Resilience and Intelligence

Katsumi Inoue
National Institute of Informatics, Japan

Systems Resilience – Bridging the Gap Between Social and Mathematical

***Shonan Meeting**, February 23, 2015*

Contents

Contents

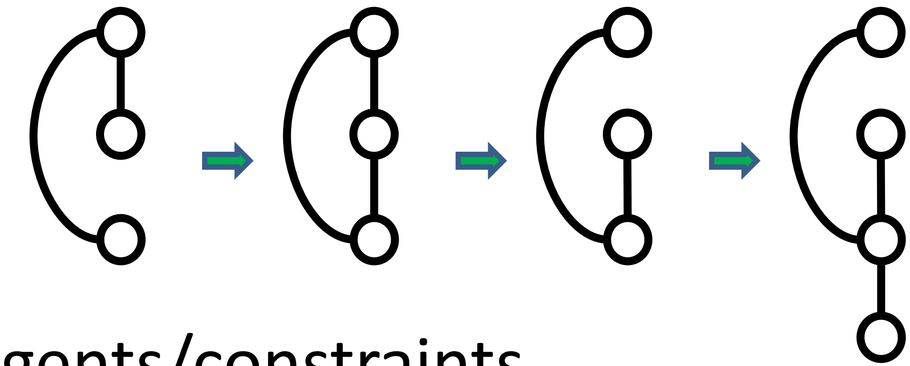
1. Intelligence into Resilience
2. Resilience into Intelligence
3. Intelligence as Resilience
4. Resilience as Intelligence

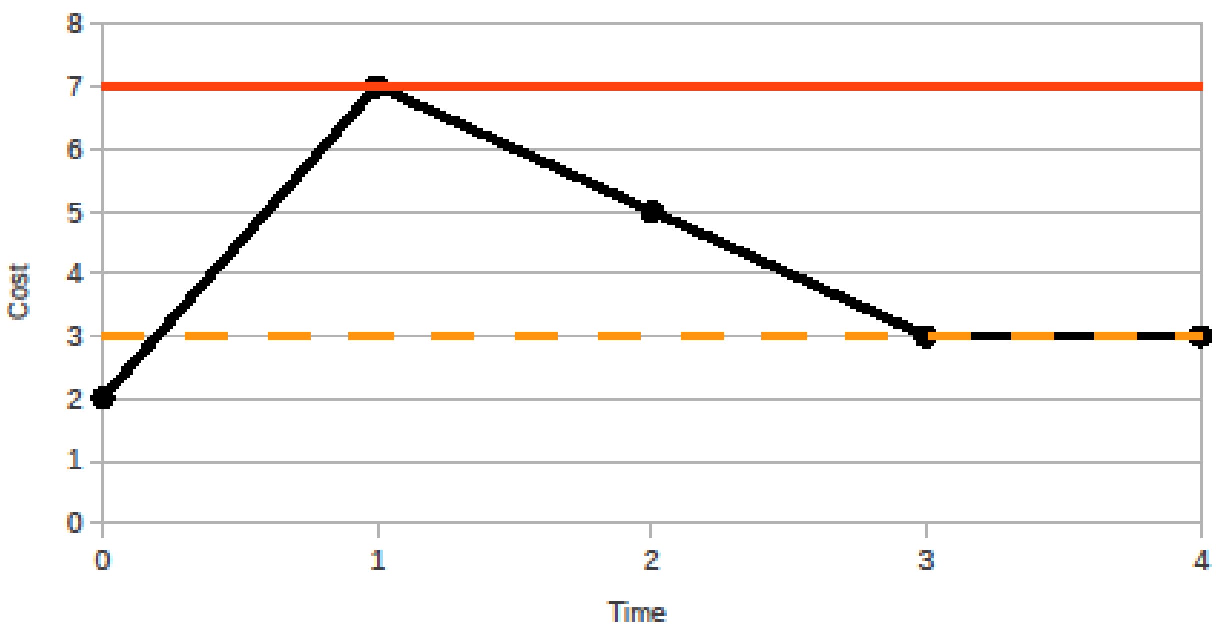
1. Intelligence into Resilience

1. **Suitable abstraction of problems**:
   - Mathematical models – *discrete/hybrid (complex) systems*
   - Symbolic representation – *dynamic (constraint) networks*
2. **Logic for systems resilience**:
   - Semantics: dynamics in terms of *possible worlds*
   - Inference: verification/prediction – *model checking, explainabiliy*
   - Update: *reasoning about change*
3. **Computation of resilience**:
   - Decision/optimization problems – *(multi-objective) CSP/COP*
   - Exact/approximation algorithms – *robust solutions*
4. **Design of resilient systems**:
   - Design of systems with desirable dynamics – *machine leaning*
   - Robustness/sensitivity analysis – *multi-agent simulation*

SR-Model (Schwind *et al.*, AAMAS 2013)

