A Fundamental Look at Models and Intelligence

Edward A. Lee

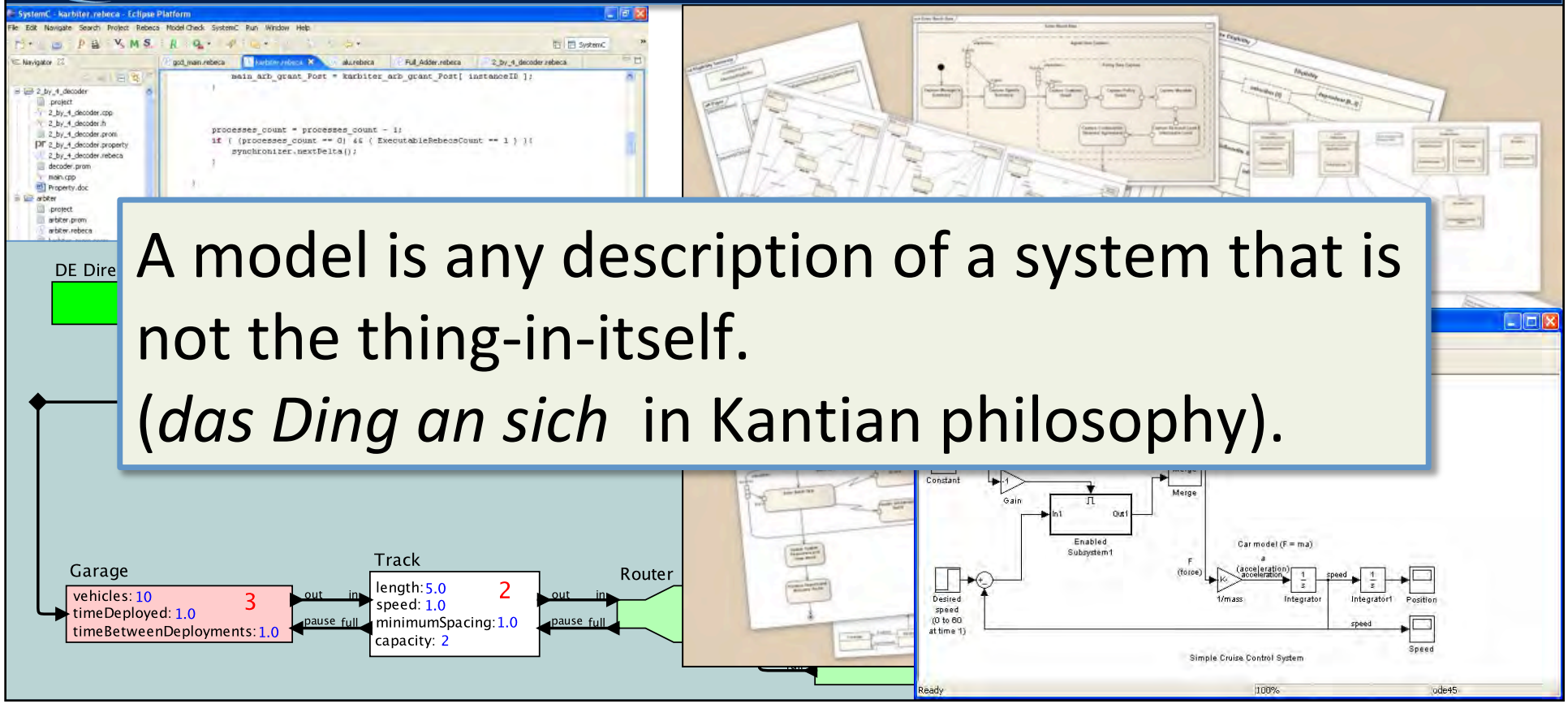
Keynote: Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS'19)

Montreal, Canada, May 28, 2019

University of California at Berkeley

What is a Model?

A model is any description of a system that is not the thing-in-itself.
(*das Ding an sich* in Kantian philosophy).



What is Intelligence?

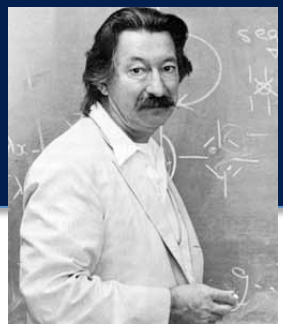
The Turing Test is all about imitating human intelligence, not about building intelligent systems.

```
ELIZA: Please tell me what's been bothering you.
YOU:   I think my computer is smarter than me.
ELIZA: Do computers worry you?
YOU:   Yes
ELIZA: You are sure.
YOU:   No, not really.
ELIZA: Are you saying no just to be negative?
```

elizabot http://www.masswerk.at/elizabot/

Joseph Weizenbaum
ELIZA, 1966

Is ELIZA Intelligent?



"[O]nce a particular program is unmasked, once its inner workings are explained in language sufficiently plain to induce understanding, its magic crumbles away; it stands revealed as a mere collection of procedures, each quite comprehensible. The observer says to himself `I could have written that.' With that thought he moves the program in question from the shelf marked `intelligent,' to that reserved for curios, fit to be discussed only with people less enlightened than he."

[Weizenbaum, 1966]

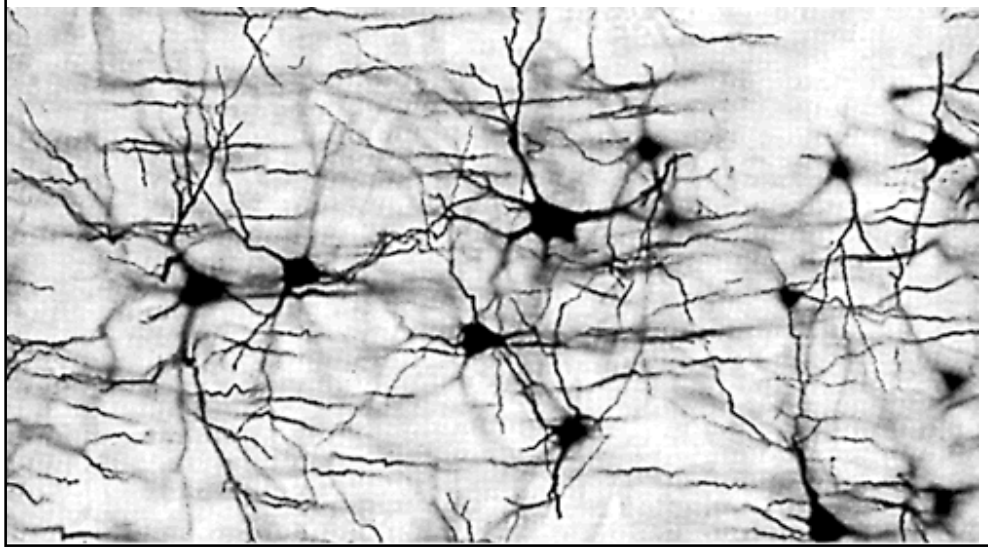
Explaining Natural Intelligence

- McColloch and Pitts (1940s)
- Rosenblatt: Perceptrons (1950s)
- Putnam: multiple realizability (1960s)
- Rumerlhart, Hinton, Williams: neural nets (1980s)
- Machine learning explosion (1910s)

