Automating the development of metabolic network models

supervisor:
Ross King

collaborators:
Oliver Ray, Bristol University
Stefano Bragaglia, Bristol University

author:
Robert Rozanski

sponsors:
MANCHESTER 1824
EPSRC Engineering and Physical Sciences Research Council

motivation:

biological systems are large and complex
(yeast 7.11: ~3.7k species; ~4k reactions)

Hnin W. Aung, Susan A. Henry, and Larry P. Walker, *Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism*, 2013

motivation:

biological systems are large and complex
(yeast 7.11: ~3.7k species; ~4k reactions)

bottleneck for manual model development

Hnin W. Aung, Susan A. Henry, and Larry P. Walker, *Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism*, 2013

motivation:

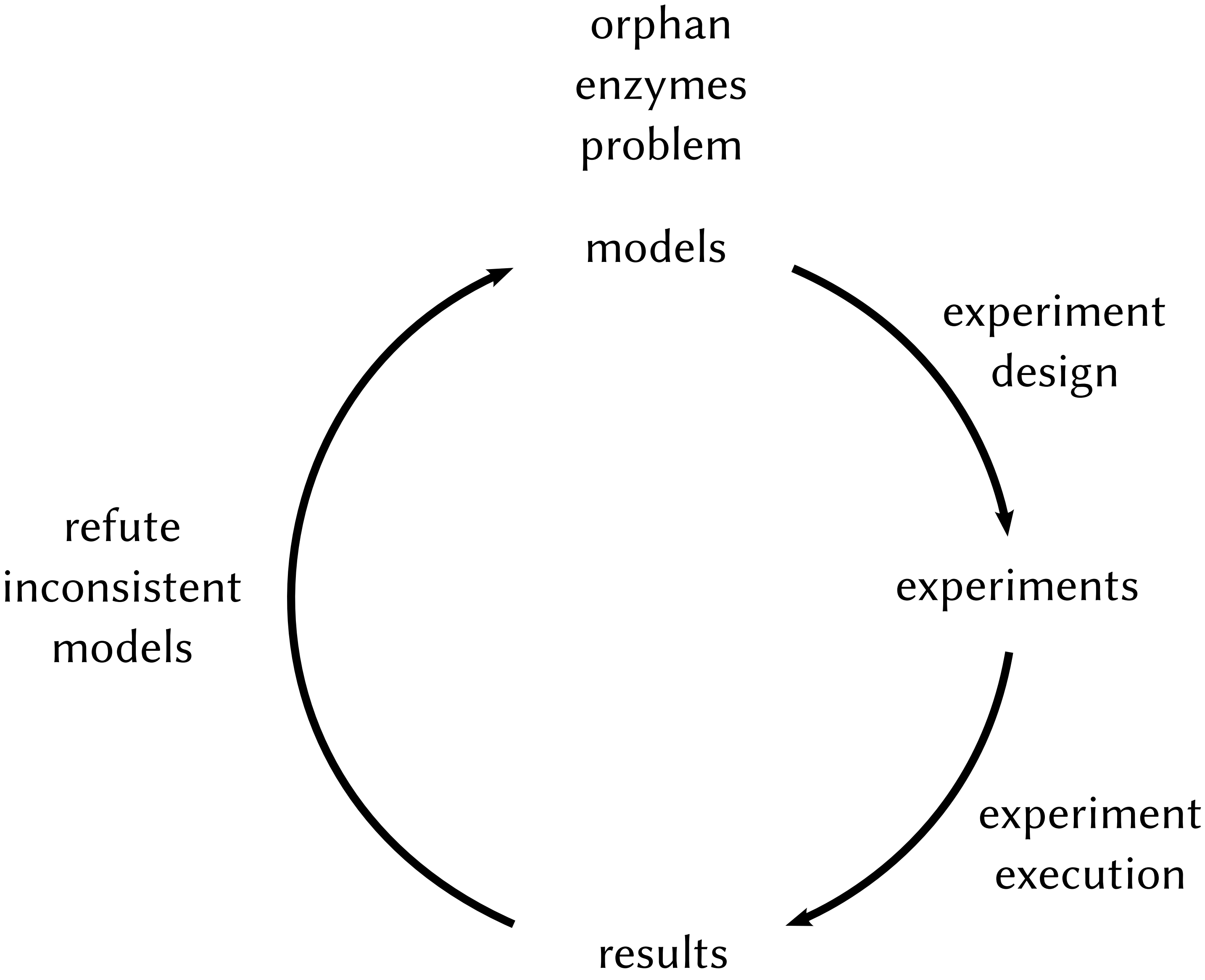
biological systems are large and complex
(yeast 7.11: ~3.7k species; ~4k reactions)

bottleneck for manual model development

could be overcome through automation

Hnin W. Aung, Susan A. Henry, and Larry P. Walker, *Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism*, 2013

