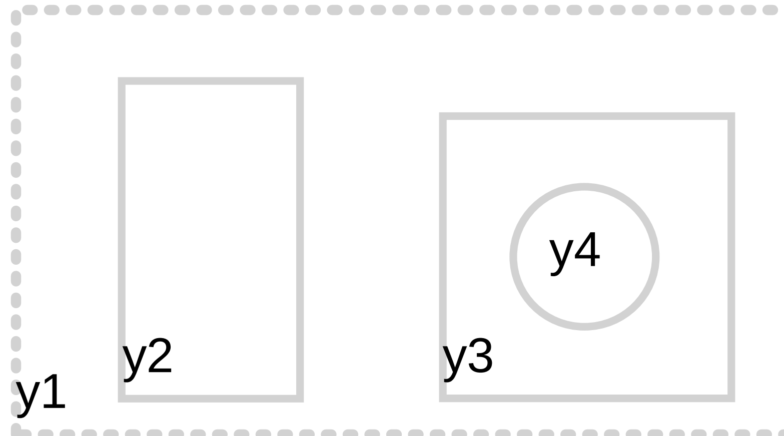
R(z2,z6)

R(z4,z6)

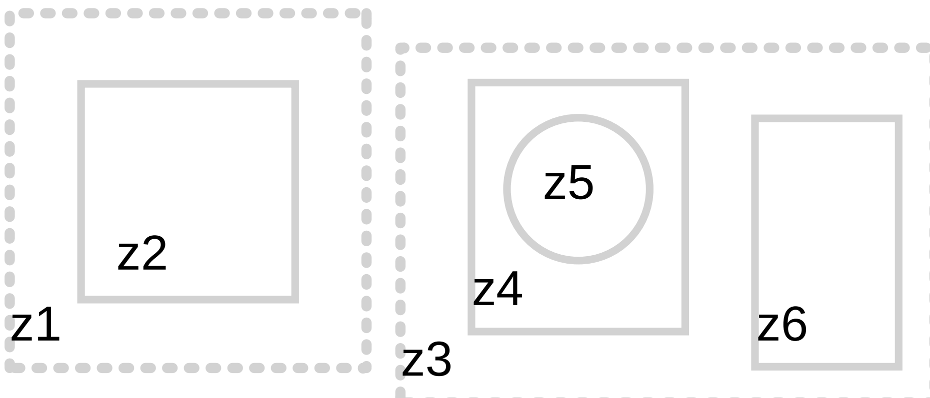
Geometric Patterns



X, in which R(x2,x4) holds



Y, in which R(y2,y3) holds



Z. Does R(a,b) hold for some a,b?

Consider some possibilities:

R(z2,z5) ✗

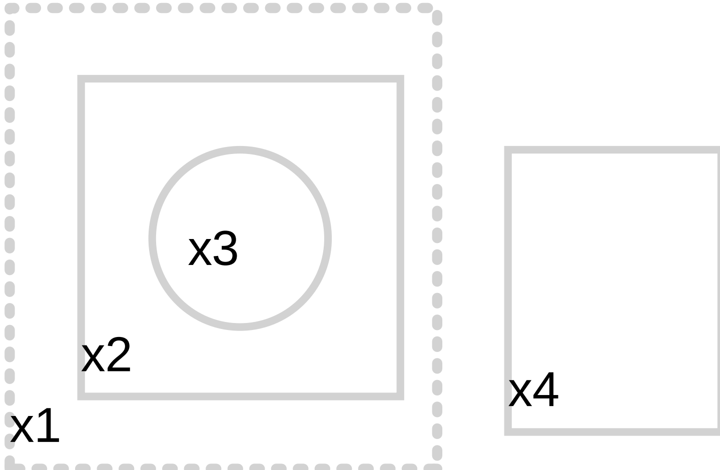
R(z1,z3) ✗

R(z2,z4) ✗

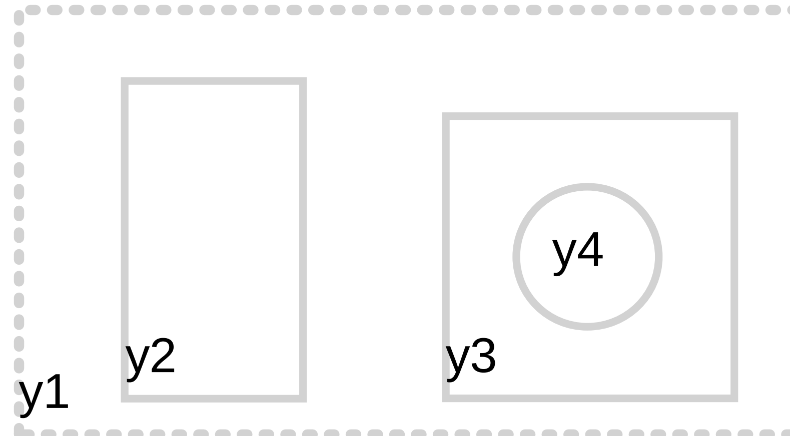
R(z2,z6)

R(z4,z6)