Mechanical man as envisioned by an unknown sixteenth-century Italian master, contemporary of Leonardo da Vinci
[Web Gallery of Art, Public Domain]

Oversimplifying Natural Intelligence

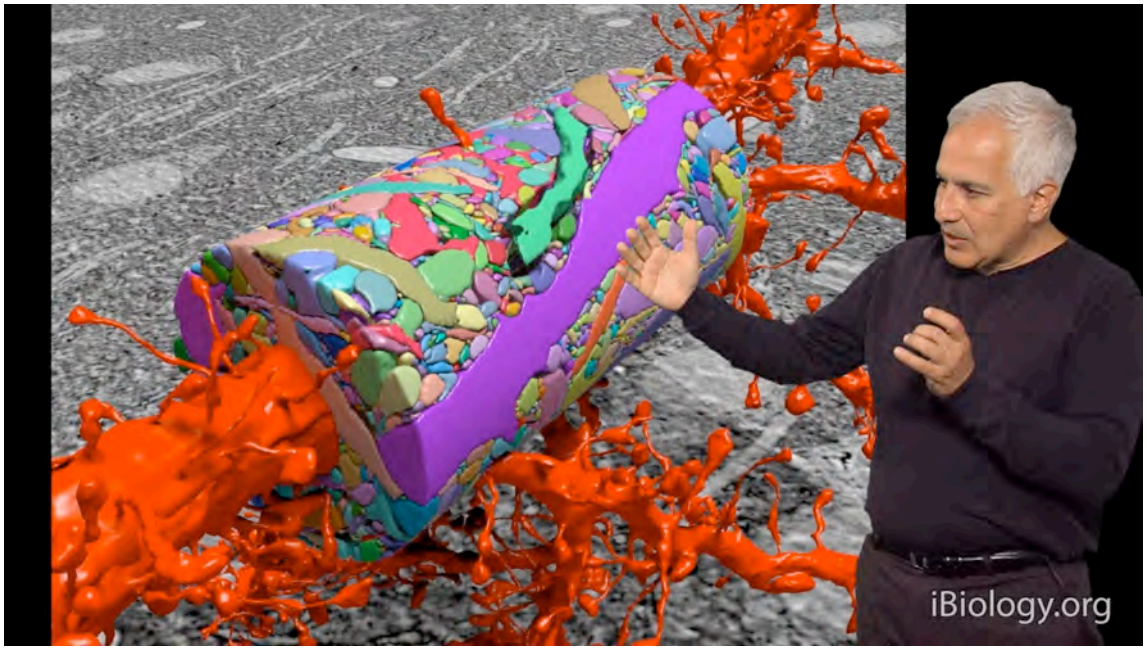
- Neurons fire discretely. (McCulloch and Pitts, 1940s)
- Neurons combine to realize logic functions.
- Logic functions can be realized on other hardware. (Putnam, 1960s)



Camillo Golgi's method (1870s) gives a misleading picture of the brain.

Explaining Natural Intelligence

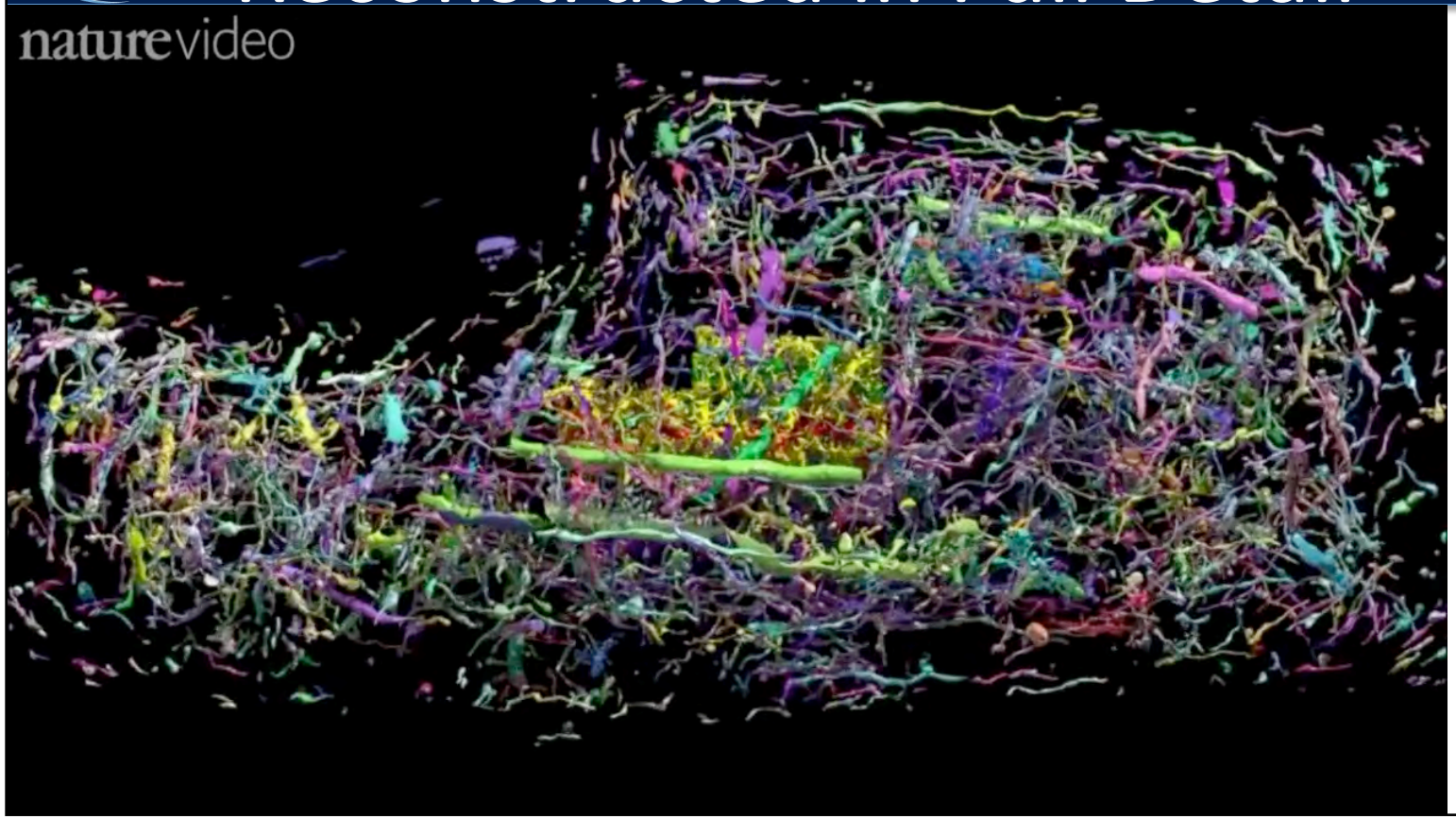
Can we understand brain function by studying the wiring diagram, even in principle?