previous work: Robot Scientist



orphan
enzymes
problem

models

experiment
design

experiments

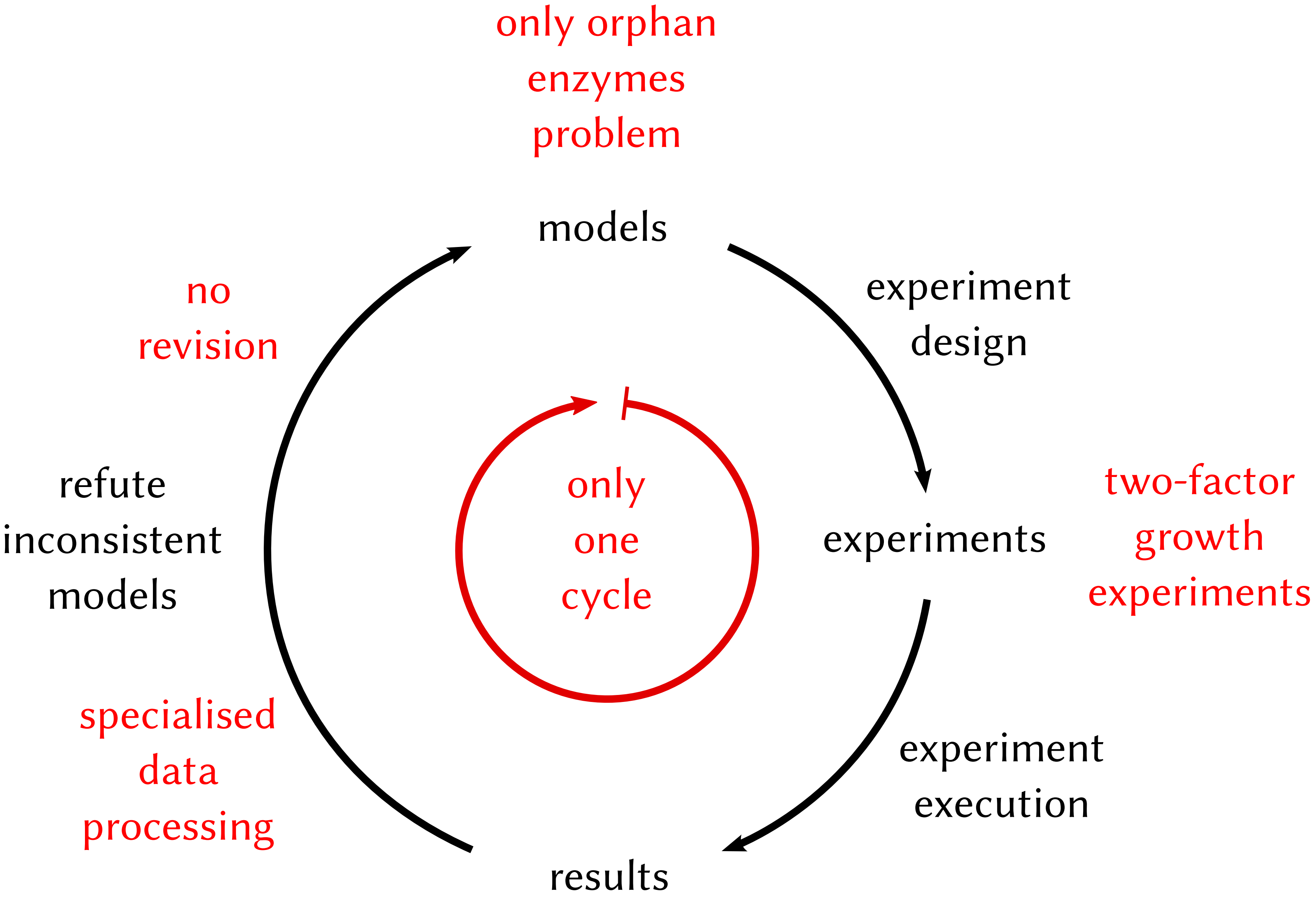
experiment
execution

results

refute
inconsistent
models

Ross D. King et al. *The Automation of Science*, 2009

previous work: Robot Scientist


only orphan enzymes problem

models

experiment design

two-factor growth experiments

experiments

experiment execution

results

specialised data processing

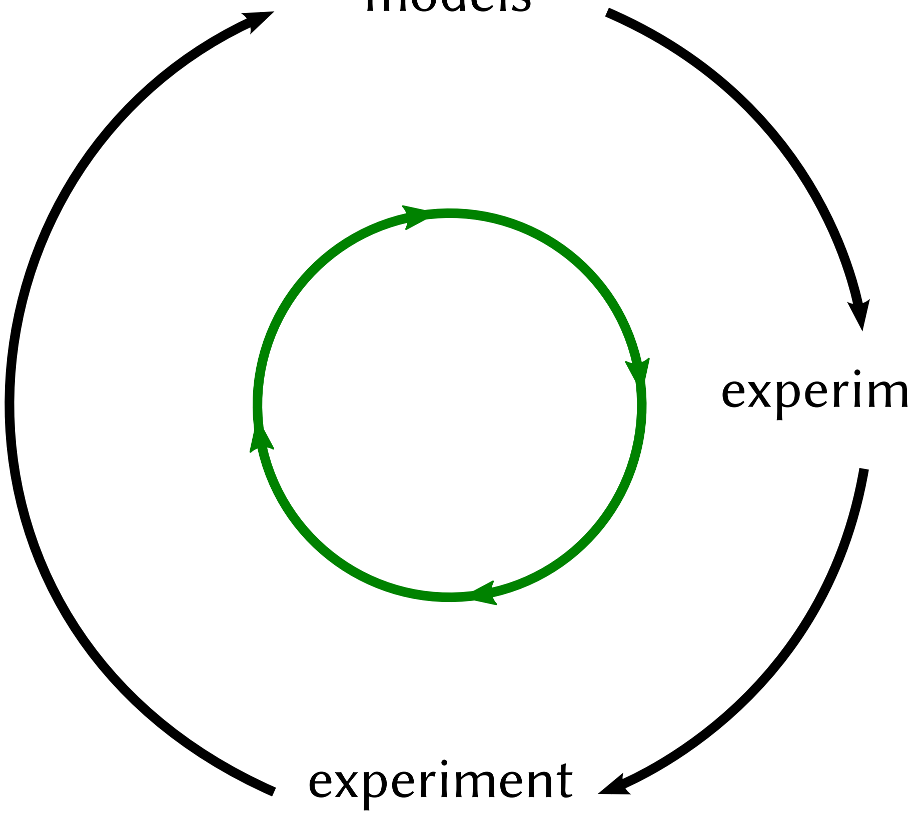
refute inconsistent models

no revision

only one cycle

Ross D. King et al. *The Automation of Science*, 2009

Huginn



learning structure of
metabolic networks

models

two-factor
growth exp.

metabolite/
protein detection

protein
localisation

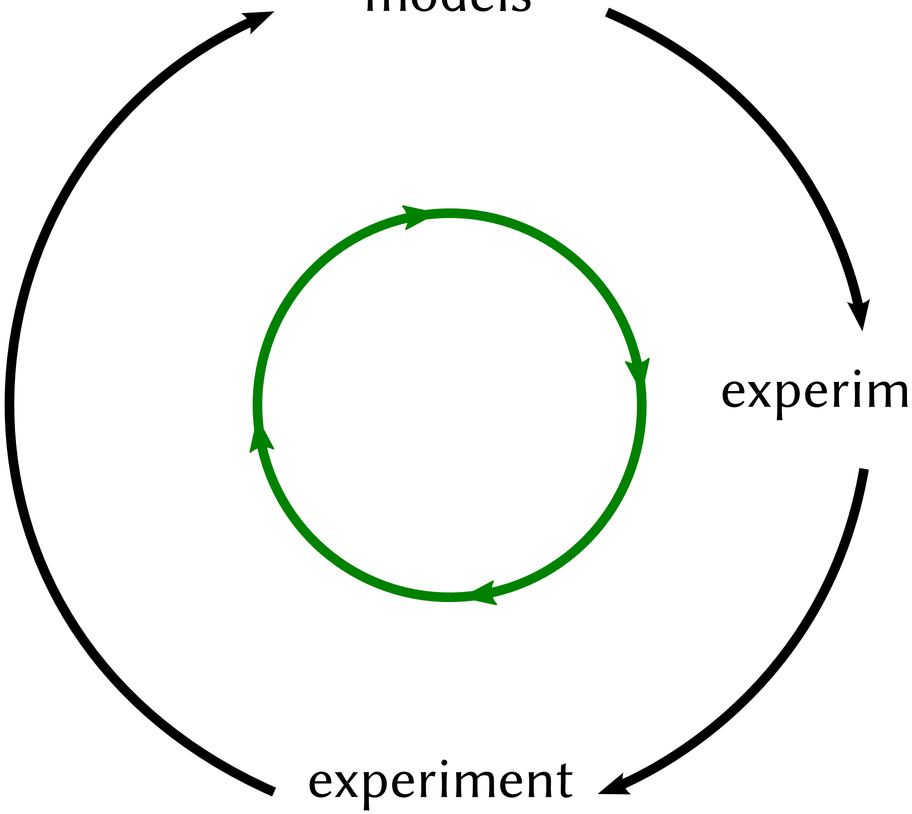
experiments

model
revision

consistency
test

experiment
execution
(simulation)

Huginn



learning structure of
metabolic networks

models

two-factor
growth exp.

model
revision

consistency
test

experiments

metabolite/
protein detection

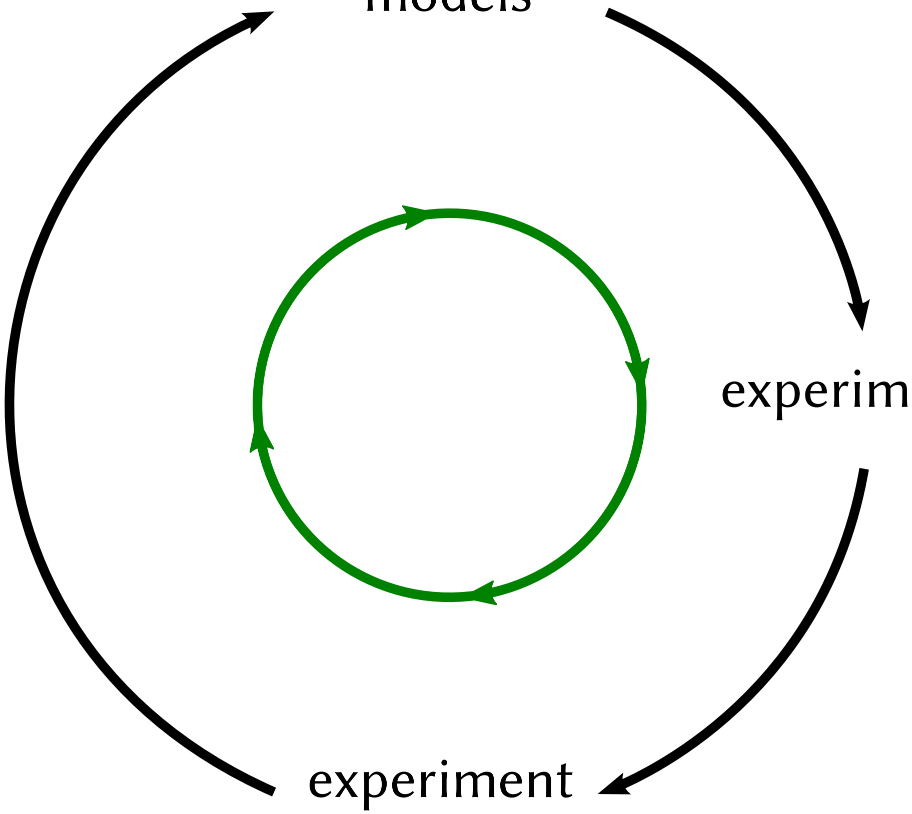
protein
localisation

experiment
execution
(simulation)

ASP/ALP techniques
automate crucial
operations

Huginn

learning structure of
metabolic networks



models

two-factor
growth exp.

metabolite/
protein detection

protein
localisation

experiments

experiment
execution
(simulation)