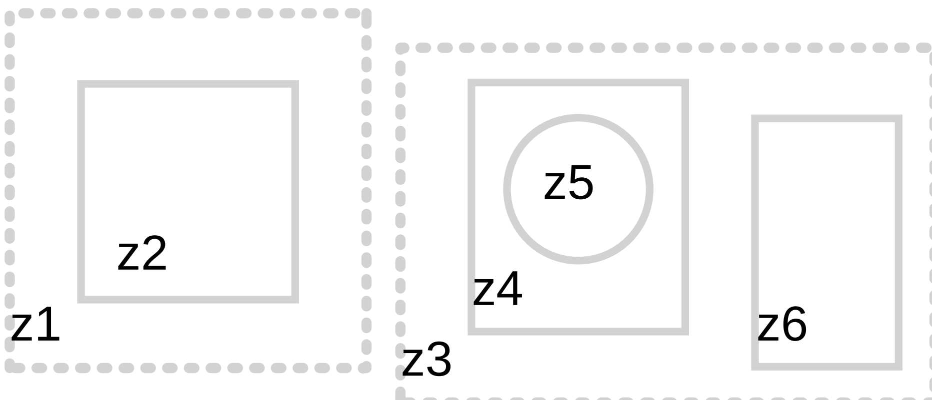
Geometric Patterns



X, in which R(x2,x4) holds



Y, in which R(y2,y3) holds



Z. Does R(a,b) hold for some a,b?

Consider some possibilities:

R(z2,z5) ✗

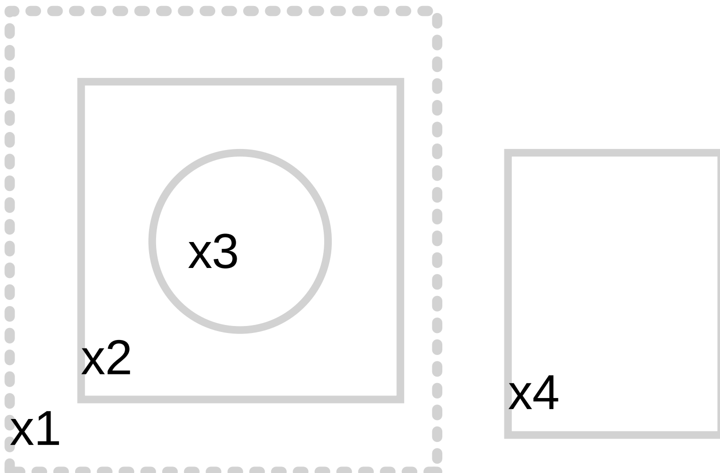
R(z1,z3) ✗

R(z2,z4) ✗

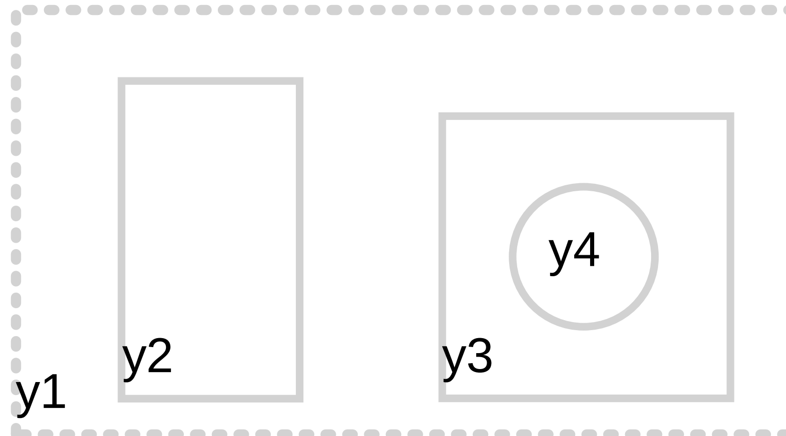
R(z2,z6) ✗

R(z4,z6)