Jeff Lichtman, Harvard

Lee, Berkeley

Crumb of Mouse Brain Reconstructed in Full Detail

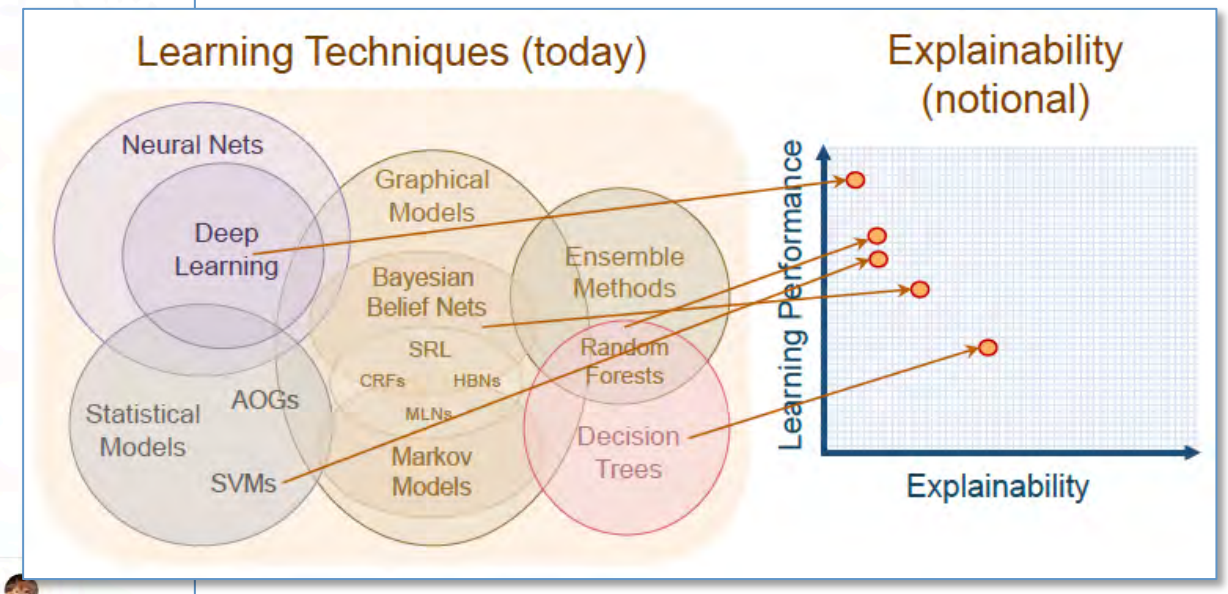


Nature News, 7/30/15, by Alison Abbott

Jeff Lichtman, Harvard

Explainable *Artificial* Intelligence

DARPA XIA Program:



David Gunning, Program Manager, XIA

Google Inception

Electric Guitar,
Acoustic Guitar,
Labrador

Marco Túlio Ribeiro, Sameer Singh, Carlos Guestrin (Univ. of Washington, 2016)

Recognizing Cars and Pedestrians

Self-driving cars need to recognize cars and pedestrians.

Do we understand how humans do this?

Lee, Berkeley

How can we build models of something
we do not understand

?

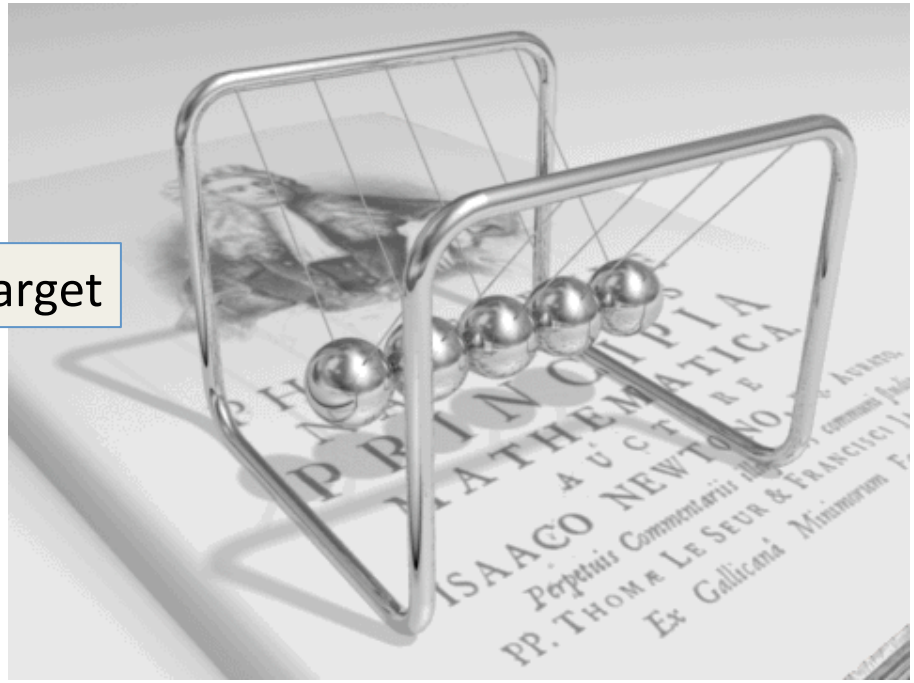
Explainable Mechanical Systems (XME)