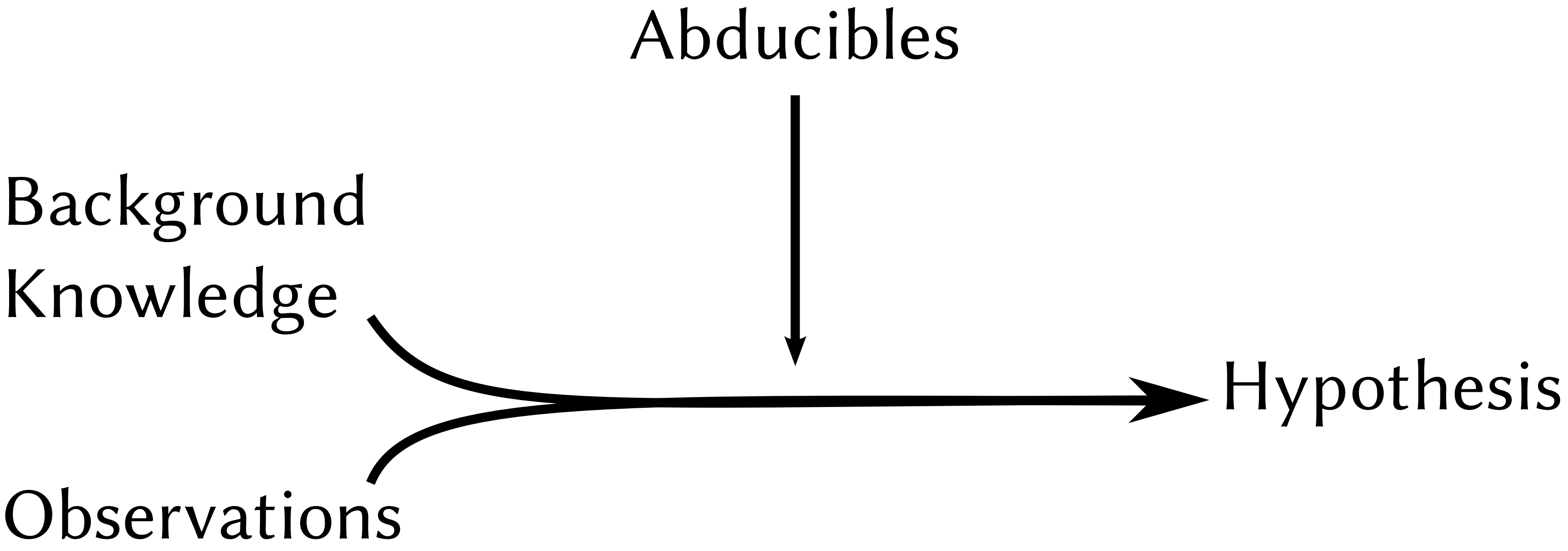
model
revision

consistency
test

ASP/ALP techniques
automate crucial
operations

inspired by XXth century
PoS and works by
Darden et al.