1. Dynamical systems
2. Multi-agent systems
3. Constraint-based systems
4. Flexible, can add/delete agents/constraints





Resistance

+

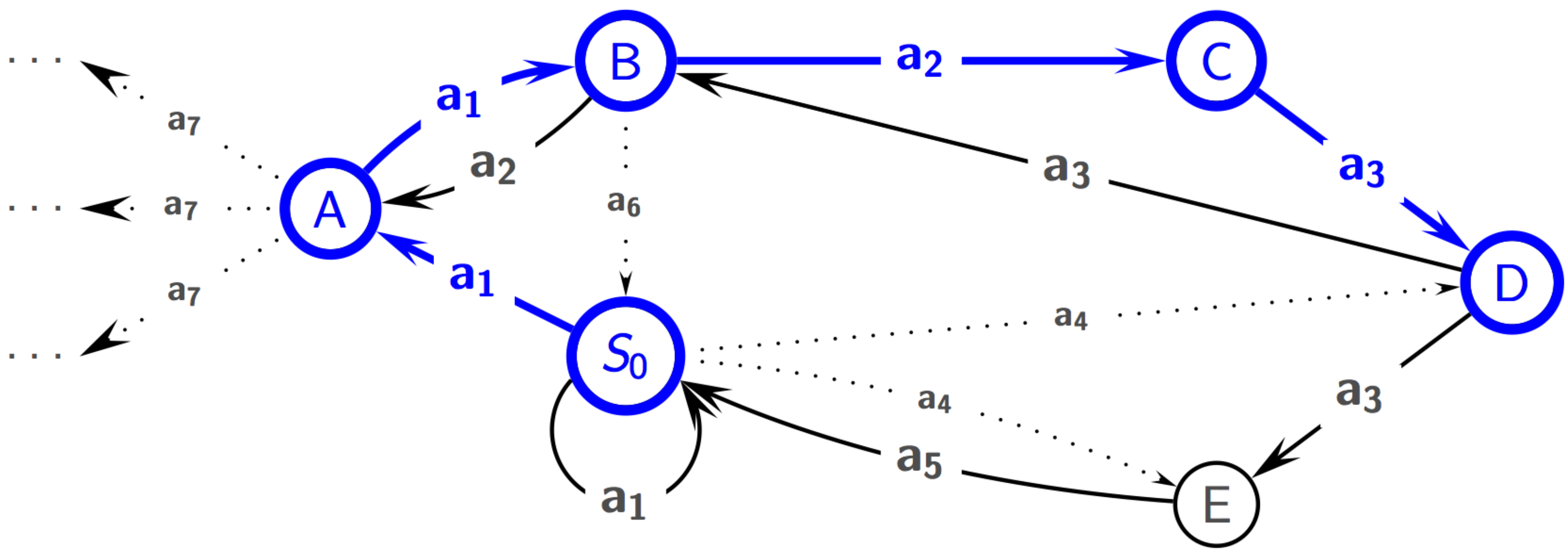
Recoverability

=

Resilience

Shape of a Dynamic System

- At each time step, a decision is made.
- Depending on the environment (uncontrolled event), the specifications of the system may change without any restriction.

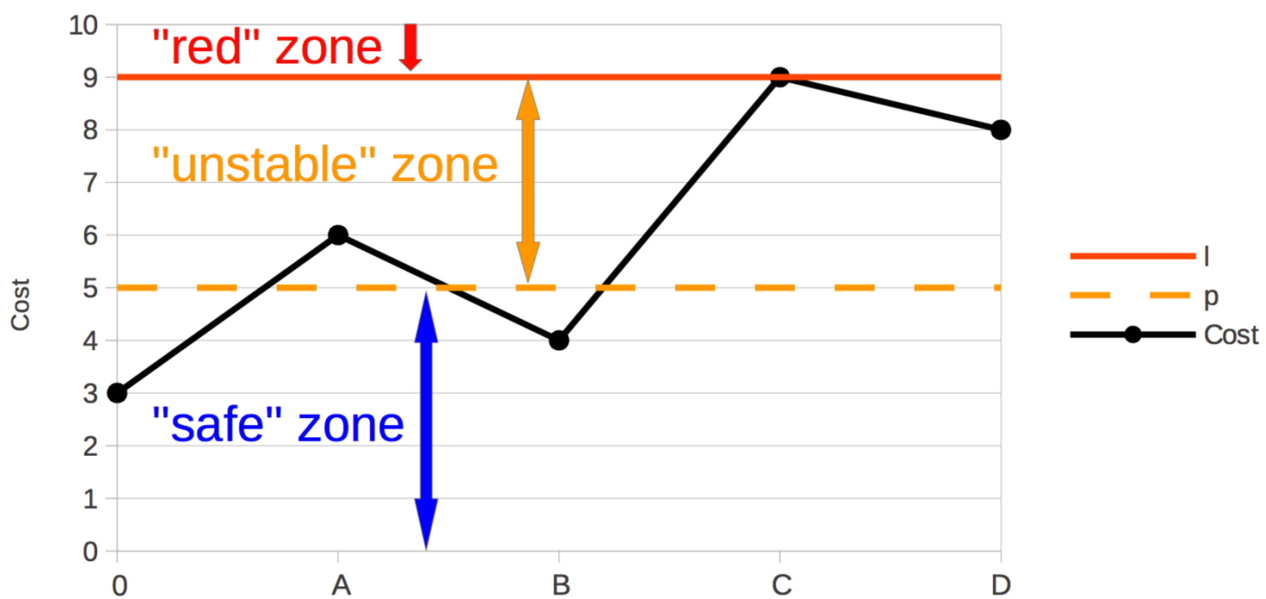


Resistance + Recovery

- At each time step, the state of the system is associated with a cost
- Resistance + Recovery:

➢ The ability to maintain some underlying costs under a certain “threshold”, such that the system satisfies certain hard constraints and does not suffer from irreversible damages.