$$x(t) = x(0) + \int_0^t v(\tau)d\tau$$

$$v(t) = v(0) + \frac{1}{m}\int_0^t F(\tau)d\tau$$

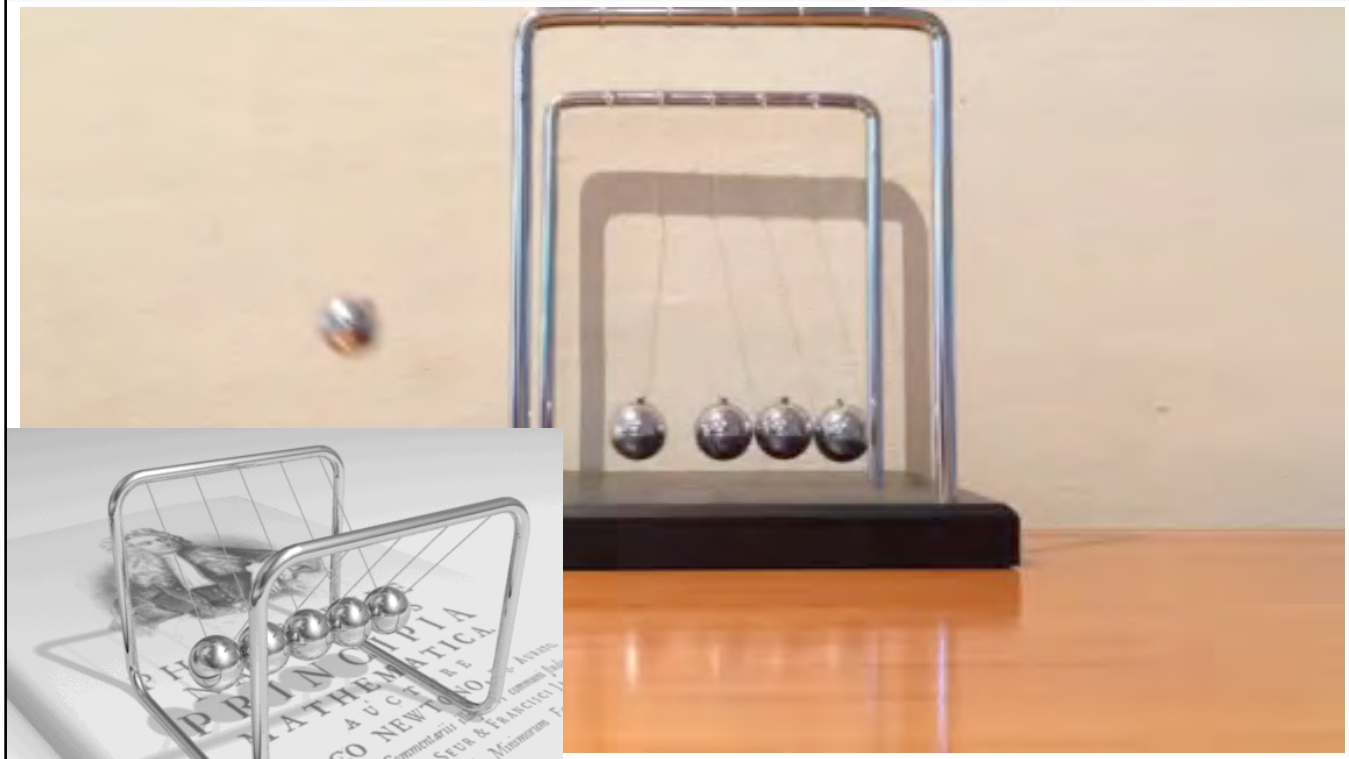
The model

The target



In this example, the *modeling universe* is calculus and Newton's laws in a time and space continuum.

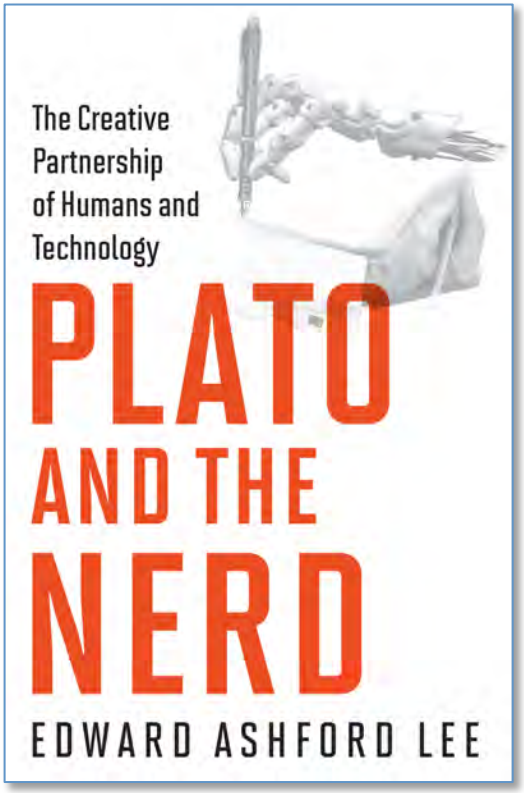
Unexplainable Mechanical Systems



A few things we need to model to explain this behavior:

- Plastic deformation
- Acoustic propagation
- Stretching of strings
- Gravity
- …

An Epiphany



Lee, Berkeley

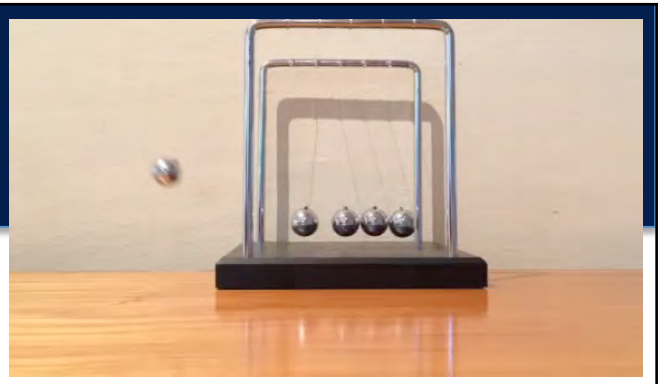
The Value of Models

- In *science*, the value of a *model* lies in how well its behavior matches that of the physical system.

- In *engineering*, the value of the *physical system* lies in how well its behavior matches that of the model.

A scientist asks, “Can I make a model for this thing?”
An engineer asks, “Can I make a thing for this model?”

Model Fidelity

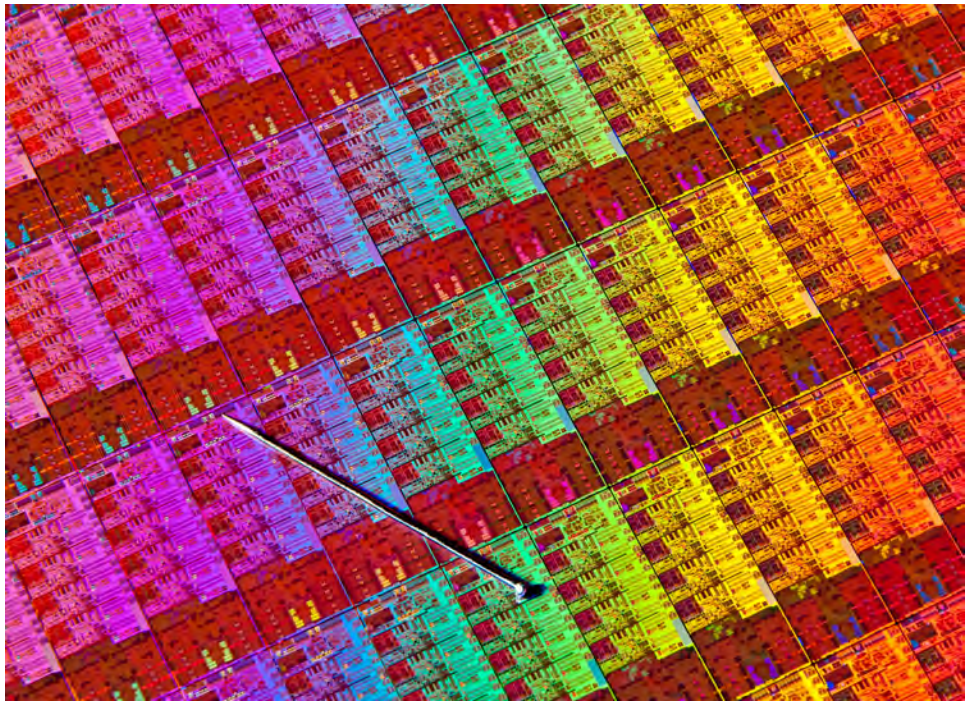


- To a *scientist*, the model is flawed.
- To an *engineer*, the realization is flawed.