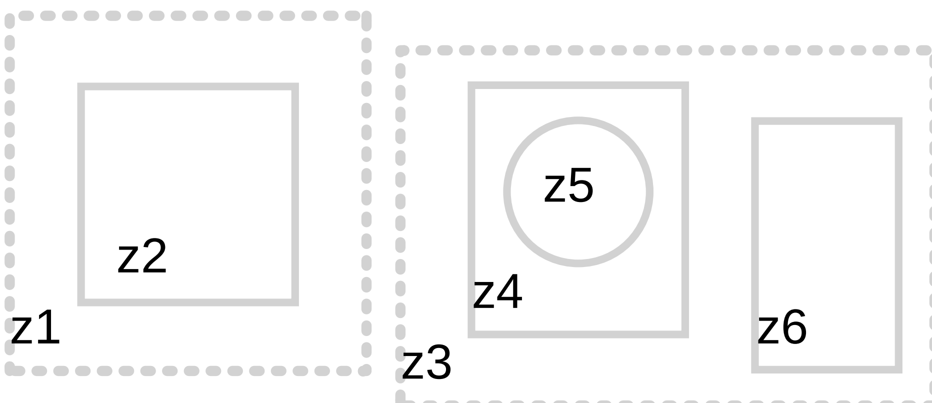
Geometric Patterns



X, in which R(x2,x4) holds



Y, in which R(y2,y3) holds

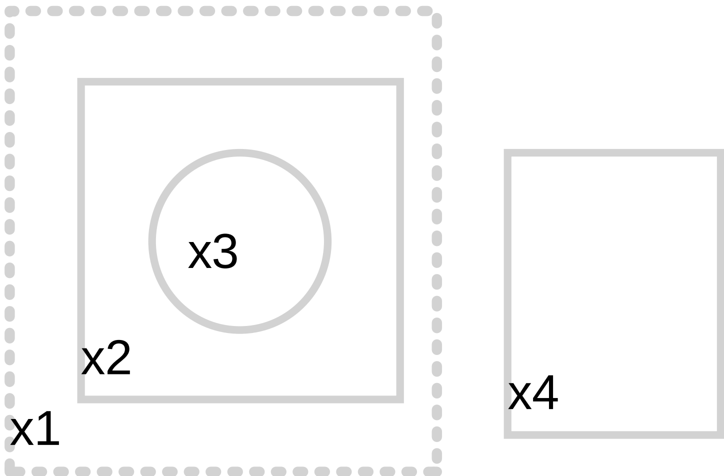


Z. Does R(a,b) hold for some a,b?

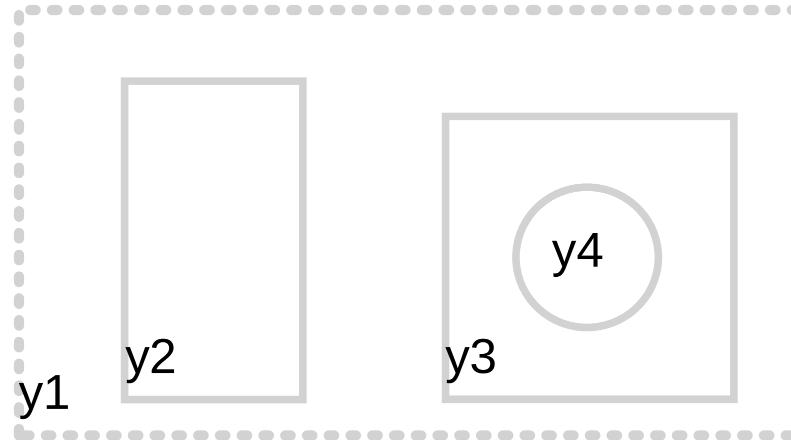
Consider some possibilities:

- R(z2,z5) ✗
- R(z1,z3) ✗
- R(z2,z4) ✗
- R(z2,z6) ✗
- R(z4,z6) ???