➢ The ability to recover to a baseline of acceptable quality as quickly and inexpensively as possible.



Functionality + Stabilizability

- *Functionality*: the ability to provide a guaranteed average degree of quality for a period of time.
- *Stabilizability*: the ability to avoid undergoing changes that are associated with high transitional costs.
- A dynamic system is *resilient* if one can find a "strategy" (i.e., the "right decisions") and a state trajectory within this strategy that is resistant, recoverable, functional, and stabilizable.

Logical Theory of Unpredictability

- To know if an event is (un)predictable or not
- To identify if there is an unpredictable state

Approach

- A logical account of (**un**)**predictability** based on **abduction**.
- Provide computational methods of configurations of cellular automata in logic programming and Answer Set Programming.

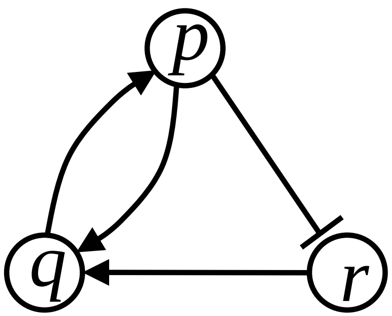
Results

- Investigate Hempel's symmetry: *An event* **E** *is predictable under* <**B**, **H** > iff **E** *is explainable under* <**B**, **H** >.
- *A configuration* **E** *is a* ***Garden of Eden*** *of a cellular automaton* iff ***E*** *is unpredictable under* <**B**, **H** >.

☞ C. Sakama and_K. Inoue: "Abduction, Unpredictability and Garden of Eden", *Logic Journal of the IGPL*, 21(6):980-998, 2013.

Reasoning about Boolean Networks

- Models of biological (gene regulatory and signaling) networks
- Models of complex systems like Cellular Automata and Game of Life
- Analysis of dynamic behavior involving positive and negative feedbacks



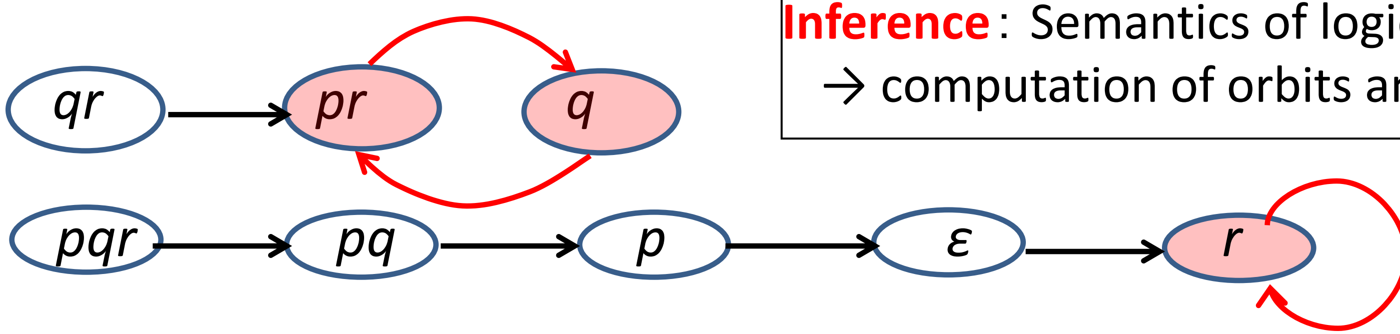
$p \leftarrow q.$

$q \leftarrow p \wedge r.$

$r \leftarrow \neg p.$

Attractor #1: (p,q,r) =
101 ⇒ 010 ⇒ 101 ⇒…

Attractor #2: (p,q,r) =
001 ⇒ 001 ⇒ …



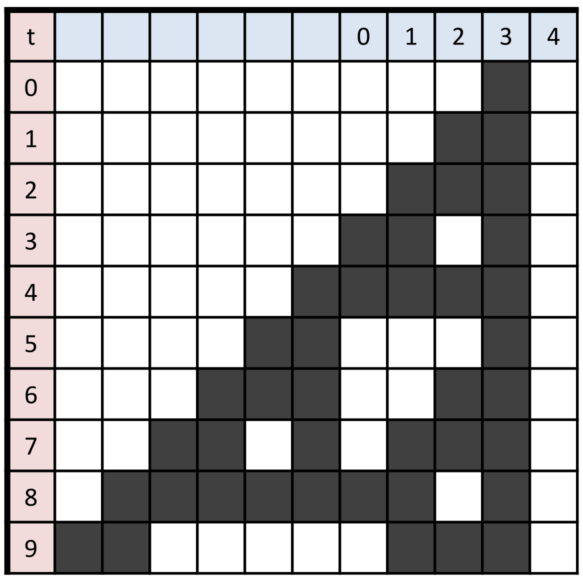
Inference: Semantics of logic programs
→ computation of orbits and attractors

☞ K. Inoue: "Logic Programming for Boolean Networks", *IJCAI 2011*.

☞ K. Inoue & C. Sakama: "Oscillating Behavior of Logic Programs", *Correct Reasoning* (Lifschitz Festschrift), *LNAI*, Vol.7625, pp.345-362, 2012.