I'm an engineer…

Perhaps we should be trying to make systems behave in more intelligent ways rather than trying to build systems that emulate human intelligence.

Consider Chip Design



Intel Haswell, each with 1.4 billion transistors

A piece of silicon that doesn't behave like the model is just beach sand.

Useful Models and Useful Things

"Essentially, all models are wrong,
but some are useful."

Box, G. E. P. and N. R. Draper, 1987: *Empirical Model-Building and Response Surfaces*. Wiley Series in Probability and Statistics, Wiley.

"Essentially, all system implementations
are wrong, but some are useful."

Lee and Sirjani, "What good are models," FACS 2018.

The Value of Simulation

"Simulation is doomed to succeed."

Could this statement be confusing engineering and scientific models?

Figure 1: Three scenes generated from a single ~20-line SCENIC scenario representing bumper-to-bumper traffic.

[Fremont, et al., Scenic: Language-Based Scene Generation, Arxiv.org, Sept. 2018]

How can we build models of something
we do not understand

?

Do we want scientific models or engineering models?

Changing the Question

Is the question whether we can build models that behave like natural intelligent systems?

Or

Is the question whether we can build systems that behave like intelligent models?

More Intelligent Systems May Not Resemble Humans at All

Self awareness:

Consider a thermostat, miswired so that the heat control is connected to the AC and vice versa.

Which one will be more intelligent?

Scientific Model-Based Design of Intelligent Systems

- Model the human brain
- Build systems based on those models

If we are successful, then every morning, I will have to argue with my smart car about the value of getting to work on time…

Does it Really Matter Whether It is A Gorilla?

A Story

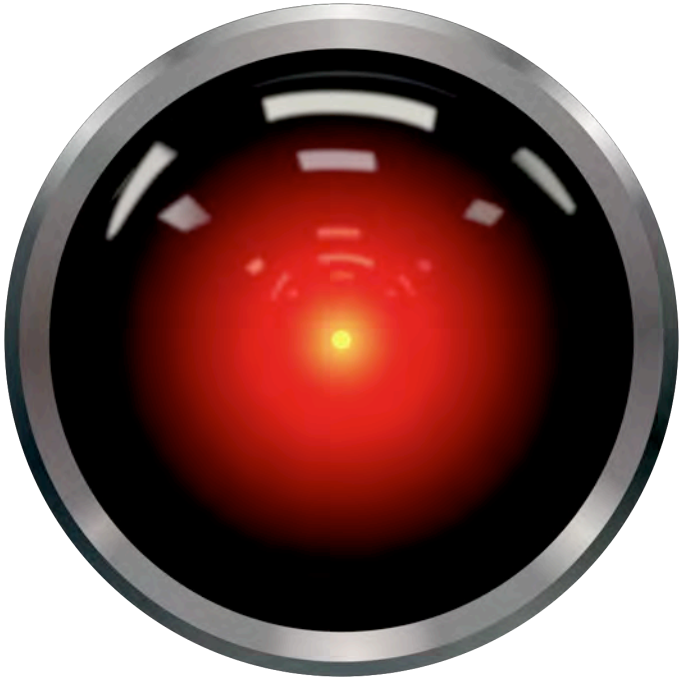
Rich Caruana
Microsoft Research

Should patients with pneumonia be admitted to the hospital or treated at home?

Found that on a training dataset, patients with a risk of asthma had a *lower* risk of dying from pneumonia than the general population.

Caruana, et al., 2015: Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)

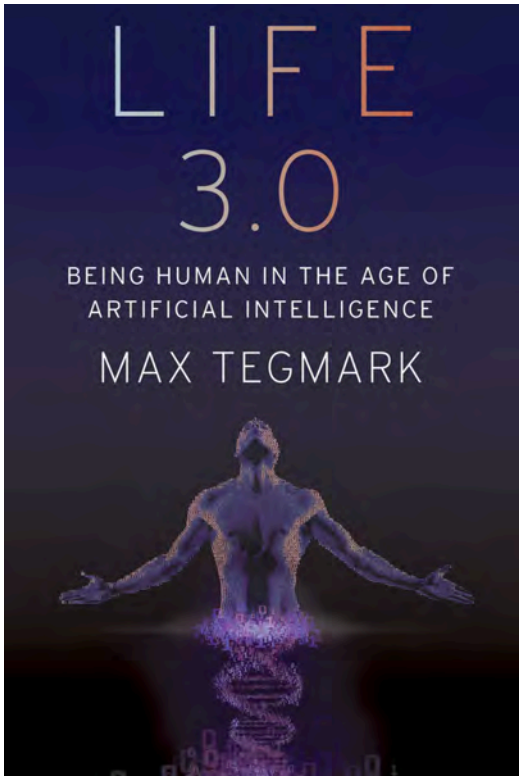
We May Not Yet Have Invented the Technology to Replicate Human Intelligence



HAL, the computer in Stanley Kubrick's 1968 movie, *2001: A Space Odyssey*

Freeing the Mind From Matter

- Are we alone?
- Teleportation?
- The singularity?
- Uploading?



Teleportation and Uploading

What happens to "I"?

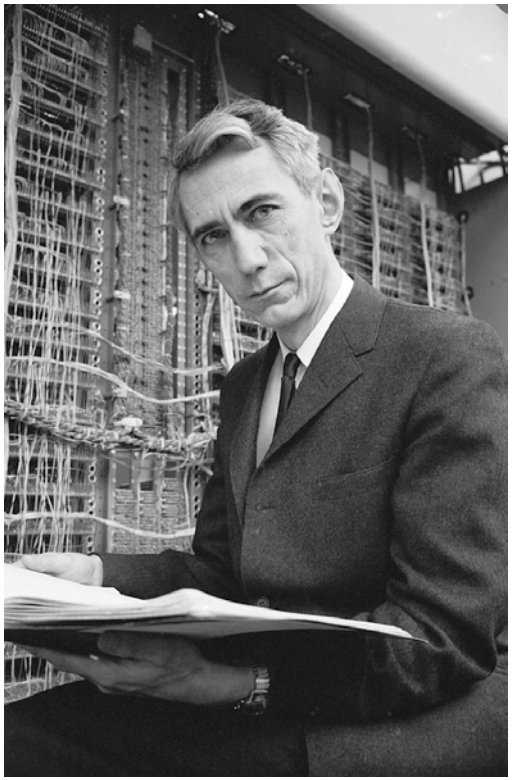
- Is the reconstruction the same "I"?
  - How can we tell?
- What if the original is not destroyed?
  - Two "I"s?
- What if a backup copy is later instantiated?
  - Two "I"s of different ages?

The Sense of Self Per Three Philosophers

What happens to "I"?

- Derek Parfit:
  - The notion of "I" makes no sense.
- Daniel Dennett:
  - "I" is a fiction, an illusion, a social construction.
- Douglas Hoftstadter
  - "I" can be in two places at once.