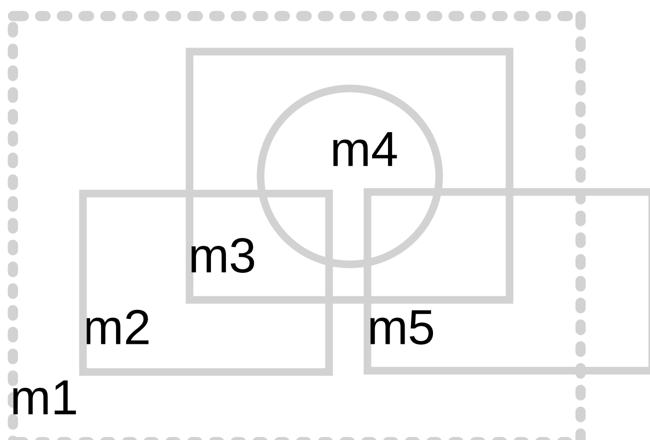
Geometric Patterns



X, in which R(x2,x4) holds



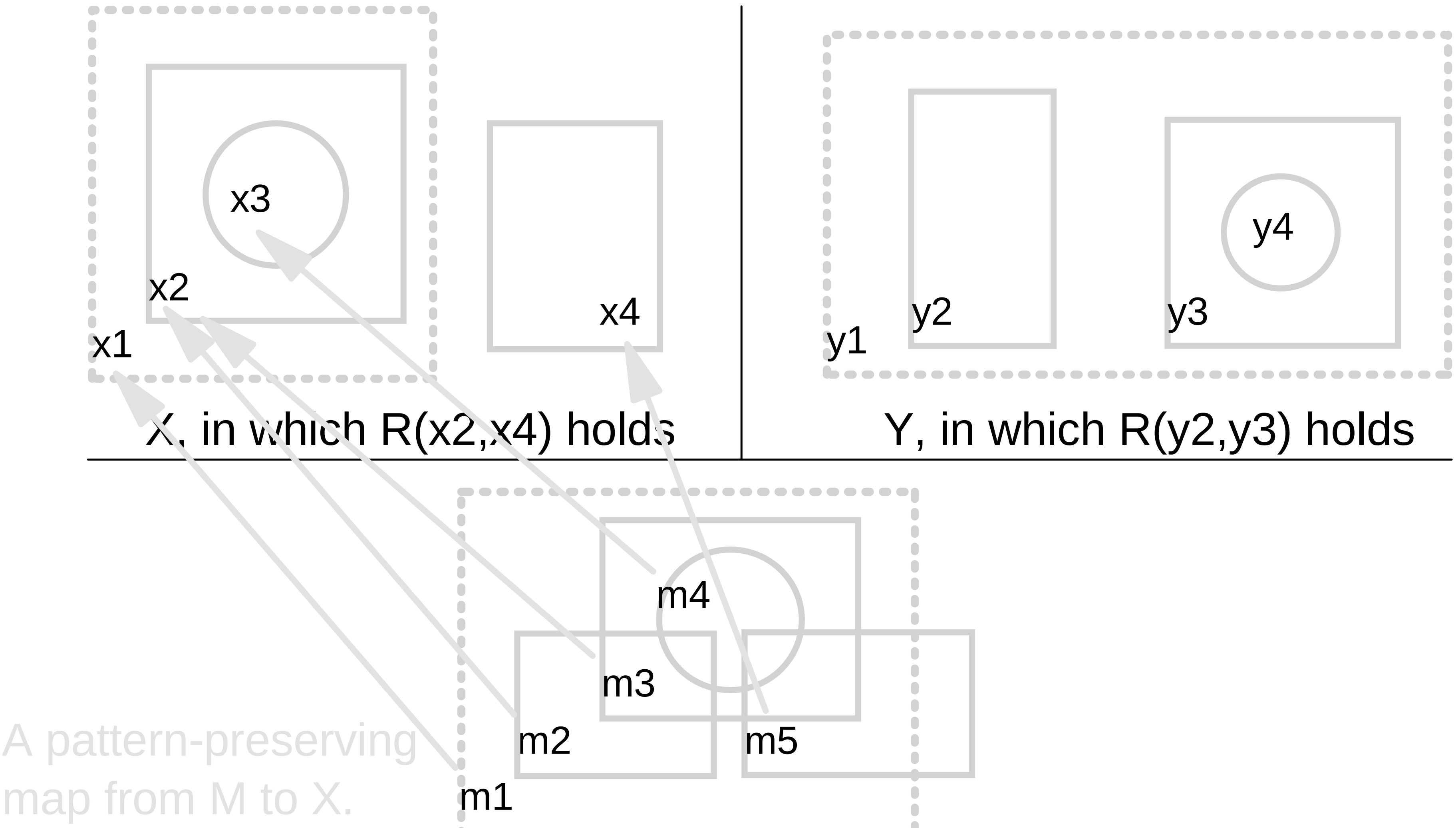
Y, in which R(y2,y3) holds



M cannot be obtained by removing from, say, X those elements inconsistent with Y.

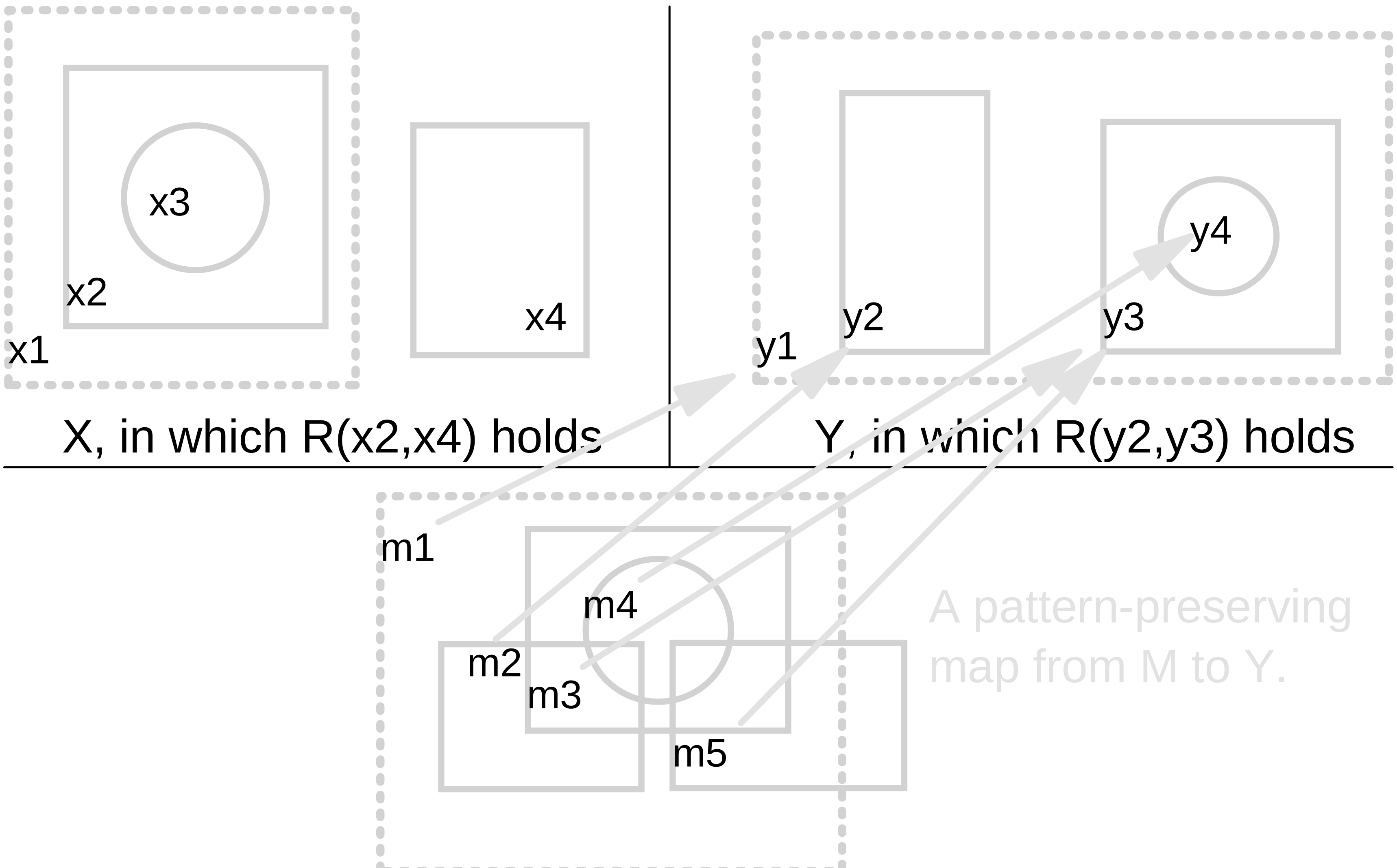
M, the miminal most specific generalization of X and Y, in which R(m2,m5) holds