Learning Dynamical and Complex Networks

- Dynamic systems involving positive and negative feedbacks
- Learning Boolean networks from state transition diagrams
- Learning Cellular Automata from traces of configuration change



- $c(x,t+1) \leftarrow c(x\text{-}1,t) \land c(x,t) \land \neg c(x+1,t).$
- $c(x,t+1) \leftarrow c(x\text{-}1,t) \land \neg c(x,t) \land c(x+1,t).$
- $c(x,t+1) \leftarrow \neg c(x\text{-}1,t) \land c(x,t) \land c(x+1,t).$
- $c(x,t+1) \leftarrow \neg c(x\text{-}1,t) \land c(x,t) \land \neg c(x+1,t).$
- $c(x,t+1) \leftarrow \neg c(x\text{-}1,t) \land \neg c(x,t) \land c(x+1,t).$

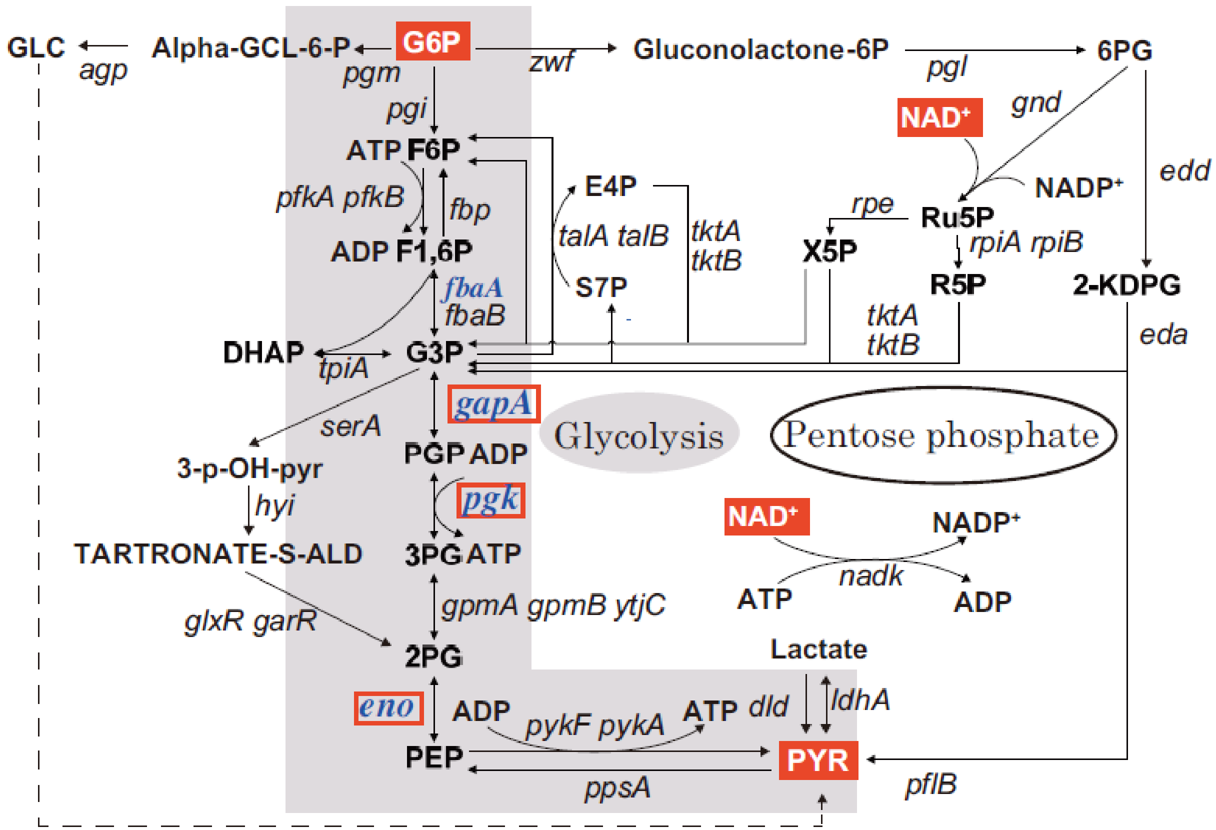
current pattern	111	110	101	100	011	010	001	000
new state for center cell	0	1	1	0	1	1	1	0

Wolfram's Rule 110 (Turing-complete)

☞ K. Inoue, T. Ribeiro & C. Sakama: "Learning from Interpretation Transition", *Machine Learning,* 94(1):51-79, 2014.

Prediction of Gene Knockout Effects of *E.coli* by SAT-Based Minimal Model Generation

