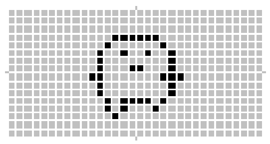
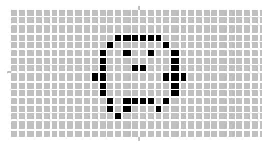
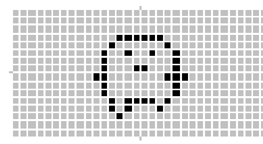
Can a Systems Biologist Fix a Tamagotchi?

Luca Cardelli

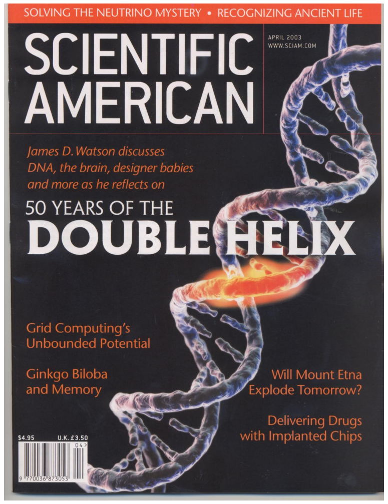
Microsoft Research

Gilles Kahn Colloquium
Paris 2007-01-12

http://LucaCardelli.name

The Chemistry of Nature

- We seem to have all the blueprints of life:
  - The structure of DNA is known since 1953.
  - The human DNA was decoded in 2001.
  - Craig Venter has decoded *his own* DNA!
  - All the information is (in principle) stored there.
  - And yet, there is very little progress in "fixing" diseases.

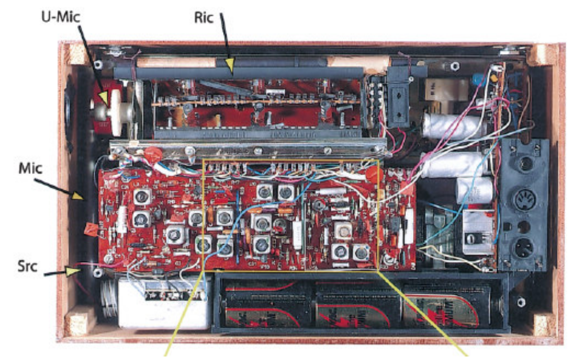


- Biology as Unknown Chemistry
  - View: Biology arises from complicated chemistry (out of standard physics).
  - Goal: We need to understand the chemistry better.
  - Approach: Groups of 10-50 biologists each studying 1 chemical for 10 years.

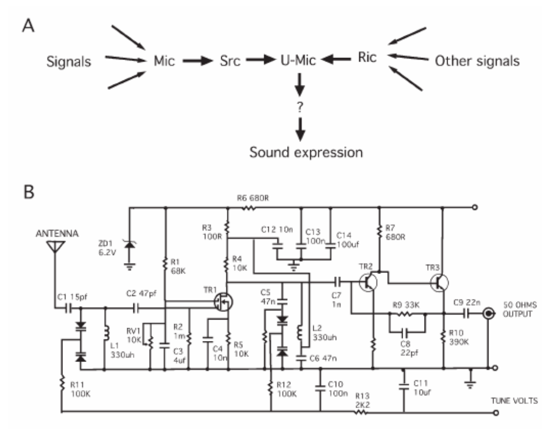
The Circuits of Nature

- Biology is more than chemistry
  - Yuri Lazebnik ("Can a Biologist Fix a Radio?"): Nature is like a radio, and biologists are like engineers who do not have a clue of electronics:

> *"We would eventually find how to open the radios and will find objects of various shape, color, and size […]. We would describe and classify them into families according to their appearance. We would describe a family of square metal objects, a family of round brightly colored objects with two legs, round-shaped objects with three legs and so on."*
>
> *"Because the objects would vary in color, we will investigate whether changing the colors affects the radio's performance. Although changing the colors would have only attenuating effects (the music is still playing but a trained ear of some people can discern some distortion), this approach will produce many publications and result in a lively debate."*