Geometric Patterns



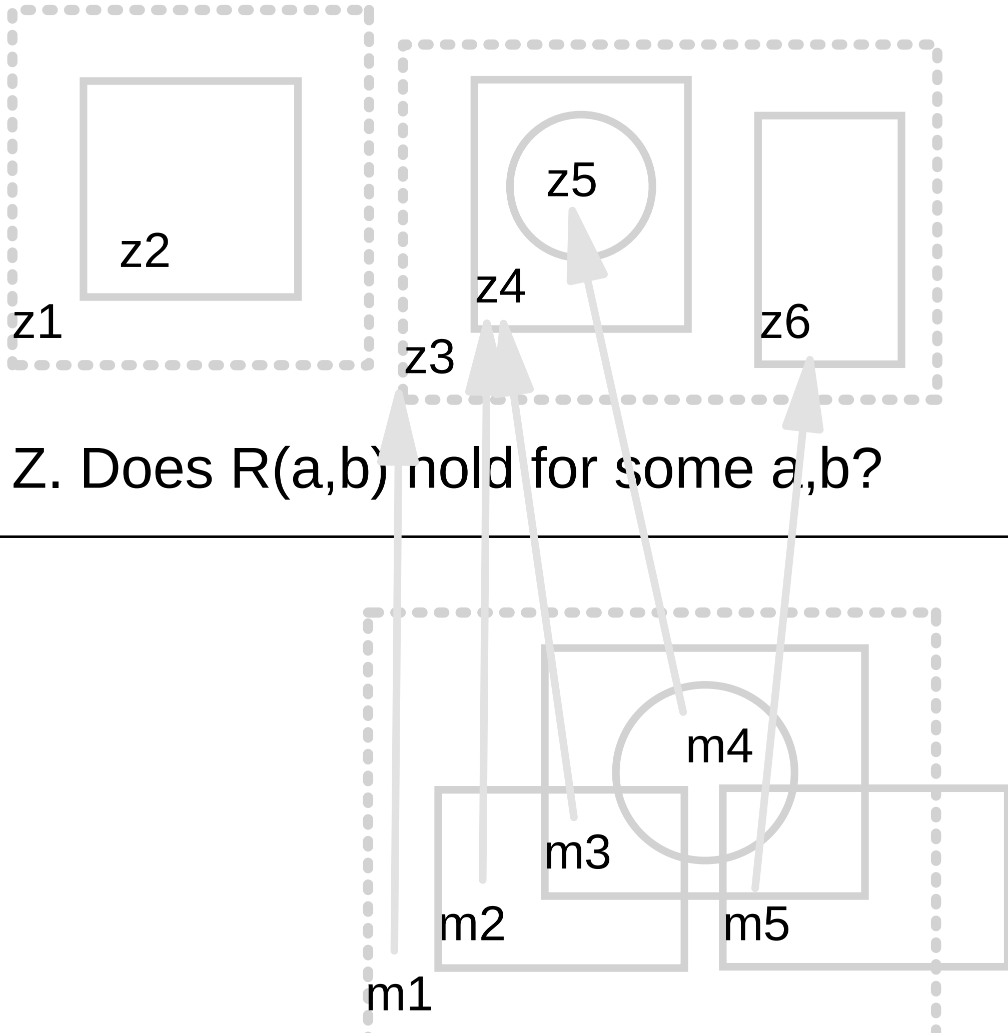
M, the miminal most specific generalization of X and Y,
in which R(m2,m5) holds

Geometric Patterns



M, the miminal most specific generalization of X and Y,
in which R(m2,m5) holds

Geometric Patterns



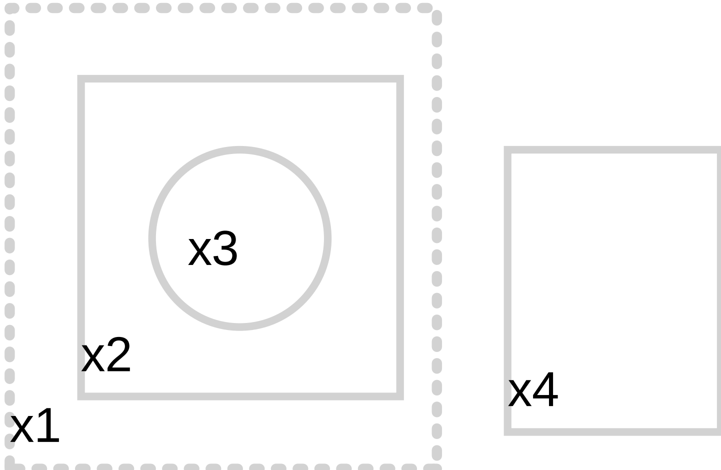
R(z4,z6) holds because
m2 → z4
m5 → z6

Z. Does R(a,b) hold for some a,b?

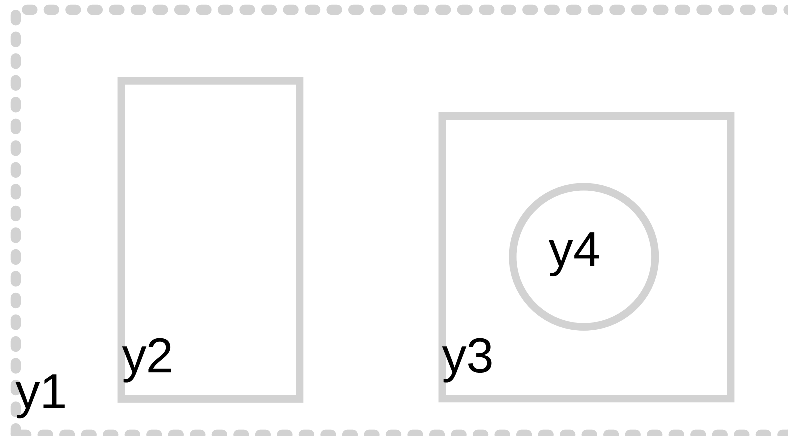
A pattern-preserving map from M to Z.