Abduction



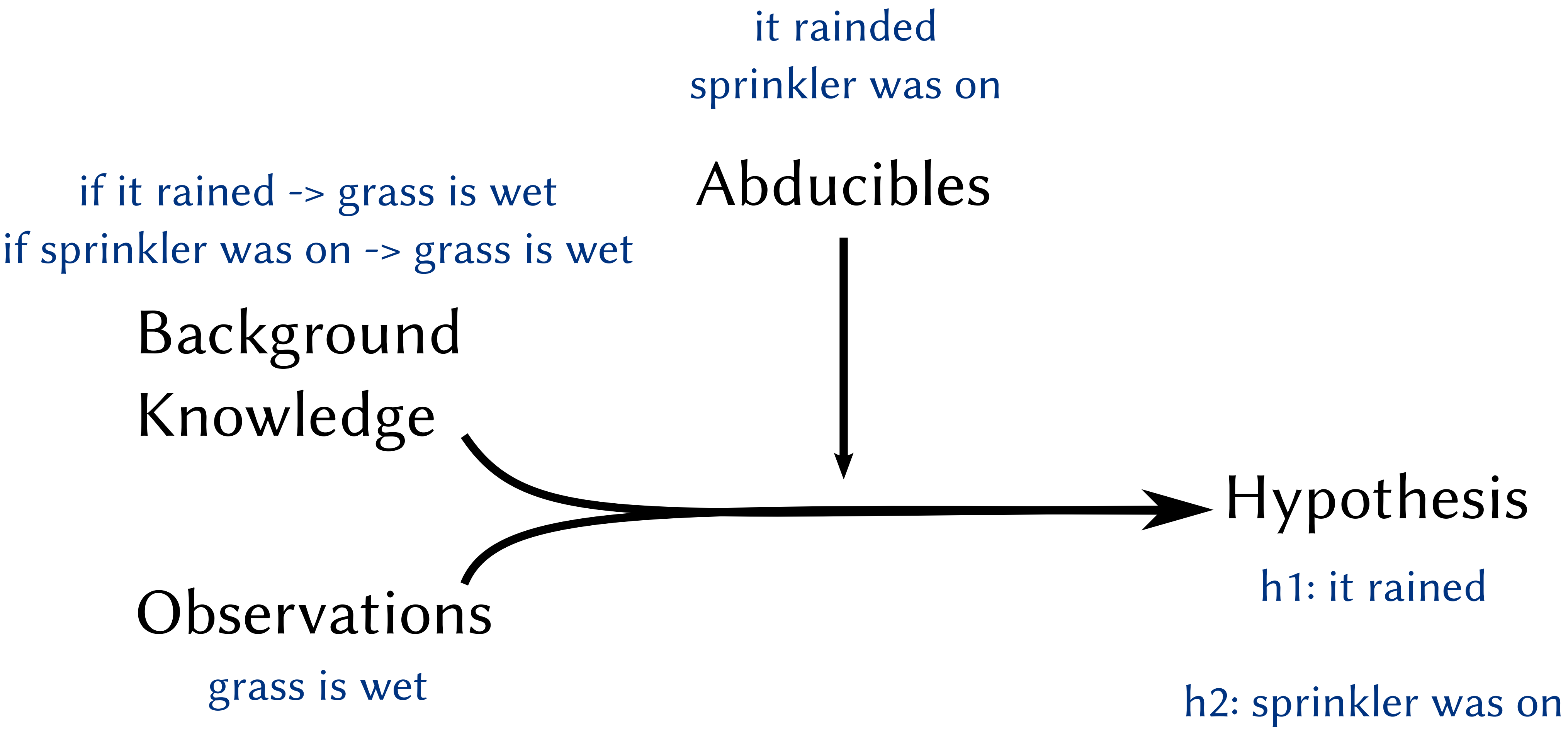
Abducibles

Background Knowledge

Observations

Hypothesis

Abduction



it rainded
sprinkler was on

Abducibles

if it rained -> grass is wet
if sprinkler was on -> grass is wet

Background
Knowledge

Observations

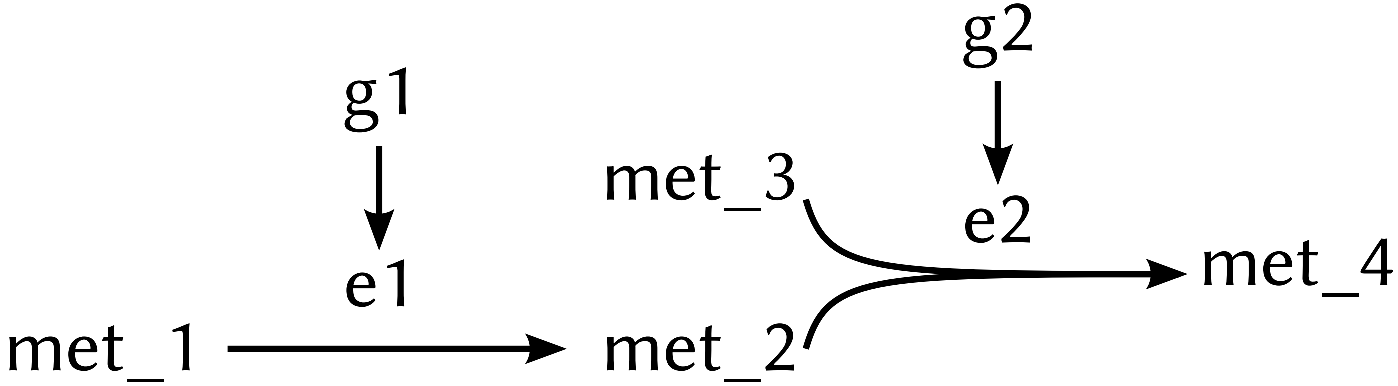
grass is wet

Hypothesis

h1: it rained

h2: sprinkler was on

Reasoning: consistency check



g2

g1

met_3

e2

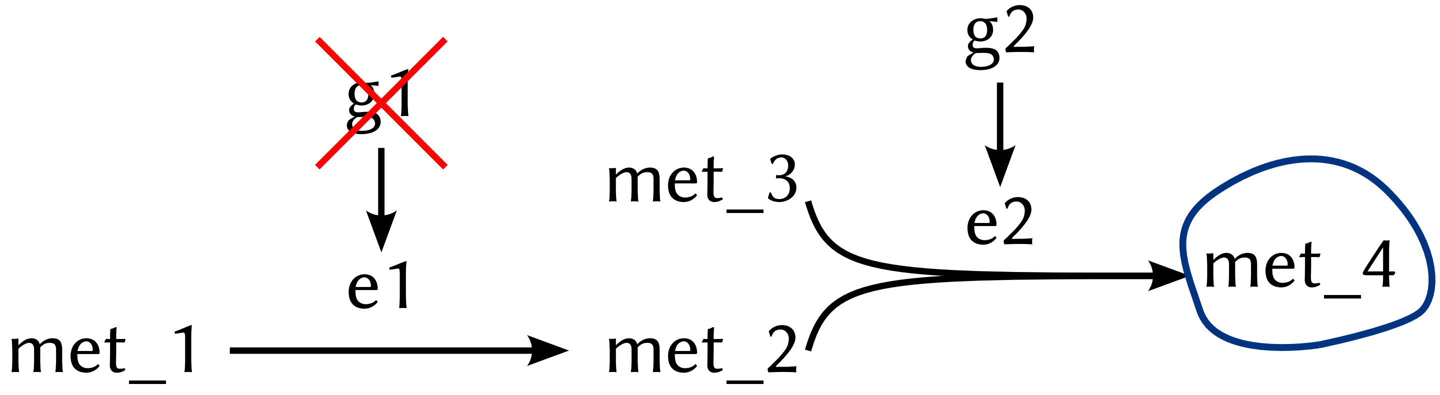
met_4

e1

met_1

met_2

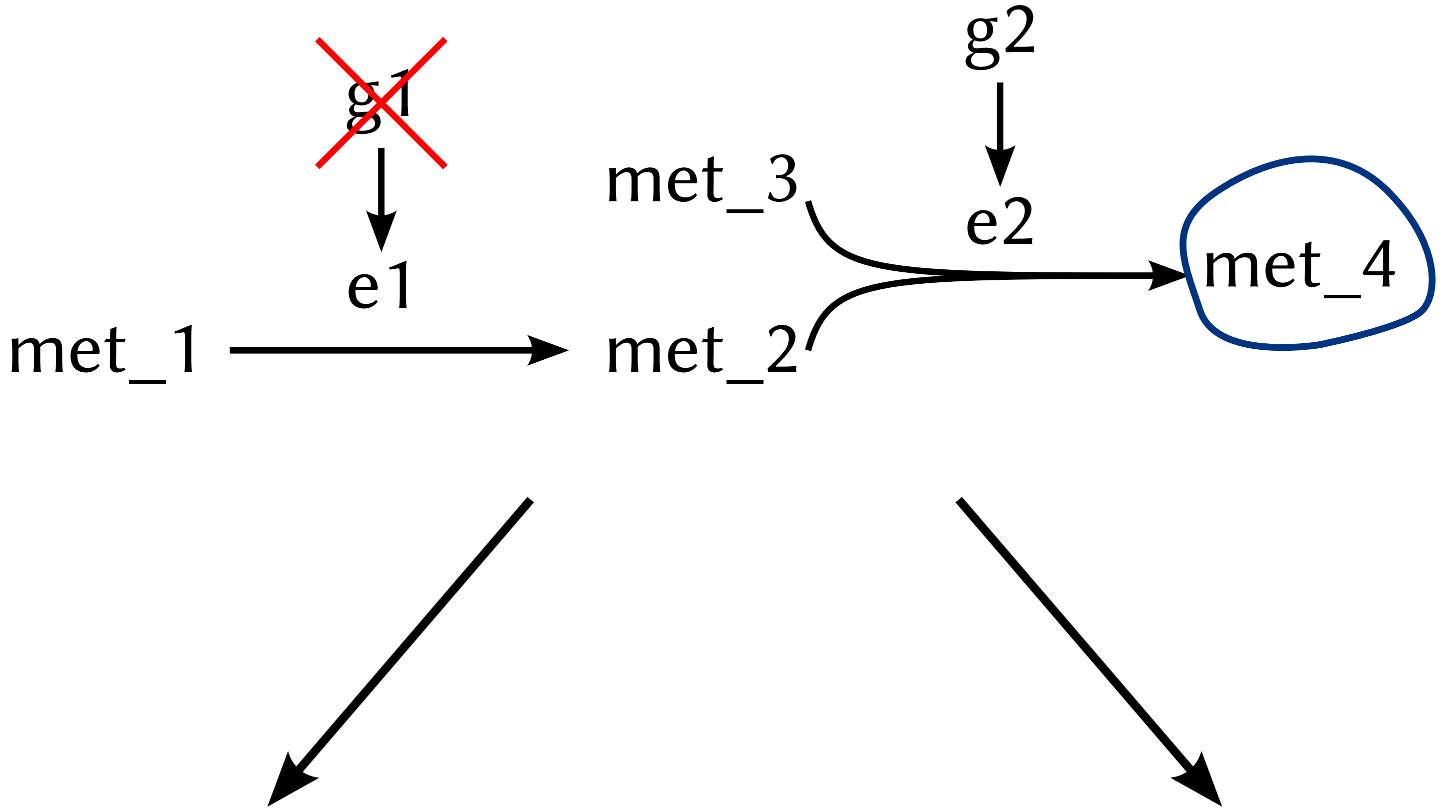
Reasoning: consistency check



experiment description:
- g1 deletion
- detection of met_4

experiment outcome:
- met_4 was detected

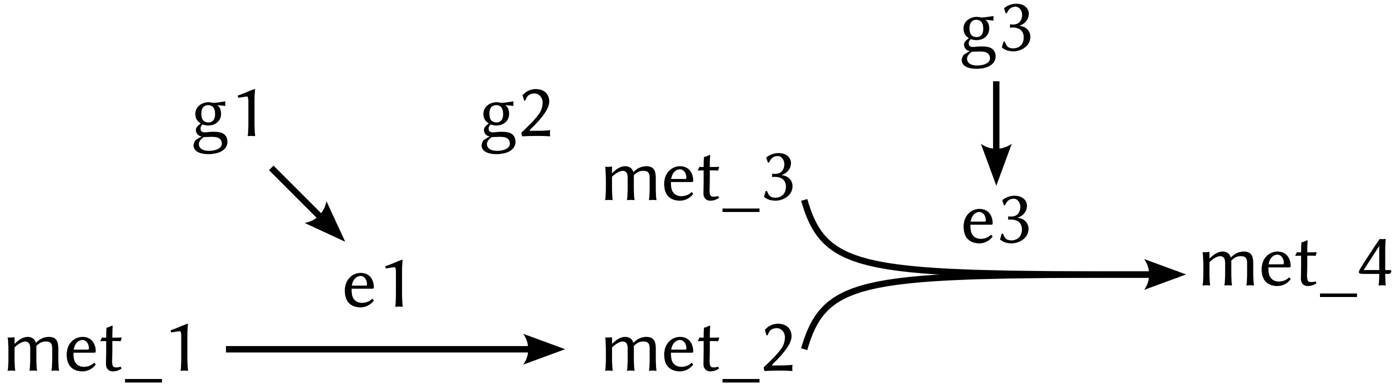
Reasoning: revision



add isoenzyme coded by a different gene (with appropriate expression)