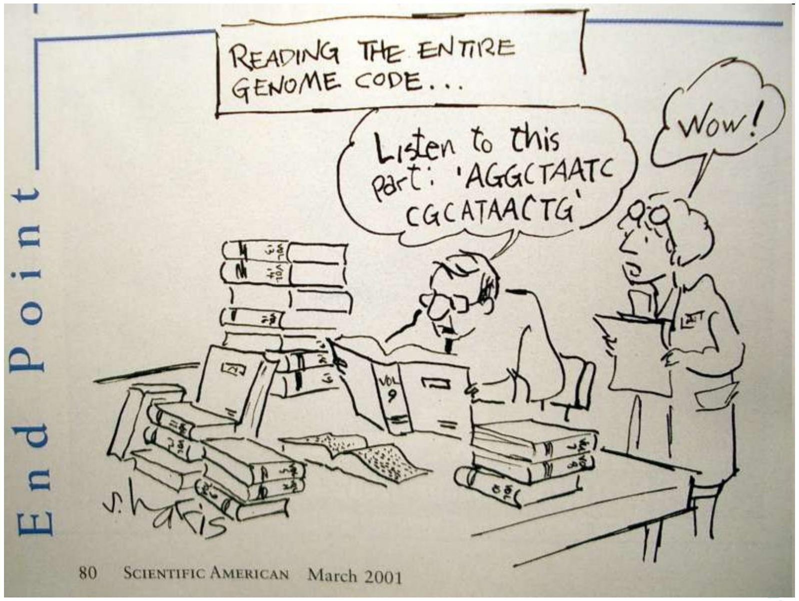
- Biology as Unknown Circuits
  - View: Biology arises from complicated *circuits* (over standard components).
  - Goal: We need to understand the circuits better.
  - Approach: "High throughput" experiments (Systems Biology).



The Programs of Nature

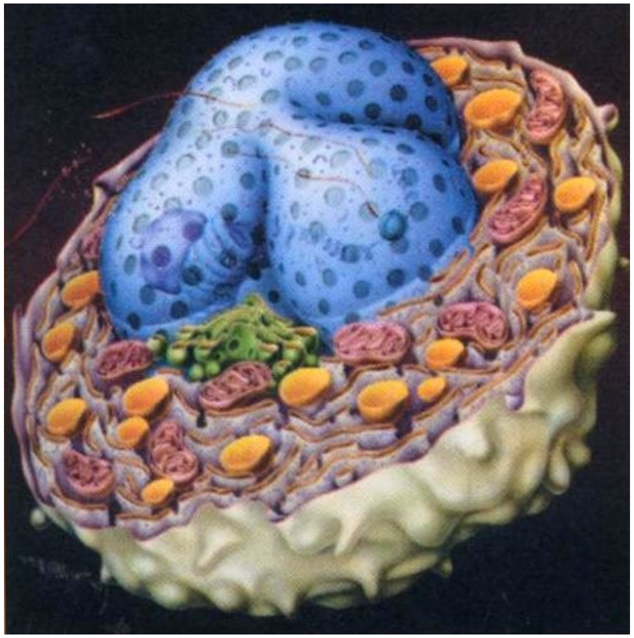
- Biology is more than circuits
  - Lazebnik's analogy is illuminating. But not accurate.
    - Biologists cannot understand radios without knowledge of electrical engineering.
    - But similarly, electrical engineers cannot understand mp3 players without knowledge of software engineering.
  - All life is software based:
    - Even the simplest bacteria have ~1M code (more than mp3 players!).
    - Hence, biologists will need to learn more than learn how to fix radios!

- Biology as Unknown Programs
  - View: Biology arises from complicated *programs* (over standard circuits).
  - Goal: We need to understand the programs better.
  - Approach: currently inadequate (comparable to software metrics).



A Technological Analogy

- The goal of Biology is:
  To reverse-engineer biological organisms.
  - That's hard.
  - Maybe impossible? Can we reverse-engineer something that was not engineered in any standard sense?

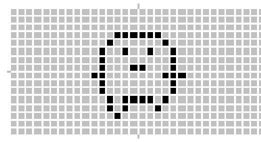


Reverse-Engineer This!

- Let's start instead with a *technological* organism.
  - We *know* it was engineered.
  - So it should surely be *possible* to reverse-engineer it.
  - The exercise will still reveal a significant subset of the difficulties of reverse-engineering biology.

T.Nipponensis

- Tamagotchi (*T.Nipponensis*) will be our "model organism".
  - It has inputs (buttons) and outputs (screen/sound)
  - It has state: happy or needy (or hungry, sick, dead…)
  - It displays a *cyberpet* that has to be "grown".