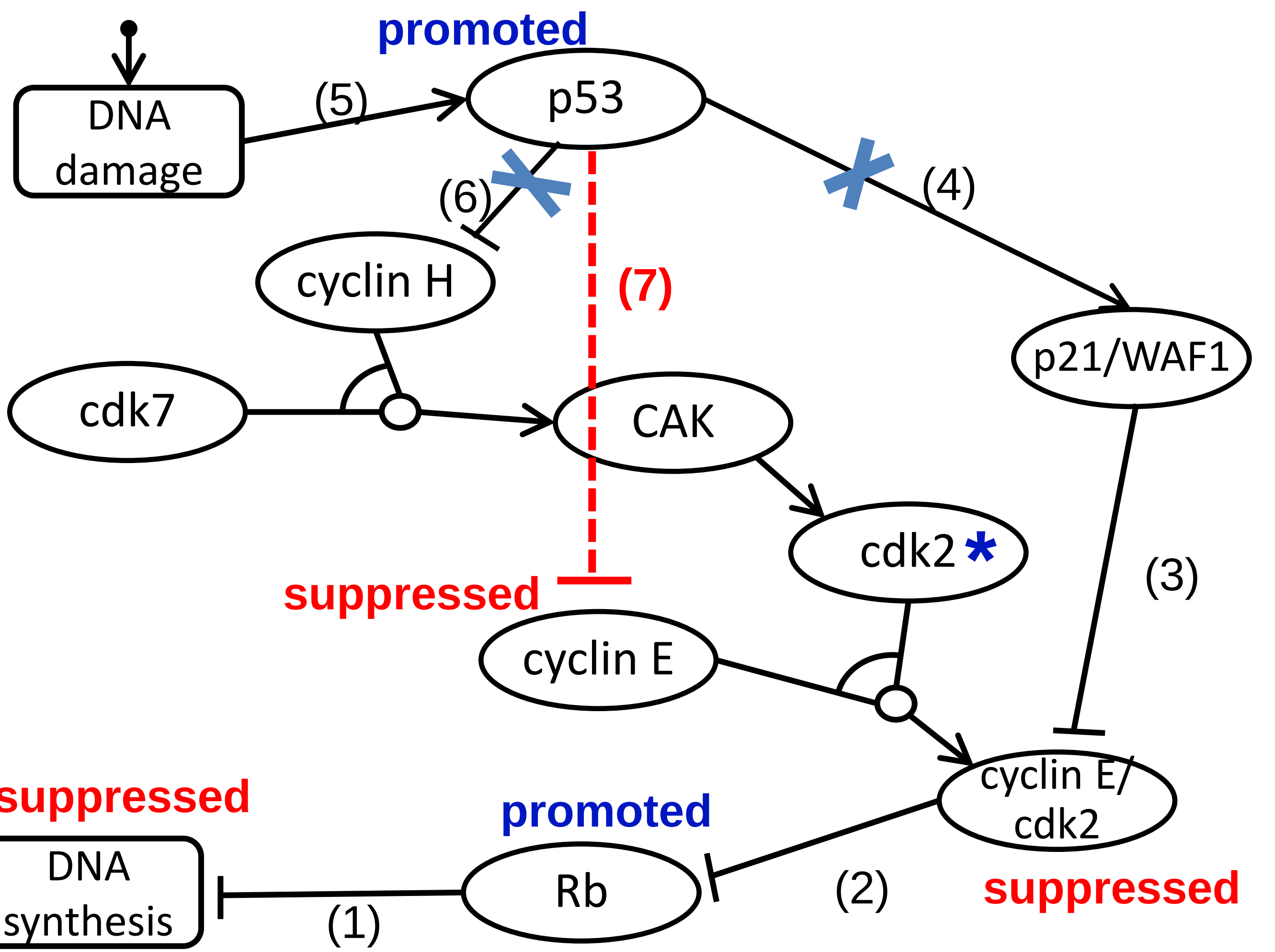
(Soh, Inoue, Baba, *et al.*: *Int'l J. Adv. Life Science*, 2012)



Source metabolites (M_S): β-D-glucose-6-phosphate, NAD^+	G6P NAD^+
Target metabolite (M_T): pyruvate	PYR
Four essential genes confirmed by KEIO collection	*fbaA* *gapA* *pgk* *eno*
Genes predicted by our method	(red box)

Biological Robustness: Pathway Completion by Meta-Level Abduction

(Inoue, Doncescu & Nabesima: *Machine Learning*, 2013)



Cell cycle with cyclin-dependent kinases (Schneider *et al.*, 2002)