M, the miminal most specific generalization of X and Y, in which R(m2,m5) holds

Geometric Patterns

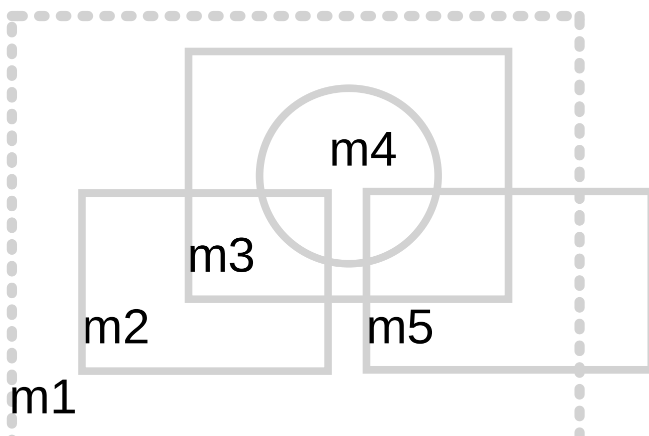


X = (x1 (x2 (x3))) (x4)



Y = (y1 (y2) (y3 (y4)))

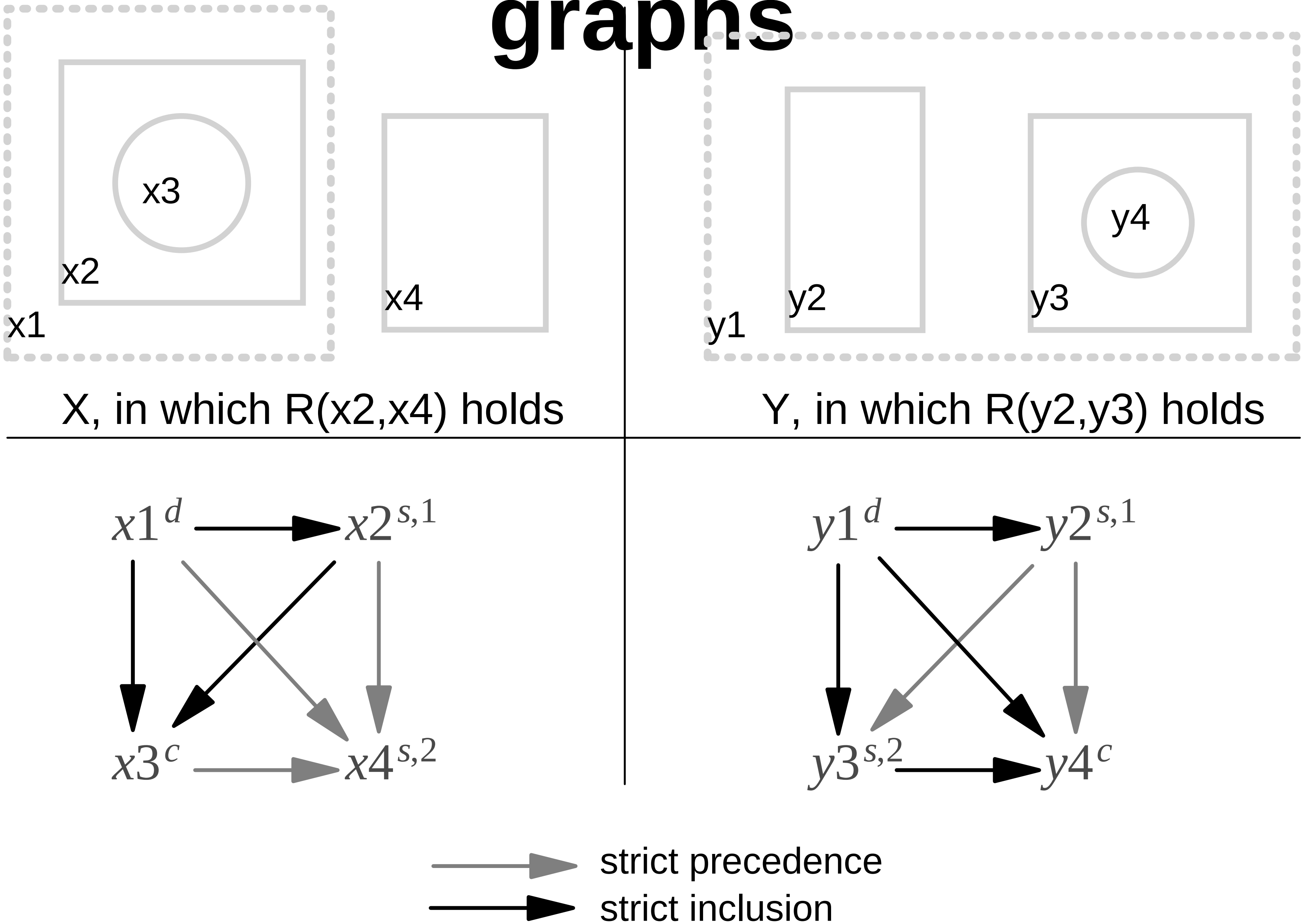
M cannot be described by inserting balanced parentheses into a string.