A Simpler Answer: "I" Is Not Digital

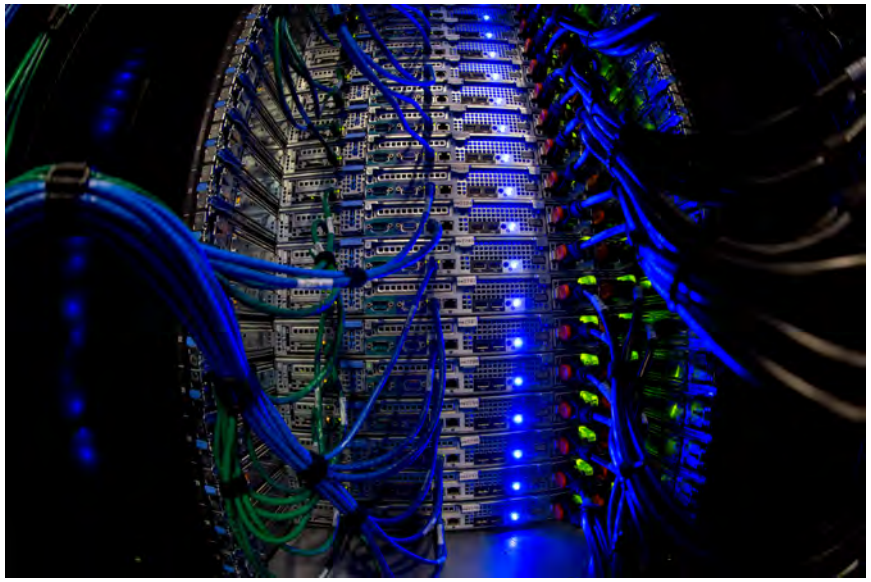


Claude Shannon

Shannon showed in 1948 a noisy channel can, in principle, perfectly convey a finite number of bits (the "channel capacity").

The converse is even more important: A noisy channel *cannot* convey more than a finite number of bits.

Digital and Computational Machines



Physical realization
(Wikimedia Foundation servers)
[Photo by Victor Grigas/Wikimedia Foundation, CC BY-SA 3.0]

Turing machine:

An abstract machine that, given a finite digital input, either computes a finite digital output or fails to halt.

Universal Turing machine:

A Turing machine that can realize any other Turing machine, given a digital representation of that machine.

A Universal Turing Machine is *Not* a Universal Machine

By Piotrus, CC BY-SA 3.0, via Wikimedia Commons

A machine that is (probably) not modeled in any useful way by a Turing machine.

It is neither digital nor algorithmic.

How Many Machines?

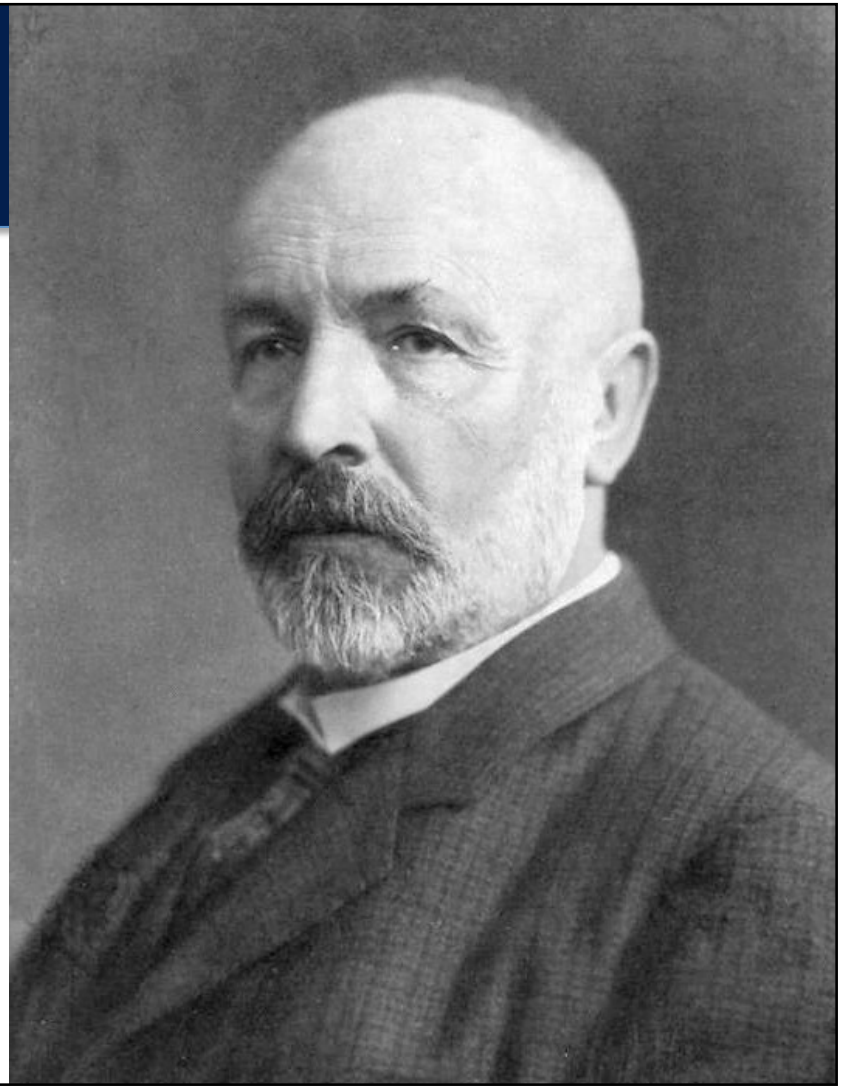
My question:

Of all machines realizable in the physical world, how many are usefully modeled as digital and algorithmic machines?

Cantor and Cardinality

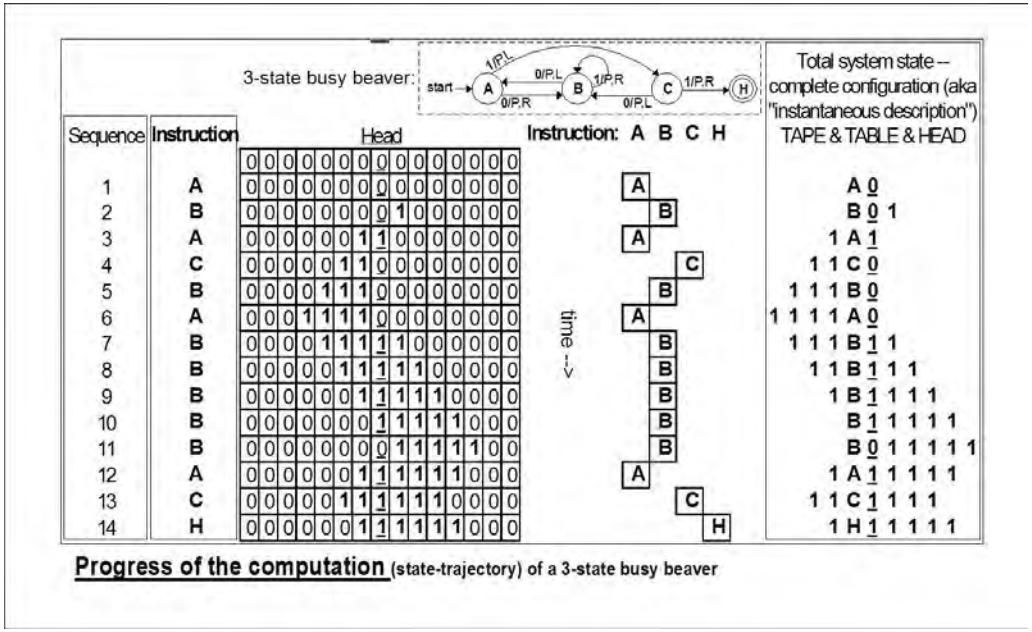
The smallest of all infinite sets are the countable sets.