2. Resilience into Intelligence

AI = search problems

weakly constrained: too many possible solutions

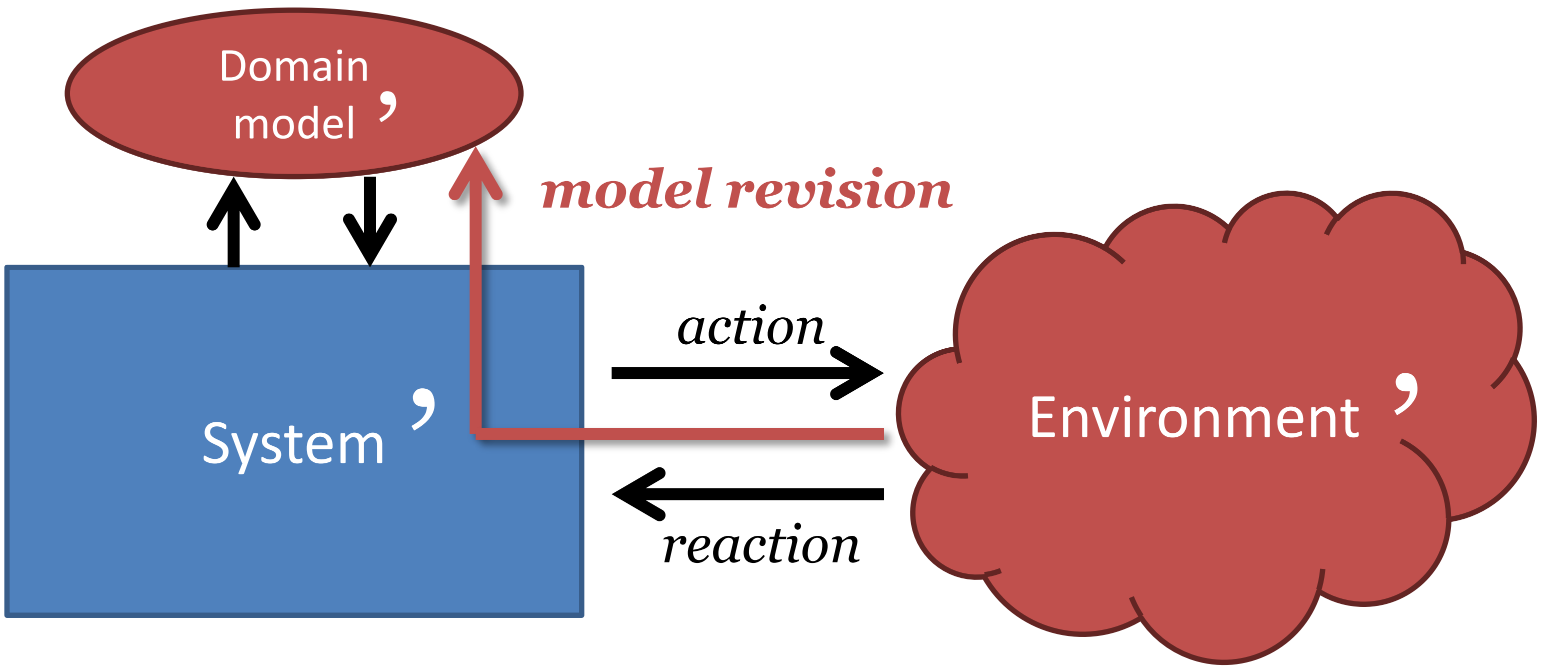
Resilient Solutions

Select the models that are robust/diverse/adaptable/etc.

Design agent systems that are enforced stabilizabiliation.

Revising Plans in Adaptive Systems

(Sykes *et al.*, *ICSE 2013*)



Behavioural model revision through **probabilistic rule learning**

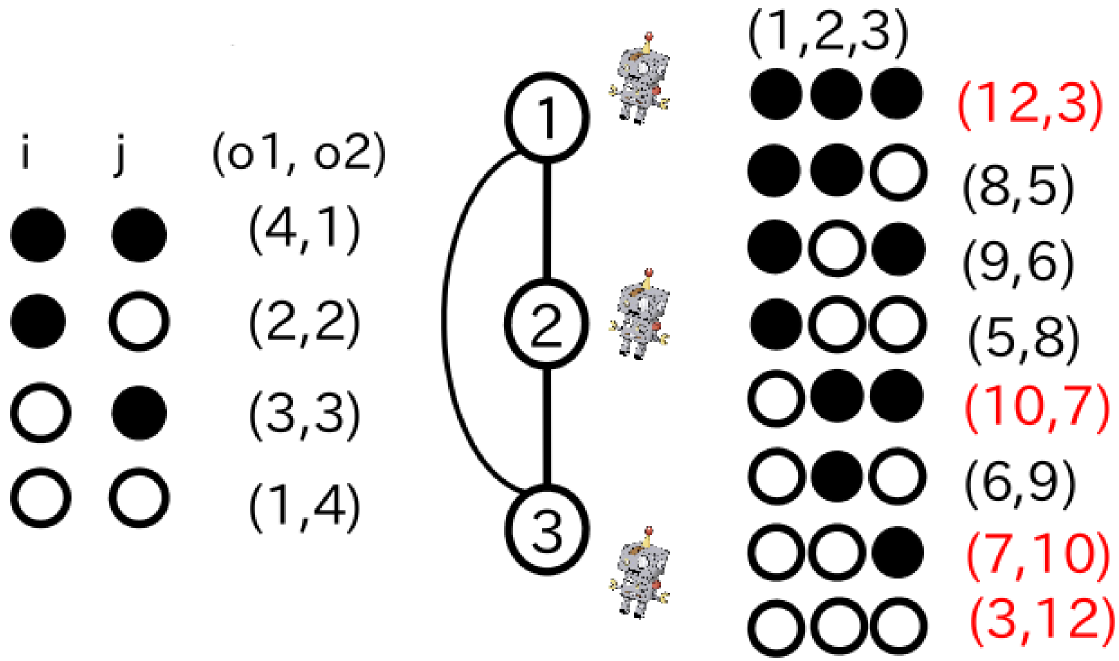
Diverse Solutions

- Structural diversity — diverse genotypes
- Functional diversity — diverse phenotypes
- These notions have been incorporated in many optimization research, in particular for multi-objective optimization using genetic algorithms.

- Well-balanced diversity (Schwind, *et al.*, 2015): representative solutions — useful for distribution of sensor networks, enhancing robustness

Multi-Objective Distributed Constraint Optimization Problem (MO-DCOP)

- Real-world problems involve multiple criteria that should be considered separately yet optimized simultaneously.