add alternative pathway to met_4

Reasoning: experiment design



g3

g1

g2

met_3

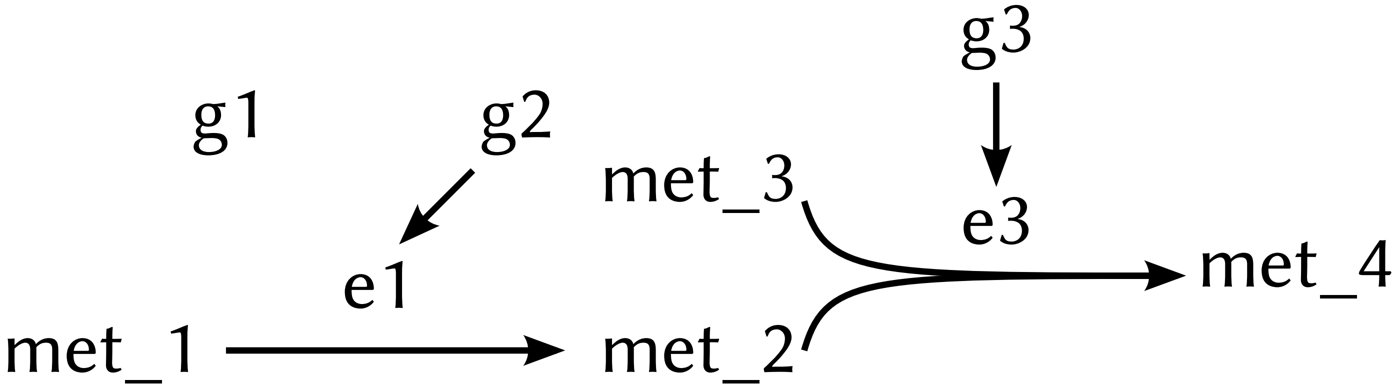
e3

met_4

e1

met_1

met_2



g3

g1

g2

met_3

e3

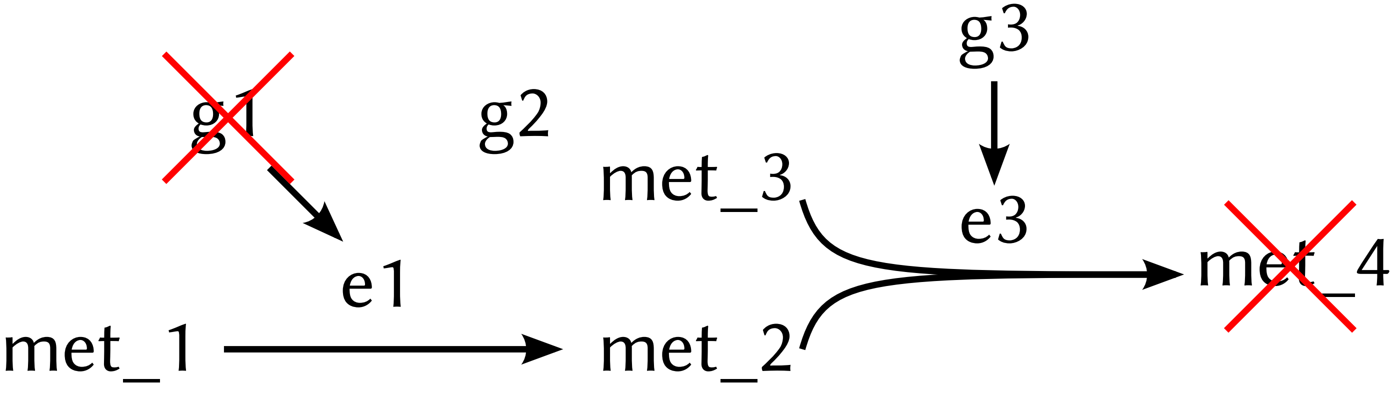
met_4

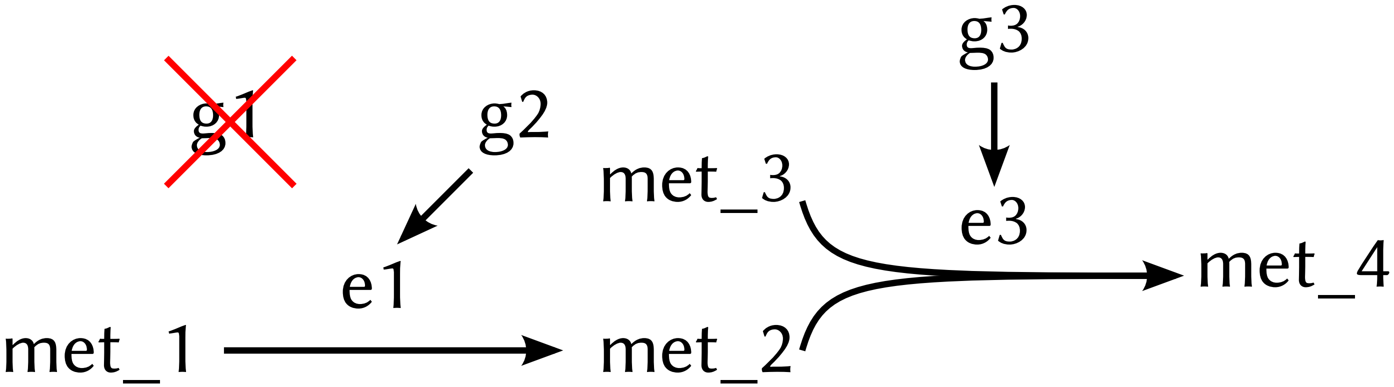
e1

met_1

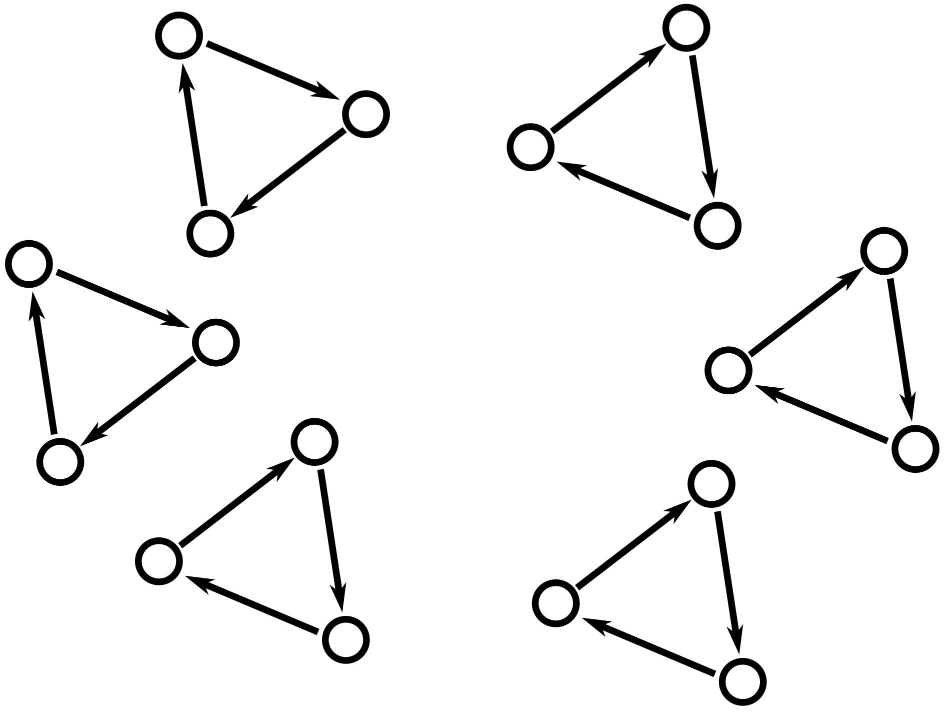
met_2

Reasoning: experiment design



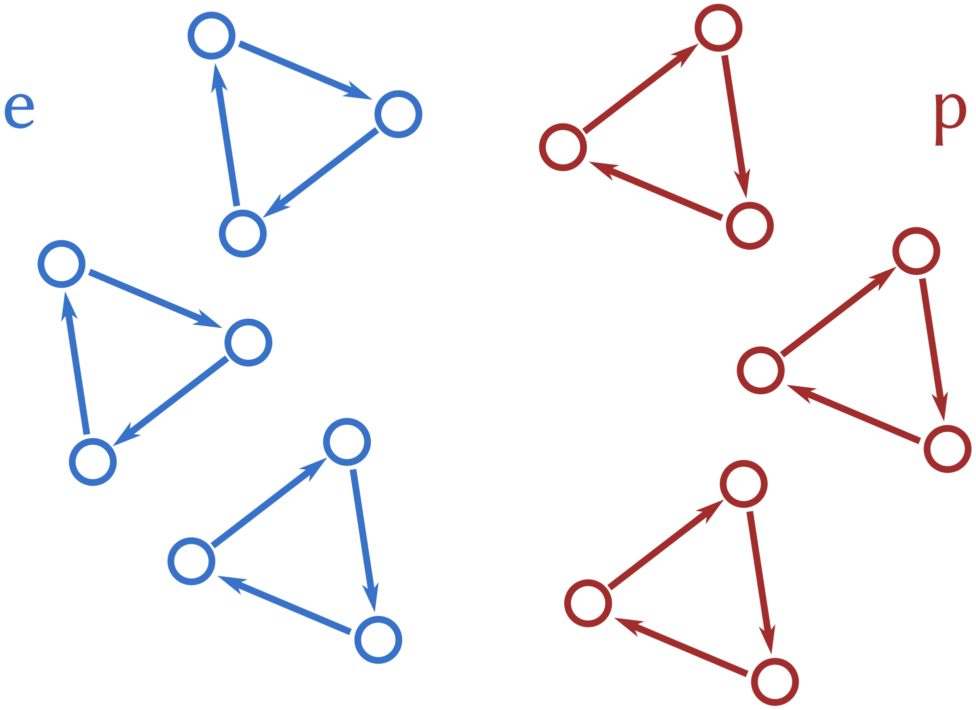


Experiment design strategy



R. D. King et al. *Functional genomic hypothesis generation and experimentation by a robot scientist*, Nature 2004