Elements of a countable set can be put into a one-to-one correspondence with the natural numbers:
0, 1, 2, 3, 4, …



Georg Cantor (1845-1918)

The Set of Turing Machines is Countable



Progress of the computation (state-trajectory) of a 3-state busy beaver

By Wvbailey CC BY-SA 3.0

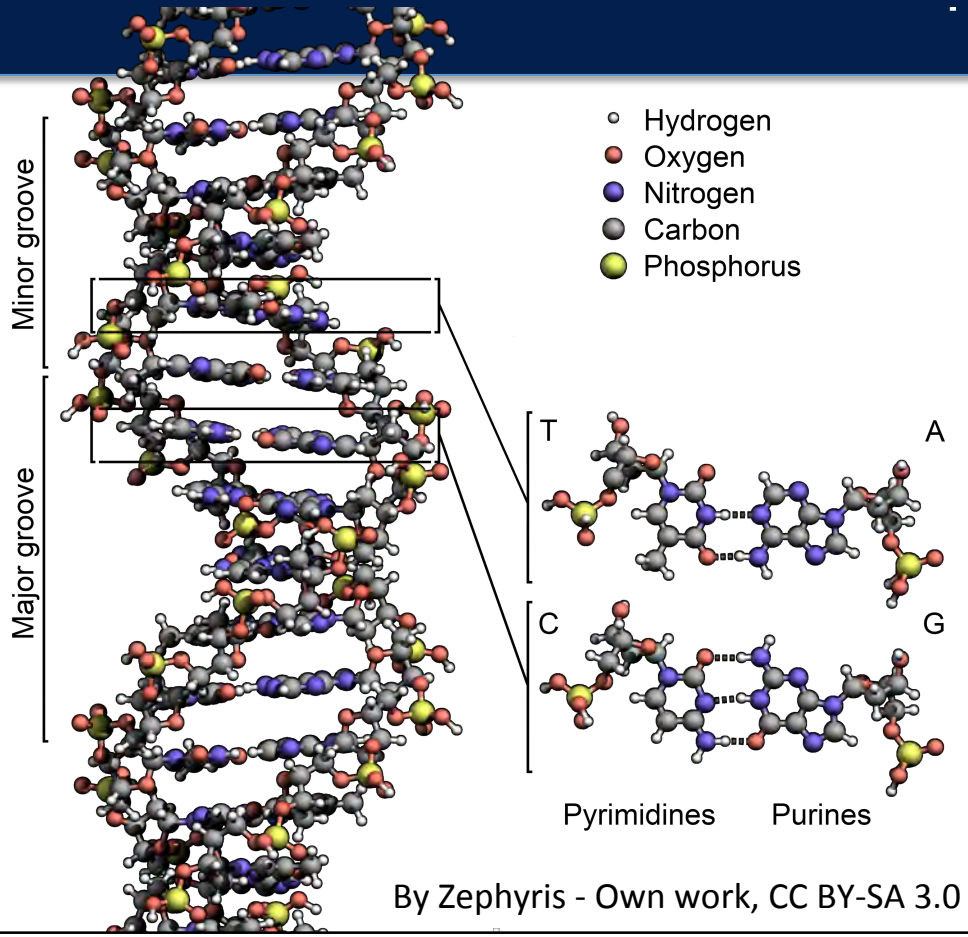
Countable subset

Uncountable set
E.g. set of all ODEs

The DNA Fallacy

Naïve assumption: DNA encodes the information needed to create a human (and hence, a mind).

Every human alive today is the endpoint of a continuous, unbroken, biological process dating back about four billion years.



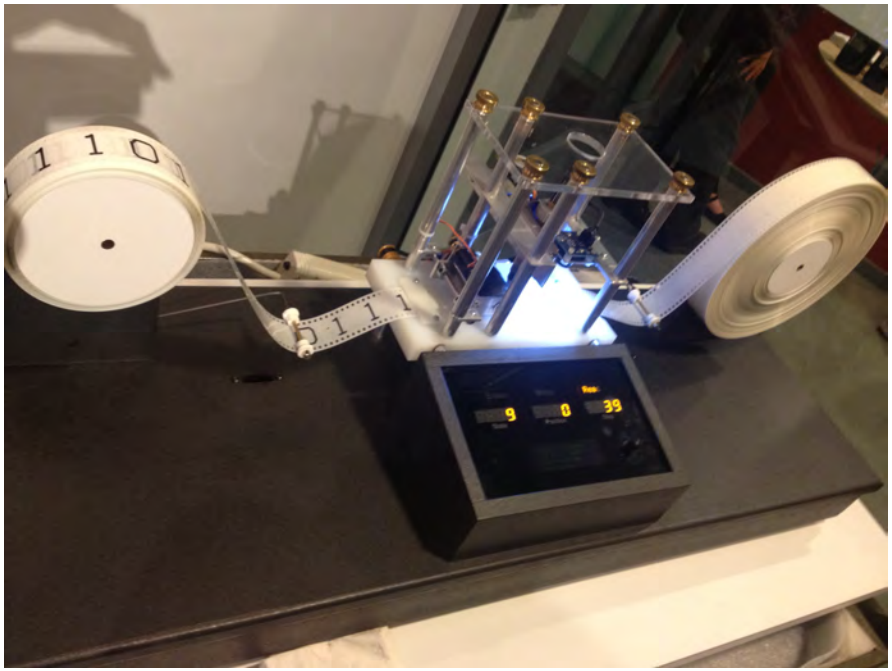
By Zephyris - Own work, CC BY-SA 3.0

The "Universal Machine" Fallacy

Naïve assumption: Turing machines are universal machines.

They are:

- Algorithmic
- Digital
- Non-interactive
- Terminating



Machine designed by Mike Davey
By GabrielF - Own work, CC BY-SA 3.0

What Are the Odds?

Is human intelligence
digital and computational?

We need evidence…

Processes in nature

Is The Human Mind Digital and Computational?

Many aspects are discrete:

- Language
- Logic
- Mathematics

All subject to the Shannon Channel Capacity Theorem.

Evidence that it is not:

- Consciousness is not externally observable.
- Language imperfectly encodes our thoughts.
- Most truths in our heads never get expressed.
- Communicated ideas are never perfect.