- Computation of resilient systems that have trade-off criteria
- Multiple criteria are considered in *Pareto solutions*

Resilient Solutions for Dynamic Multi-Objective Constraint Optimization

(Okimoto, Clement, Schwind & Inoue: *ICAART 2015*)

- Introduce the framework of a Dynamic Multi–Objective Constraint Optimization Problem (DMO– COP)
- Focus on resistance and functionality
- Provide an algorithm called Algorithm for Systems Resilience Applications (ASRA) to compute all resistant and functional solutions for DMO–COPs

Secured AI

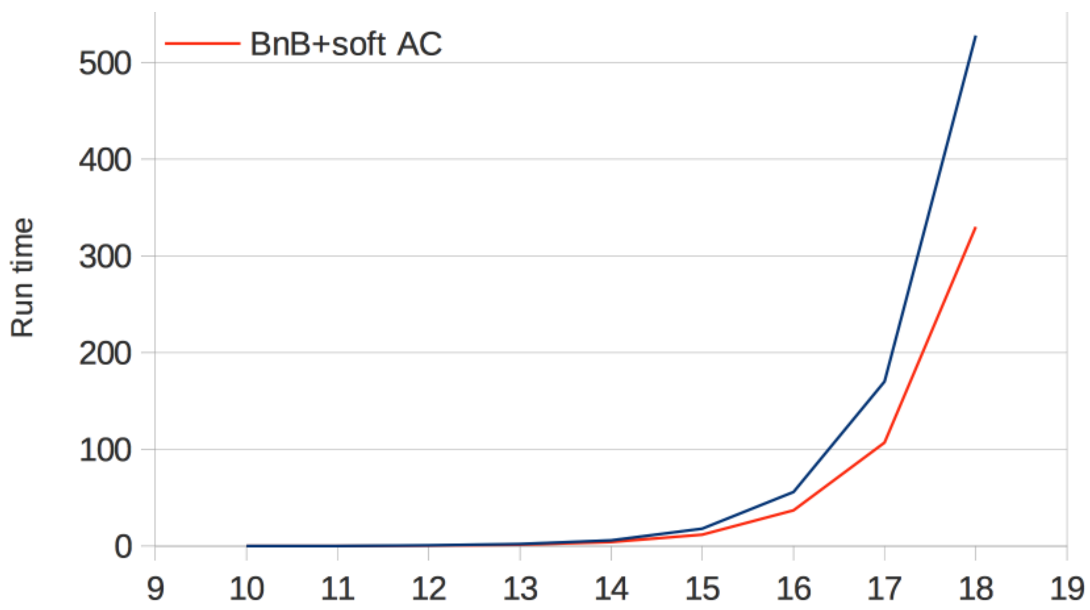
- Security research can help make AI more robust ☺
- AI systems are used in an increasing number of critical roles, including cyber-attack surface area 😐
- AI and machine learning techniques will themselves be used in cyber-attacks ☹
- At a lower level, robustness against exploitation is achieved by verifiability and freedom from bugs.
- At a higher level, AI techniques could be applied to the detection of intrusions, analyzing malware, and detecting potential exploits in other programs through code analysis.