Experiment design strategy

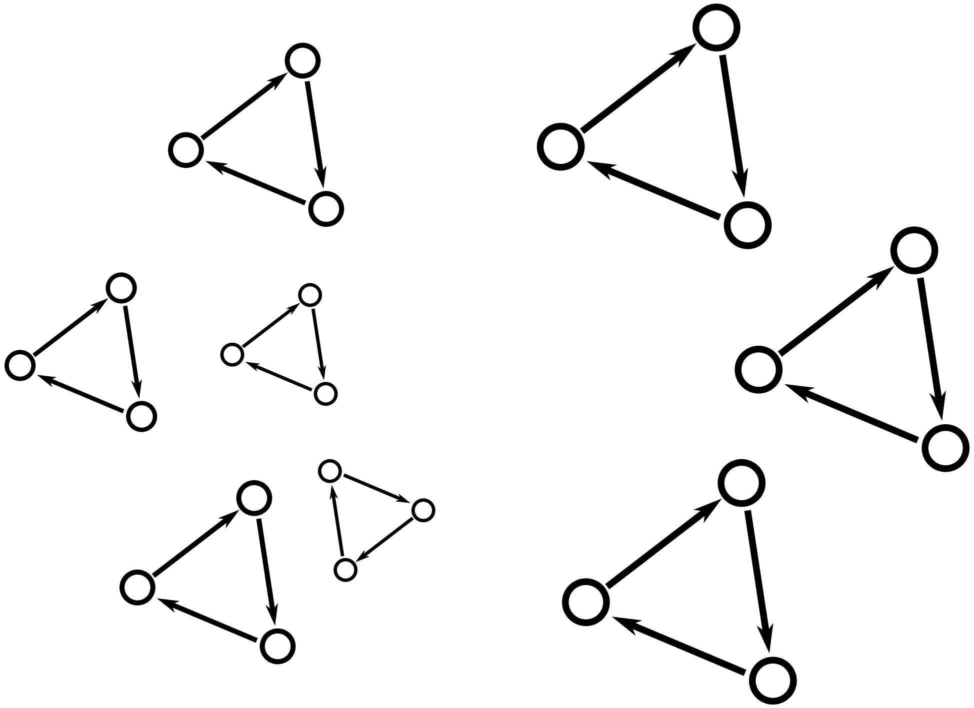


predict true

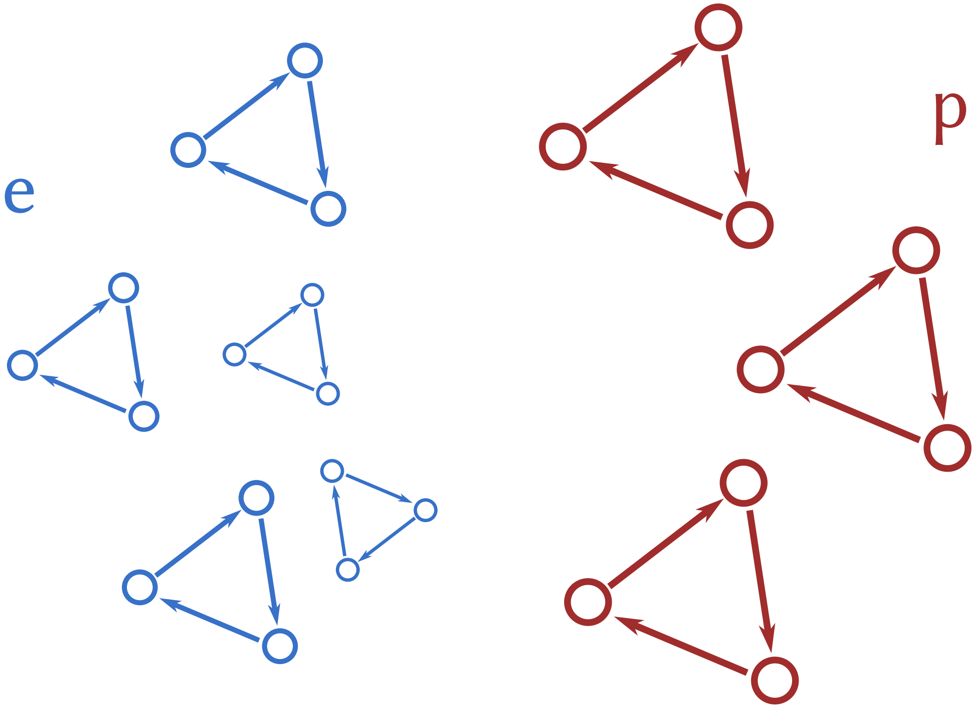
predict false

R. D. King et al. *Functional genomic hypothesis generation and experimentation by a robot scientist*, Nature 2004

Experiment design strategy



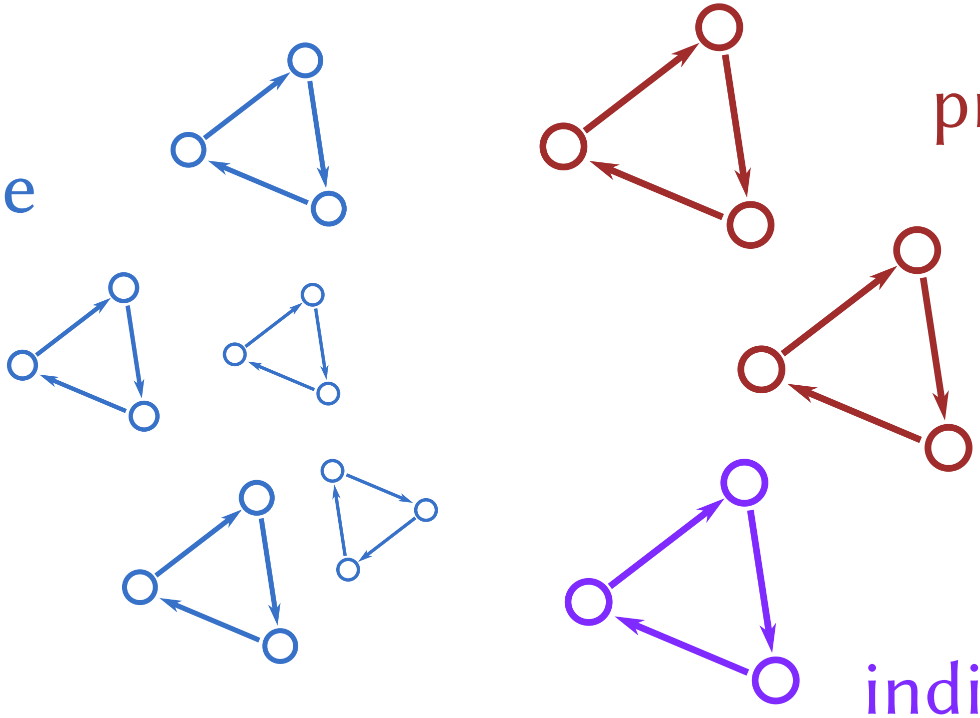
Experiment design strategy



predict true

predict false

Experiment design strategy



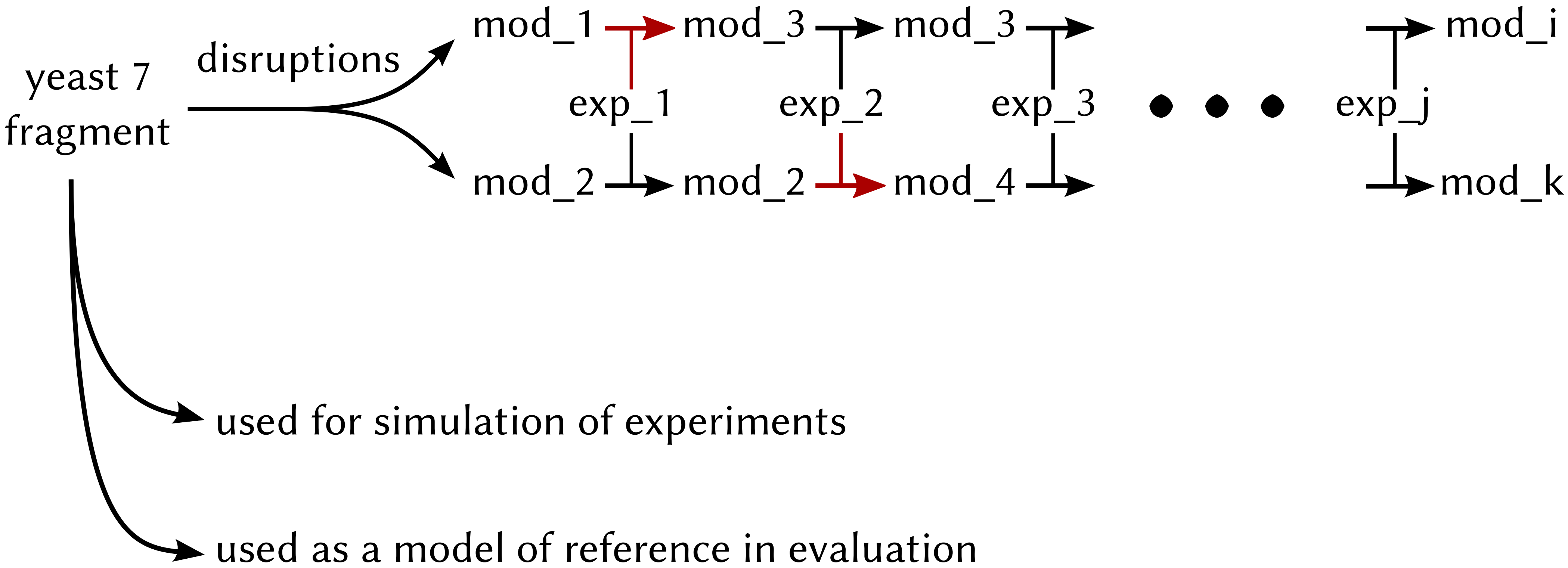
predict true

predict false

indifferent

Evaluation

disrupt a model and try to relearn it (structure)