- *A good choice because:*
  - Its behavior rests in its software, not its circuits.
  - It is uncooperative and somewhat annoying.
  - It blurs the lines between organisms and devices.
  - Not a simple automaton:
    - Nondeterministic.
    - Stochastic.

T.Nipponensis

BANDAI

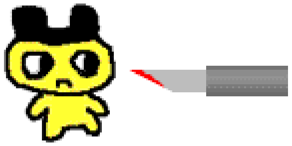
Q: How often do I have to exercise my Tamagotchi?
A: Every Tamagotchi is different. However we do recommend exercising at least three times a day

The Scientific Method

- Our task: Reverse-engineer *T.Nipponensis.*
  - Not by industrial espionage.
  - Not by rumors, speculation, or blogs.
  - But by the Scientific Method.

- The Scientific Method (SM) consists of:
  - Formulating falsifiable hypotheses about phenomena.
  - Running reproducible experiments that test the hypotheses.
  - Building models that explain all past experiments and predict the results of new experiments.

- Important applications of the Scientific Method:
  - Provide scientifically-based cyberpet consulting services.
  - Engineer and sell our own *T.Français.*
  - Fix broken Tamagotchi.



Approach 1: Understanding the Principles

- The assumption:
  - The organism underwent assembly and optimization by design or evolution.
  - In any case, general *principles* (deliberate or emerging) organize it.
  - Discovering such principles will help us in modeling the organism.

> *"Grouping facts so that general laws or conclusions may be drawn from them" -Darwin*

- Typical principled questions (first 2 not normally asked by biologists):

The creator

- Q1 of 4: Who created it?
  - Aki Maita. But it does not help:
  - Cannot hire creator as consultant (not SM).
  - Creator does not necessarily understand creation.

> *"Aki's own Tamagotchi seldom lives longer than its baby stage."*
> *-Apple Daily*

- Q2 of 4: Where is the documentation?
  - There is no publicly available design manual (trade secret).
  - Even if we had it, it would be in Japanese.

Approach 1: Understanding the Principles

- Q3 of 4: What is its function? What does a Tamagotchi compute?
  - It does not appear to compute anything in particular.
  - How can we understand its principles if we cannot say what it *does*?

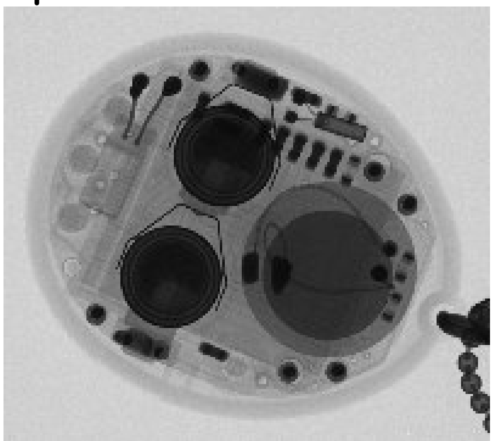
- Q4 of 4: Why does it have 3 buttons?
  - Evolved from extinct 2-button devices?
  - Is there a scaling law relating number of buttons to device size?
  - No rational explanation is immediately obvious.

- Therefore
  - Principled understanding of *T.Nipponensis* fails.

Approach 2: Understanding the Mechanism

- The assumption:
  - The organism is a *mechanism* driven by causes and effects.
  - Let's understand how it "ticks", and never mind the principles.

To understand "all of the forces that animate nature and the mutual positions of the beings that compose it" -Laplace



Tamagotchi X-Ray

- Typical mechanistic questions:

- Q1 of 4: What are the parts?
  - Hmm, very few parts, no moving parts.
  - Most interesting "part", the cyberpet, moves, but is not a part (cannot be extracted).

- Q2 of 4: How are the parts connected?
  - Parts connected by wires; pulling the wires has no clear effect.
  - And how is the cyberpet connected? If we cannot understand even that...

Approach 2: Understanding the Mechanism

- Q3 of 4: How does it react to perturbations?
  - Biospeak for “throwing a wrench in the clockwork”.
  - Removing or replacing parts make it stops working.
  - This approach works better in system with moving parts.
- Q4 of 4: How is it assembled?
  - Assembled in top-secret facilities at unknown locations not accessible to scientists.
  - By robots that are much harder to understand than *T.Nipponensis* itself.
- Therefore
  - Mechanistic understanding of *T.Nipponensis* fails.