Possible Mechanisms in the Brain that are Beyond Digital/Computational

- Timing
- Interaction
- Chaos (induced by feedback loops)
- Nondeterminism
- Chemistry
- Embodiment

Most Chaos is Beyond Computable

Naïve assumption:
If we can write down the equation, it must be computable.

Lorenz attractor:

$$\begin{aligned}\dot{x}_1(t) &= \sigma(x_2(t) - x_1(t)) \\ \dot{x}_2(t) &= (\lambda - x_3(t))x_1(t) - x_2(t) \\ \dot{x}_3(t) &= x_1(t)x_2(t) - bx_3(t)\end{aligned}$$

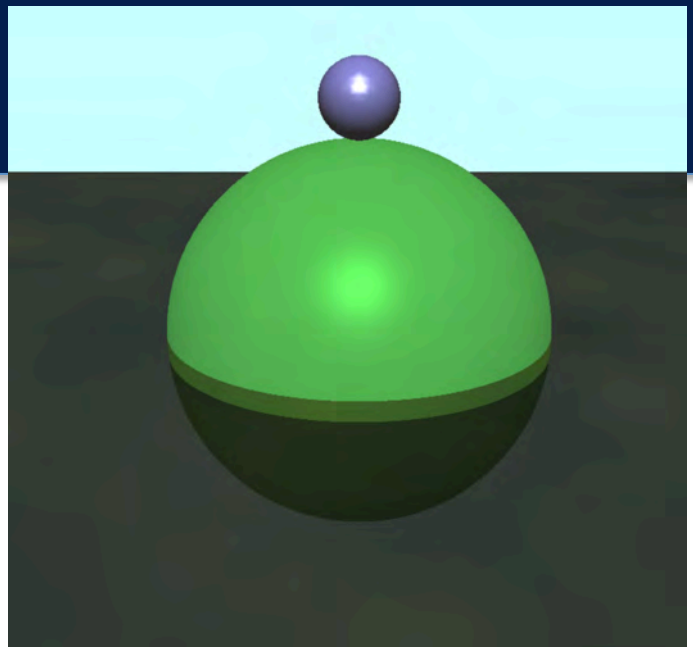
This is a chaotic system, so arbitrarily small perturbations have arbitrarily large consequences.

Edward Lorenz

Determinism

Naïve assumption:
Newtonian mechanics is deterministic.

Pierre-Simon Laplace



Metastable system that obeys all of Newton's laws but is nondeterministic.

Norton, J. D. (2007). Causation as Folk Science.
In Causation, Physics, and the Constitution of Reality
Oxford, Clarendon Press:

Lessons from Psychology: Embodied Cognition

"The mind simply does not exist as something decoupled from the body and the environment in which it resides."

[Thelen, E., 2000: Grounded in the world]

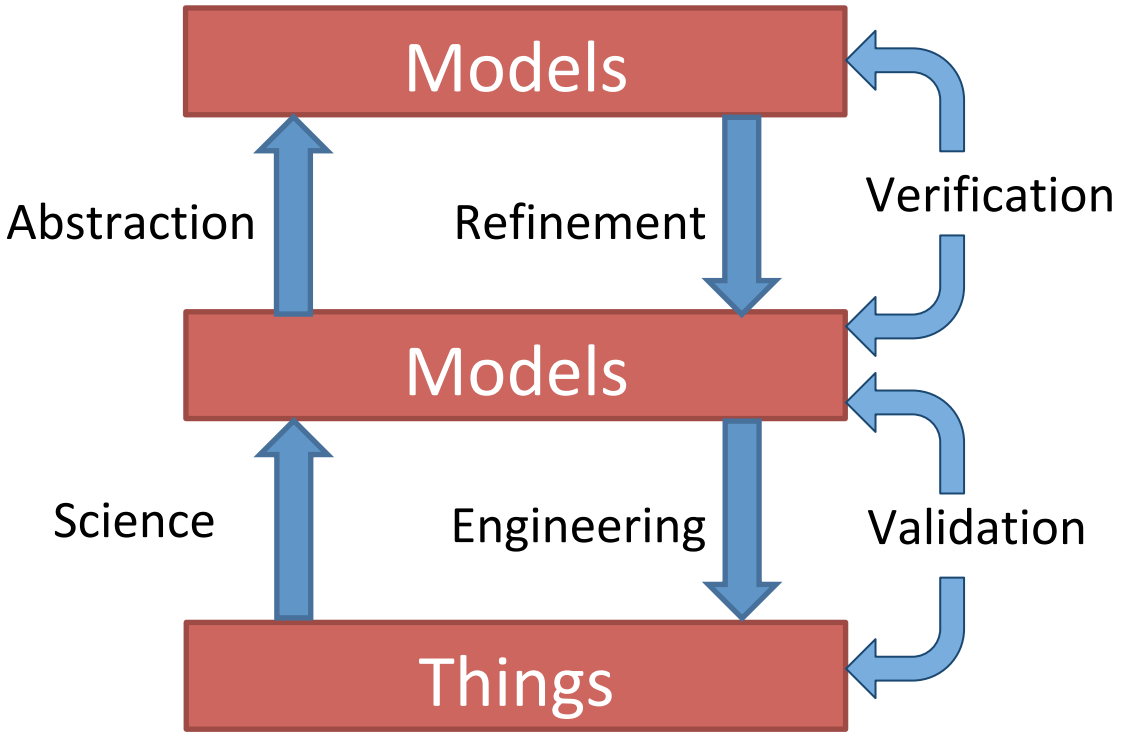
Human-like AI will more likely arise from cyber-physical systems and cyber-human systems, not cyber ones.

Esther Thelen (1941—2004)
Developmental psychologist,
Indiana University, pioneer of embodied cognition.

Towards Engineering-Model-Based Design of Intelligent Systems

Per Boehm:

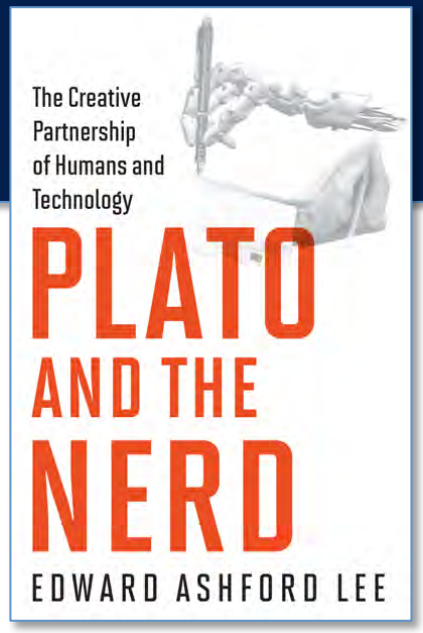
- Am I building the right product? (validation)
- Am I building the product right? (verification)



Conclusions

We can (and do) build models of things we don't understand.

The pertinent question is not whether our models accurately reflect intelligence in humans, but rather whether we can build physical artifacts that behave like intelligent models.



MIT Press, 2017