iCyPhy

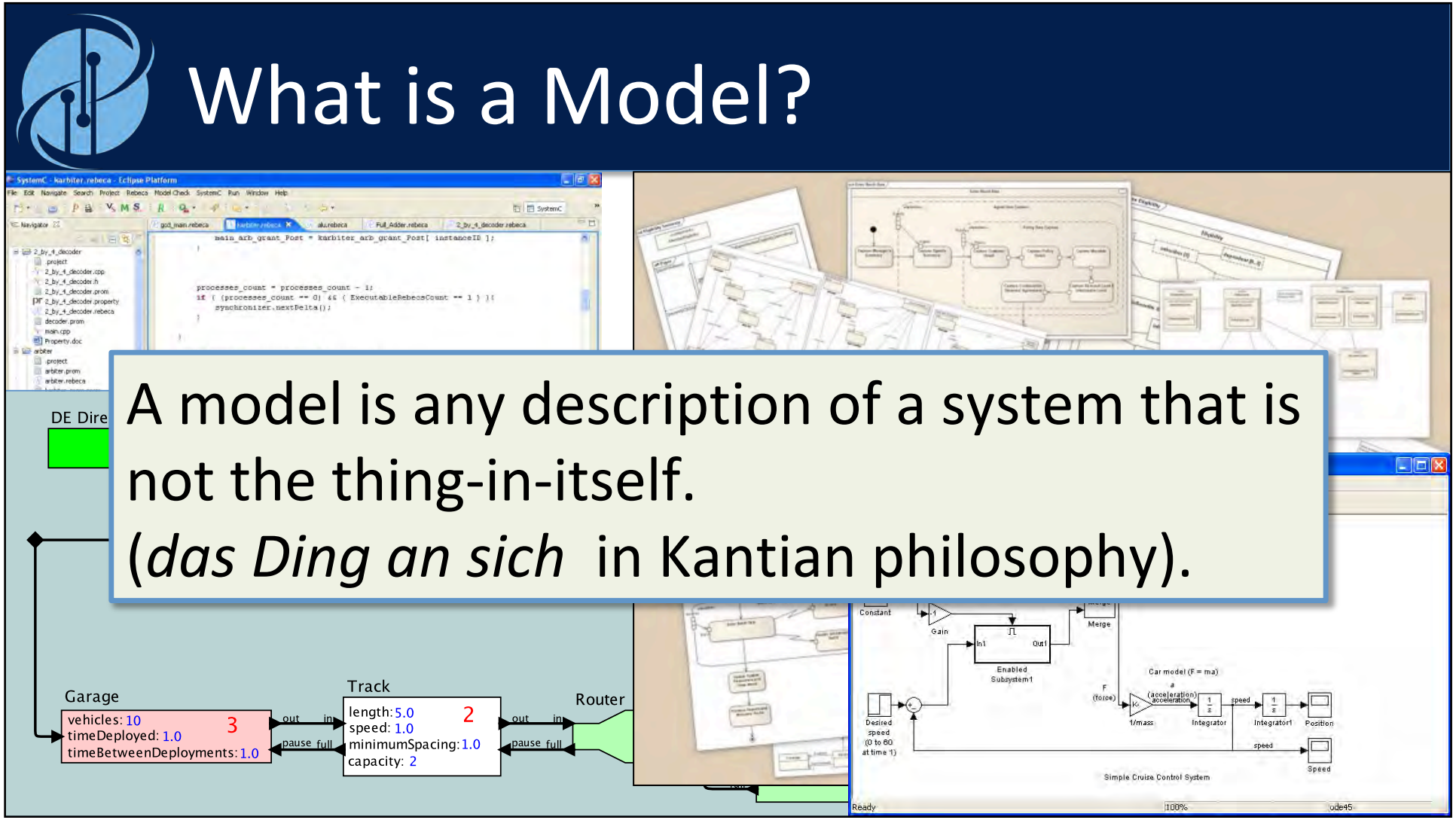
A Fundamental Look at Models and Intelligence

Edward A. Lee

DATE Special Day on Model-Based Design of Intelligent Systems

Florence, Italy, March 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?