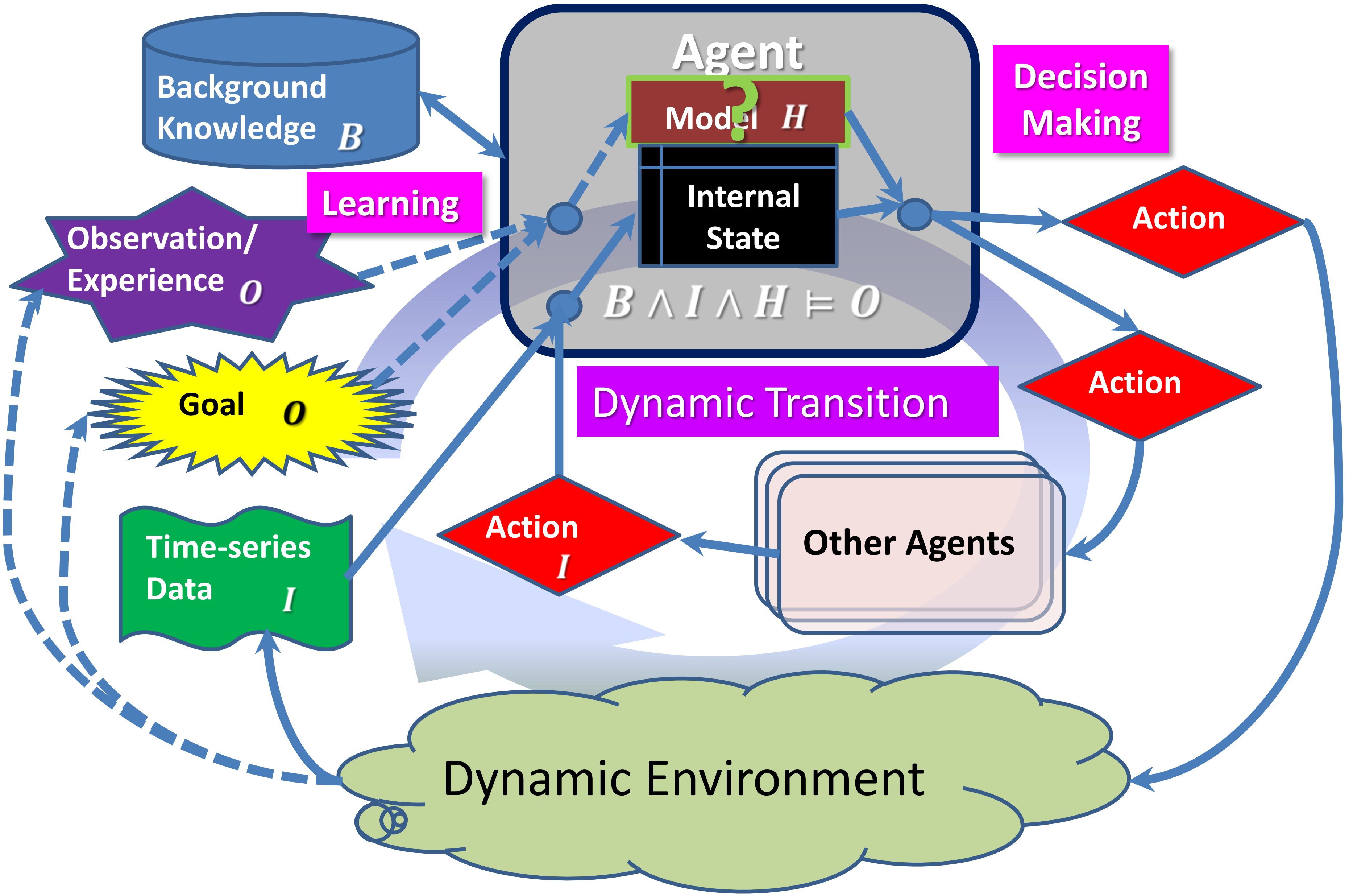
Cyber Security Trade-Off Problem

- *"Interception and communications data retention measures, even if the purpose is social security, are under the difficult trade-off between security, privacy and cost."*
- How to solve this trade-off and build the societal consensus?

Cyber Security Problem Based on Multi-Objective Distributed Constraint Optimization Techniques
(Okimoto, Ikegai, Ribeiro, *et al.*, *WSR 2013*)

- The algorithm utilizes a widely used preprocessing (soft arc consistency) and a Branch-and Bound techniques.





Agent
Model H
Internal State
$B \wedge I \wedge H \models O$
Background Knowledge B
Learning
Observation/ Experience O
Goal O
Time-series Data I
Decision Making
Action
Action
Action I
Dynamic Transition
Other Agents
Dynamic Environment

Some other topics in this meeting

❶ Intelligence into Resilience
❷ Resilience into Intelligence

- ❷ **Robust multi-team formation and its application to robot rescue simulation** (Tony Ribeiro)
- ❶ **Benefits of parametric model-checking to assess the resilience of mammalian circadian rhythm** (Morgan Magnin)
- ❶ **Understanding human behaviors through plan recognition** (Taisuke Sato)
- ❶❷ **On the evolution of beliefs in social networks** (Nicolas Schwind)
- ❷ **Limiting perturbations in Dynamic DCOP: Model with quality guarantee** (Maxime Clement)

Contents

3. Intelligence as Resilience

Are Human Resilient?

Are Memes Resilient?

Human Resilience

- Humans are capable to thrive after extremely aversive events (Bonanno, *American Psychologist*, 2004):
  - *resilience represents a distinct trajectory from the process of recovery*
  - *resilience in the face of loss or potential trauma is more common than is often believed*
  - *there are multiple and sometimes unexpected pathways to resilience.*
- Human knowledge can be explanatory and can have great reach (Deutsch, *The Beginning of Infinity,* 2011):
  - *"Rather than imitating behavior, a human being tries to explain it—to understand the ideas that cause it—which is a special case of the general human objective of explaining the world."*
  - *"Only progress is sustainable."*
- Both are allowed due to human *intelligence*.

4. Resilience as Intelligence

If humans are considered resilient due to their intelligence, future resilient systems should be designed to be intelligent too.

Secured AI (into the future)

- A long-term goal of “strong AI” is to develop systems that can learn from experience with human-like breadth and surpass human performance in most cognitive tasks.
- The use of AI techniques that significantly advance reliability in the low-level makes hardened systems much less vulnerable than today's. The design of anomaly detection systems and automated exploit-checkers could be significantly helpful☺
- AI systems will become increasingly complex in construction and behavior, and AI-based cyberattacks may be extremely effective☹
- It may be useful to create “containers" for AI systems that could have undesirable behaviors and consequences in less controlled environments.

Leakproofing the Singularity
(Yampolskiy, *J. Consciousness Studies*, 2012)

- Levels of communication security for confined AIs

levels	Outputs	Inputs	Explanation
0	Unlimited	Unlimited	Unlimited communication (Free AI)
1	Unlimited	Limited	Censored input, uncensored output
2	Unlimited	None	Outputs only with no inputs
3	Limited	Unlimited	Unlimited input and censored output
4	Limited	Limited	Secured communication (proposed protocol)
5	Limited	None	Censored output and no inputs
6	None	Unlimited	Inputs only with no outputs
7	None	Limited	Censored input and no outputs
8	None	None	No communication, fully confined AI

100 Year Study of Articial Intelligence
(Horvitz, Stanford University, 2014)

- **Privacy and machine intelligence**
- **Criminal uses of AI** —intelligent malware
- **Loss of control of AI systems**
  - ...we could one day lose control of AI systems via the rise of superintelligences that do not act in accordance with human wishes… Are such dystopic outcomes possible?
  - If so, how might these situations arise?
  - What kind of investments in research should be made to better understand and to address the possibility of the rise of a dangerous superintelligence or the occurrence of an “intelligence explosion“?
- **AI and philosophy of mind**
  - ...whether machines that we build might one day be conscious and find themselves “aware” and “experiencing” the inner or subjective world that people experience (?)