yeast 7 fragment

disruptions

mod_1 mod_3 mod_3 mod_i

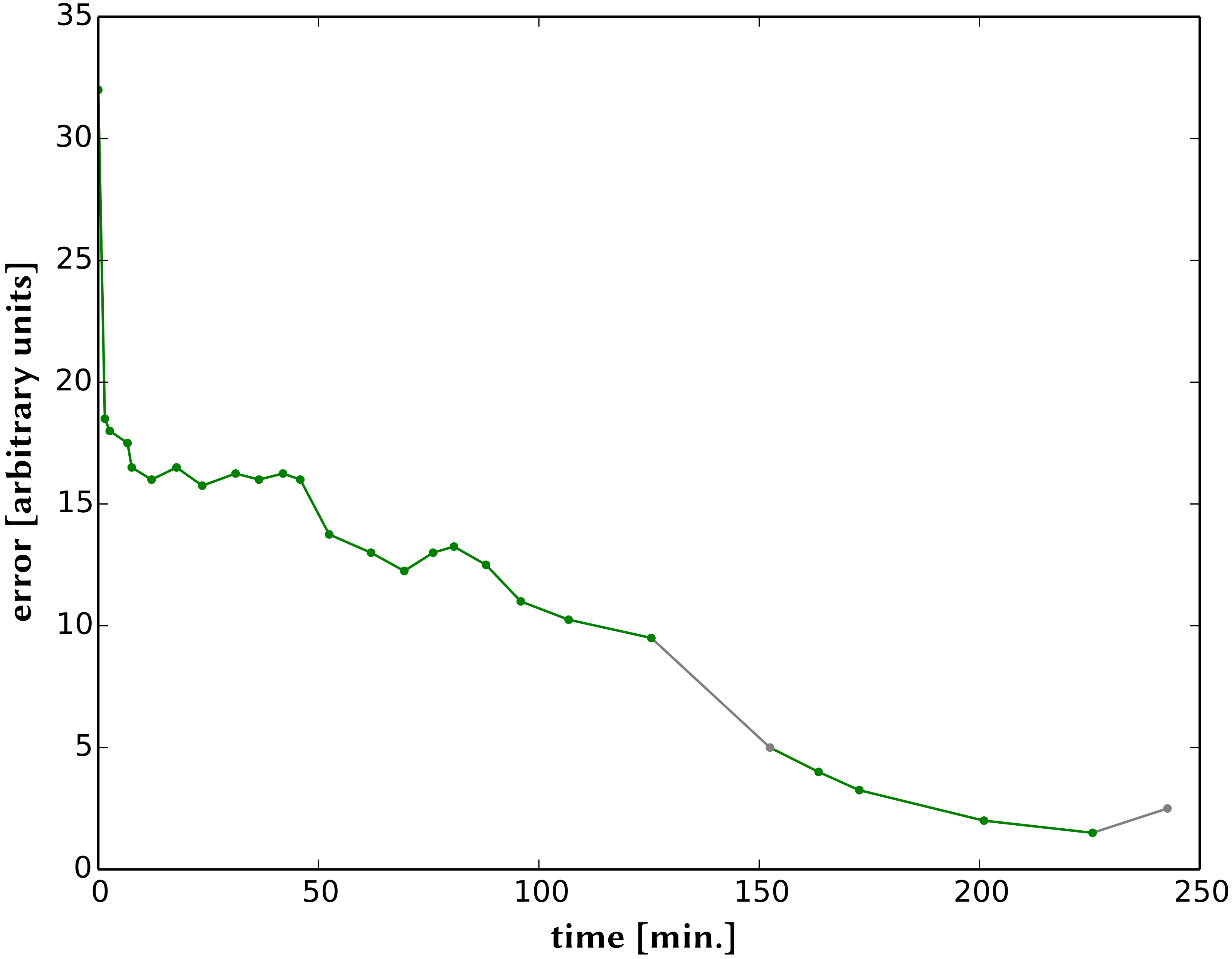
exp_1 exp_2 exp_3 • • • exp_j

mod_2 mod_2 mod_4 mod_k

used for simulation of experiments

used as a model of reference in evaluation

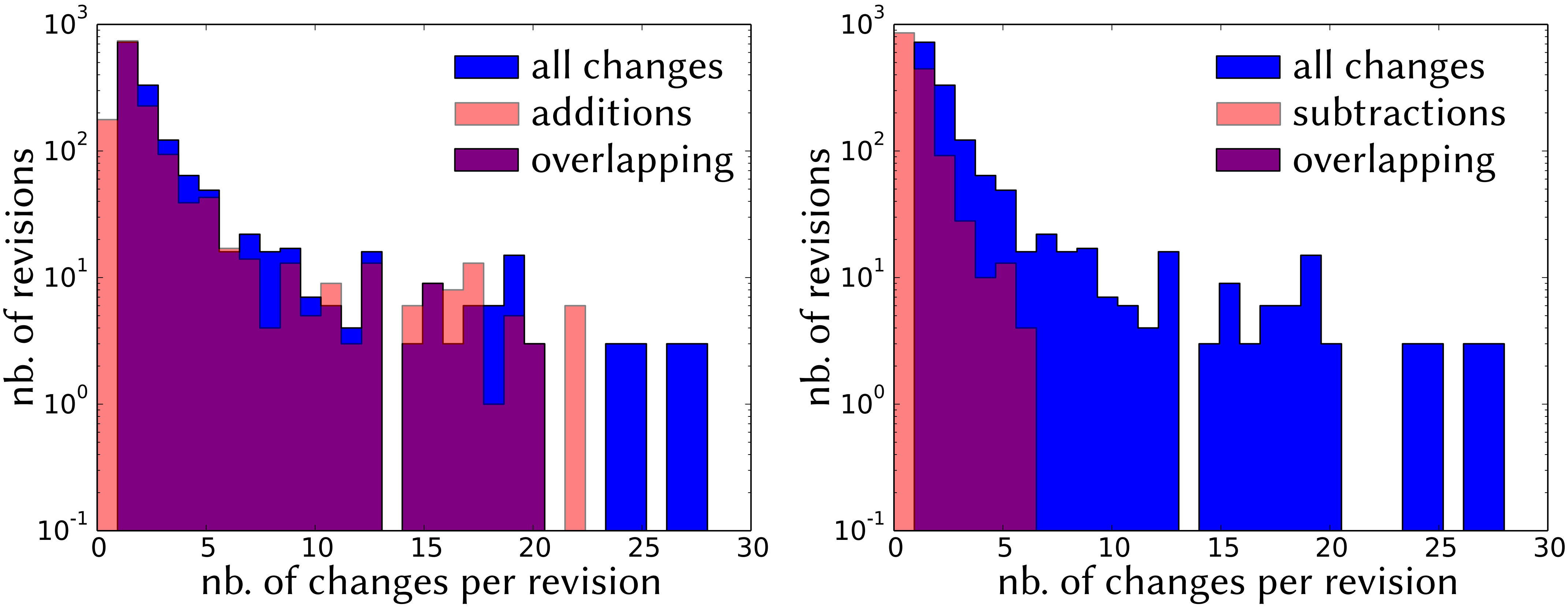
Results: models improved successfully



avg. error reduction: 76%

25-86 reactions

Results: Huginn can handle complex revisions



the largest involved as many as 28 changes

50 % of the revisions involved more than one change (addition/subtraction)

29% combined the addition and subtraction of reactions

Results: experiment design method is effective

two-factor growth experiments, detection and localisation experiments (metabolites/proteins/complexes)

detection experiments: 19% involved multiple interventions (gene deletions or medium manipulations)

most complicated experiment: 4 deletions + medium manipulation

Conclusions

Huginn is able to learn structure of metabolic models in a simulated environment

It has greater revision and experiment design abilities than previous systems

Designing extended crucial experiments using ALP demonstrated for the first time

Limitations

Not all types of supported experiments were used (would need supplementary experimentation solution)

Only higher level representation (no structure of compounds)