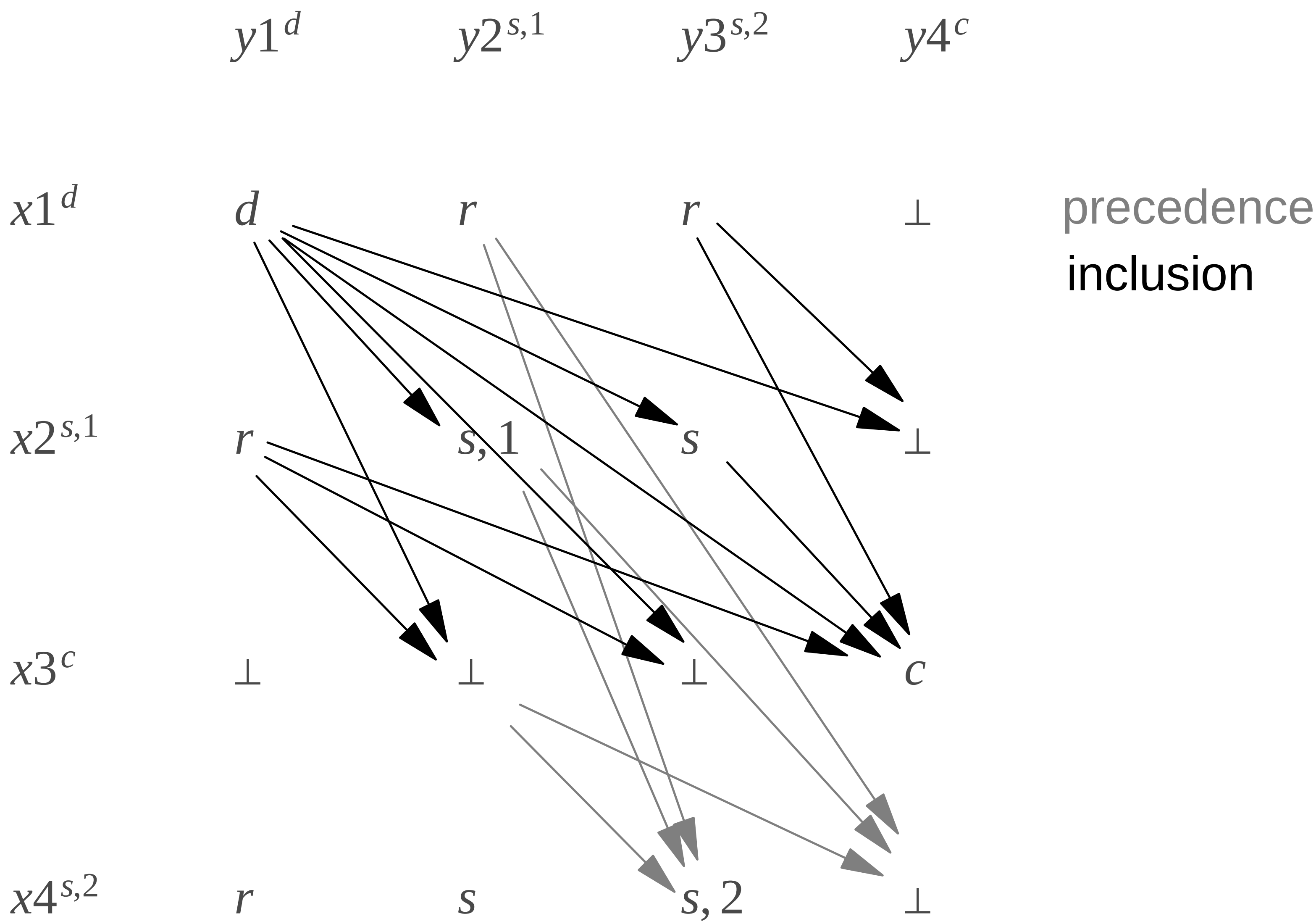
The best generalization of a set of trees may fail to be a tree.