Tamagotchi Surgery

Approach 3: Understanding the Behavior

- The assumption:
  - It is too difficult or premature to understand the mechanism.
  - But perhaps we can still model and predict the *behavior*.
  - And never mind who built it, how, or why.

"Measure what is measurable, and make measurable what is not so" -Galileo

- Typical behavioral questions:

Cyberpet Growth Chart

- Q1 of 4: How does it react to stimuli?
  - *By design* no consistent reactions! Nondeterministic and stochastic.
  - Experiments are not reproducible! (on individuals)

- Q2 of 4: How does it behave in a population?
  - Shake 1000 of them in a basket and plot resulting states.
  - Reproducible and publishable. But not terribly insightful.

Approach 3: Understanding the Behavior

- Q3 of 4: How does it communicate?
  - Antenna is *implicated* in communication (cannot conclusively determine its function behaviorally).
  - Proprietary protocol.

Communicating Tamagotchi

- Q4 of 4: How does it react to shock?
  - Extreme stimuli can test hypotheses about normal behavior.
  - Interesting results, but unrelated to cyberpet development.
  - May void your warranty.

- Therefore
  - Behavioral understanding of *T.Nipponensis* fails.

Approach 4: Understanding the Environment

- The assumption:
  - It is not possible to understand an entity in isolation or even as a population.
  - We must understand its relationships to its natural environment.

> *"The [Tamagotchi] modifies the conditions on the planet to make conditions on the planet more hospitable to it" -The Gaia Hypothesis*

- Typical environmental questions:

T.Nipponensis in its Natural Environment

- Q1 of 2: How did it evolve?
  - Japanese socio-economic environment: no consensus models.
  - It competed with now extinct devices (proprietary information).
  - Archeology (of Japanese dumps) may eventually help.
- Q2 of 2: How does it behave in its natural environment?
  - Kids hands and backpacks: impossible to reproduce in laboratory.
  - Hard to glimpse at the screen while in kids' hands.
  - Bulky remote tracking equipment affects behavior.
- Therefore
  - Environmental understanding of *T.Nipponensis* fails.

Approach 5: Understanding the Math

- The assumption:
  - The inexplicable effectiveness of Mathematics at explicating Nature.

    "Number rules the Universe" -Pythagoras

- Typical mathematical questions:

- Q1 of 1: What differential equations does *T.Nipponensis* obey?
  - Hmm...

- Therefore
  - Mathematical understanding of *T.Nipponensis* fails.

Reverse-Engineering Methods

- Why does the Scientific Method "fail" for *T.Nipponensis*?
  - We are trying to reverse-engineer a sophisticated piece of *software* (embedded in commodity hardware).
  - Which is very different from reverse-engineering, e.g., a radio.
  - No progress can be made just by looking at the cheap hardware, until we start disassembling the software!

- Reverse-engineering Software
  - How to do it:
    - Tracing, breakpointing, core-dumping, packet-sniffing, etc.
  - How *not* to do it (although it *might possibly* help):
    - Averaging over millions of core dumps. (*Cf.* proteomics)
    - Deleting exactly one line of code in turn. (*Cf.* genomic knockout)
    - Making plots of the number of stack frames over time. (*Cf.* transcriptomics)
    - Monitoring the power supply and heap size (*Cf.* metabolomics)
    - Measuring bandwidth over the connections (*Cf.* systems biology)

Standing Aghast

- Reverse-engineering the Software of Nature
  - What are the chances?
  - Probably incredibly hard:
    - 750MB (human genome) of which 40MB protein-coding.
    - Control structures still not understood.
    - Multiple levels of obscure encodings (although no deliberate encryption).
    - Riddled with internal (software) worms and viruses.
    - Full of legacy libraries, dead code, multiple coexisting versions.
    - Subject to random editing (evolution).
    - Very clever, highly optimized algorithms.
    - Self-modifying.
    - No specification! No documentation! No comments!
- Can a biologist fix a Tamagotchi?
  - Has to!

"The problem of biology is not to stand aghast at the complexity but to conquer it" -Sydney Brenner