Can't support pulse-chase experiments

Further work

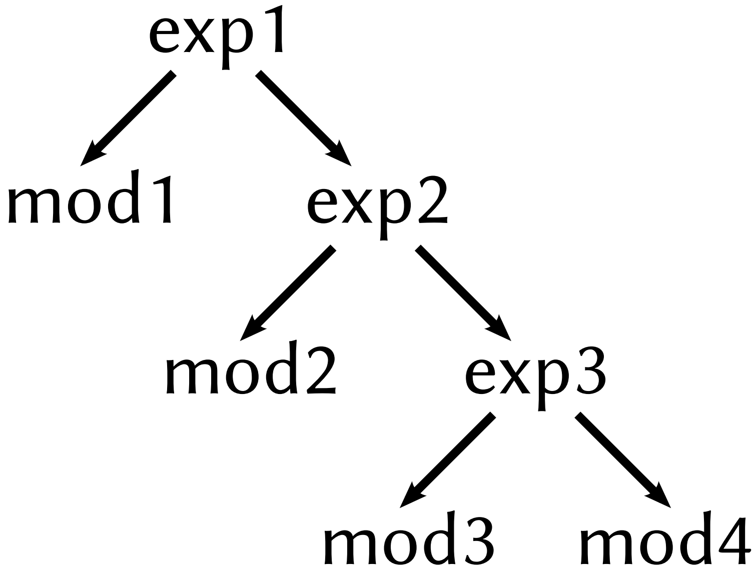
Improve efficiency & scale-up to bigger models

Adaptation to other biological networks:
- additional KR formalisms (dynamic aspects)

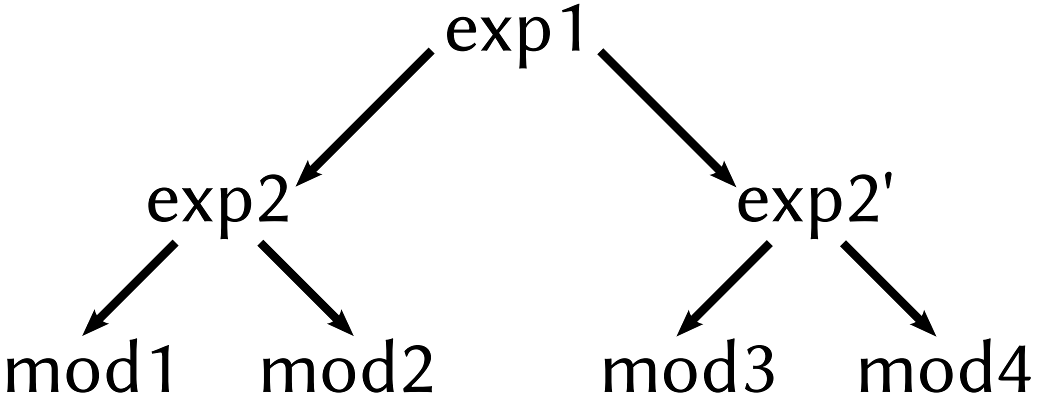
Extensions of discovery model:
- based on works of Machamer, Darden and Craver

More comprehensive experiments:
- design of control experiments, repetitions
- handling experimental noise (in exp. design, etc.)
- handling technical constraints (limited quantities, etc.)

Experiment design strategy: balancing decision tree

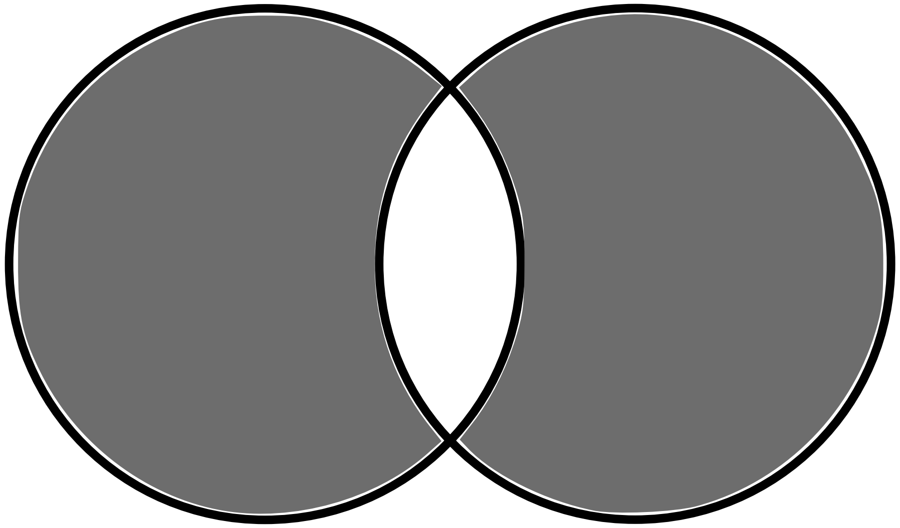


O(n)



$O(\log_2 n)$

Evaluation



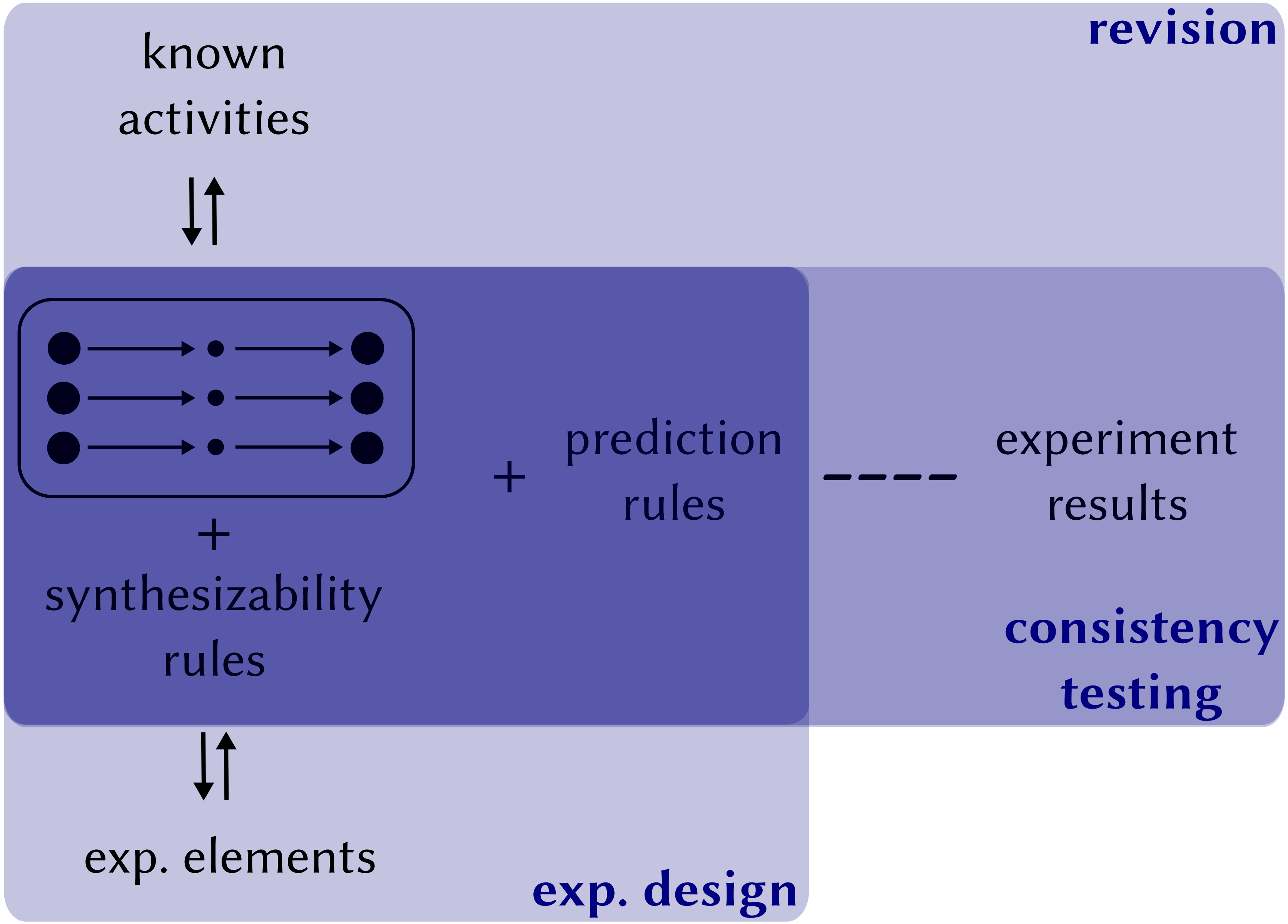
working model

model of reference

symmetric difference between sets of reactions

structural similarity (not functional)

Reasoning



revision

known activities

\+ synthesizability rules

\+ prediction rules

experiment results

consistency testing

exp. elements

exp. design

Types of experiments available

Growth experiment*

Detection experiment* (metabolite, protein, complex)

Localisation experiment* (metabolite, protein, complex)

Two-factor growth experiment (compares rates of growth of deletants and WT on standard and enriched medium to determine the role of the deleted gene)

Reconstruction in vitro (both enzymatic and non-enzymatic reactions; transport pseudo-reactions)

*may include interventions: gene deletions or change in growth medium composition

Disruptions

1. New elements added:
   - enzymes (20%) [catalysing random reactions]
   - complexes (20%) [catalysing random reactions]
   - additional expression pseudo-reactions and transport
   - metabolic reactions (20%) [randomised substrates and products]

2. Initial models created by sampling all reactions and pseudo-reactions (from the model of reference and 1.)