M, the miminal most specific generalization of X and Y, in which R(m2,m5) holds

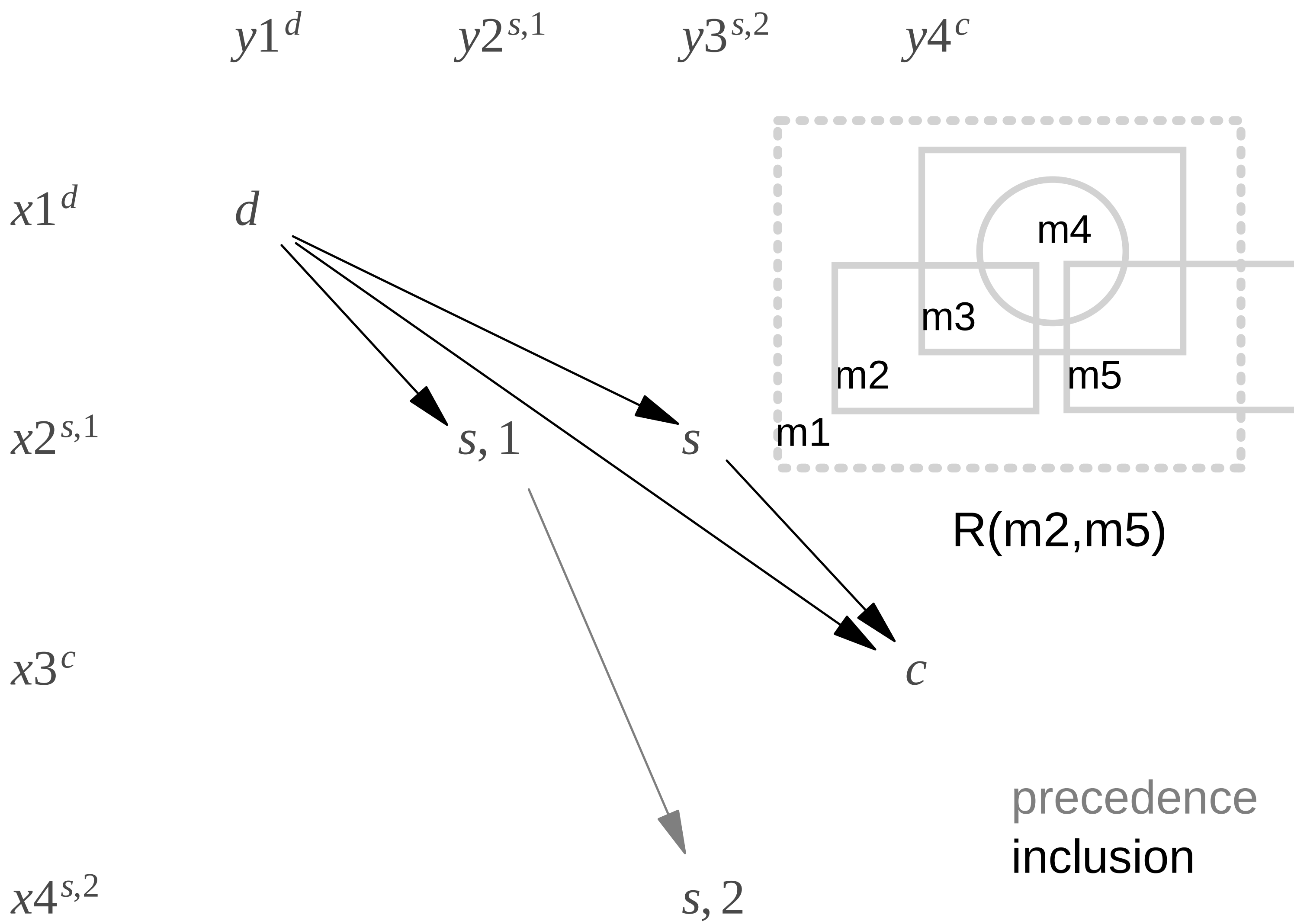
X and Y as directed acyclic graphs



A Sequence of Retracts: $M_1 = X \times Y$



$M = M_{12}$ has no proper retraction.



Completed paper:

Precedence-Inclusion Patterns and Relational Learning

IBM Research Report RC22786.

Theorem:

Every finite set of finite precedence-inclusion patterns has a minimal most specific generalization that is unique up to isomorphism.

Other subjects:

New symmetric monoidal closed categories.

How to use category theory to come up with the right definition.

"Power set" operators for patterns.

Definition & properties of infinite treelike patterns.

Backup slides

Nonempty Power Set Patterns

Definition:
P has finitely limited chains if, for all relation symbols $\sigma \in O$, the strict partial order $\prec_{\sigma,P}$ either has no infinite descending chains (i.e., is well-founded) or has no infinite ascending chains (i.e., is co-well-founded).

Theorem:
Let $\Sigma = \langle O, L, A \rangle$ be a pattern signature and let P be a Σ-pattern having finitely limited chains. Then $N(P)$, the set of nonempty subsets of P, can be made into a Σ-pattern such that

$$X \prec_{\sigma,N(P)} Y \text{ iff } (\forall x \in X)(\exists y \in Y)(x \prec_{\sigma,P} y) \text{ and } (\forall y \in Y)(\exists x \in X)(x \prec_{\sigma,P} y) .$$

Associativity of Tensor and Hom

logic	$(p \wedge q) \to r$	$p \to (q \to r)$
nonnegative integers	$(p \cdot q) \to r$	$p \to (q \to r)$
sets, K-spaces	$(P \times Q) \to R$	$P \to (Q \to R)$
vector spaces, abelian groups,	$(P \otimes Q) \to R$	$P \to (Q \to R)$

The last also holds for P-I patterns.

Tensor product $P \otimes_{|\Sigma|} Q$ in $|\Sigma|$:

$$\langle x, y\rangle \prec_{\sigma, P\otimes Q} \langle x', y'\rangle \text{ iff } \begin{array}{lll} (x \prec_{\sigma,P} x' & \text{and} & y \prec_{\sigma,Q} y'), \text{ or} \\ (x = x' & \text{and} & y \prec_{\sigma,Q} y'), \text{ or} \\ (x \prec_{\sigma,P} x' & \text{and} & y = y'). \end{array}$$

Theorem:

Let $\Sigma = \langle O, L, A\rangle$ be a pattern signature in which L is a bounded complete lattice. For all nonempty Σ-patterns $P, Q,$ and R, there is a natural isomorphism

$$f \mapsto f' : \ ((P \otimes_{|\Sigma|} Q) \to_{|\Sigma|} R) \quad \longrightarrow \quad (P \to_{\Sigma} (Q \to_{\Sigma} R))$$

given by $$f'(x)(y) = f\langle x, y\rangle.$$