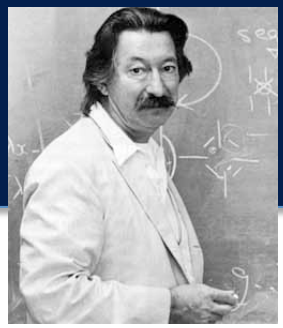
The Turing Test is all about imitating human intelligence, not about building intelligent systems.

```
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]

Your Speaker in 1966

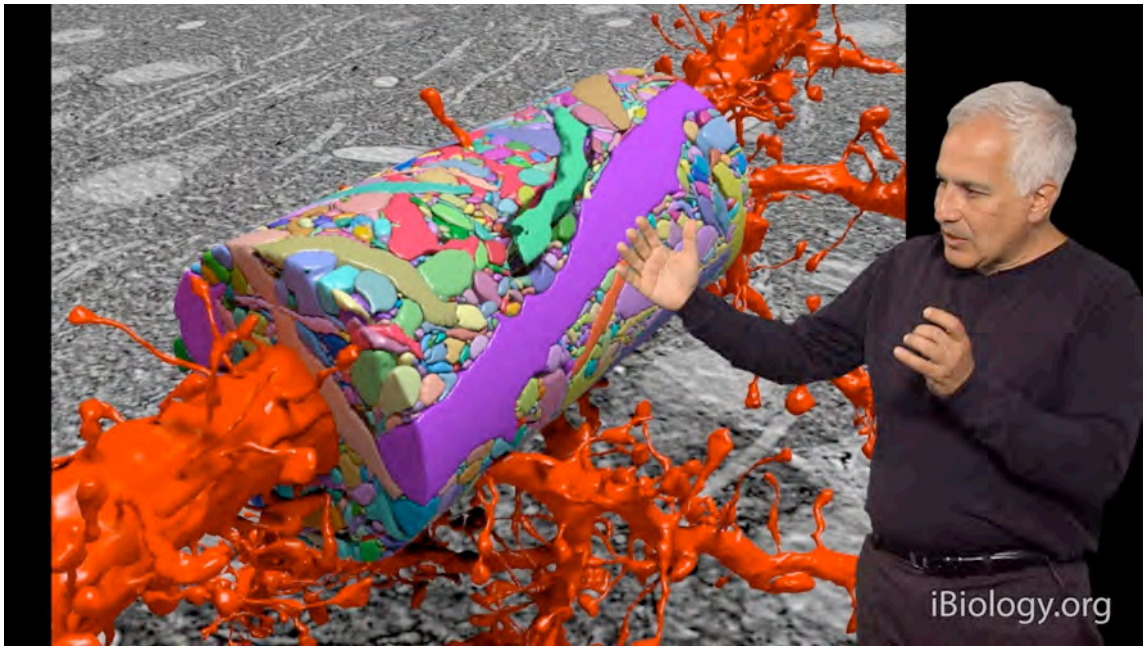
Explaining Natural Intelligence

- 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]

Explaining Natural Intelligence

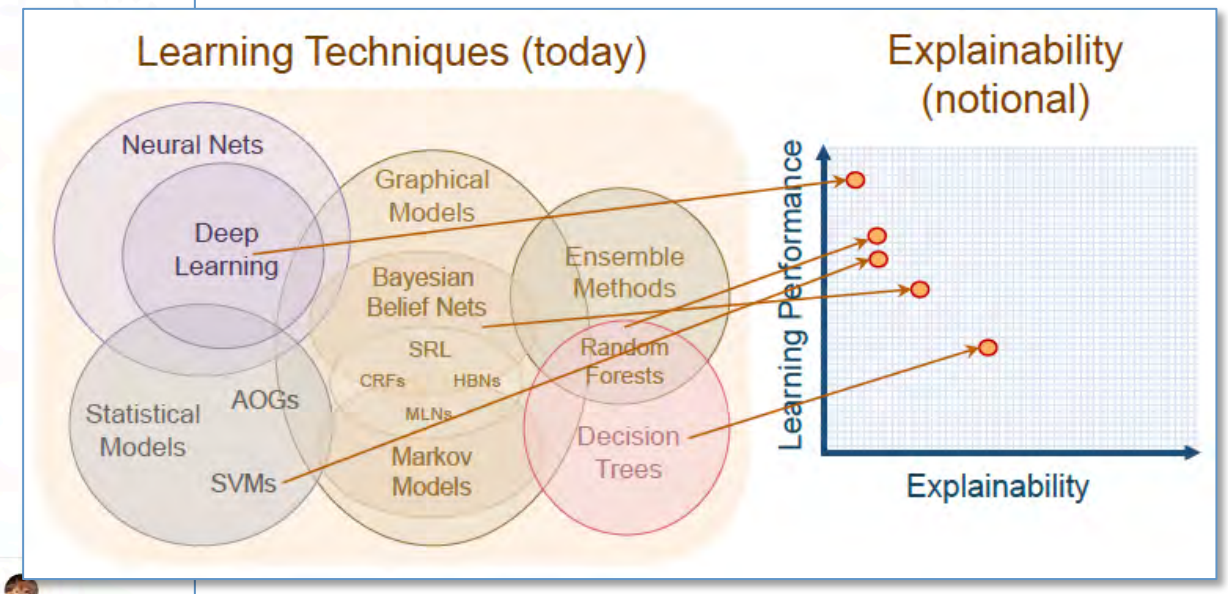
Can we understand brain function by studying the wiring diagram, even in principle?



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?

How can we build models of something
we do not understand

?

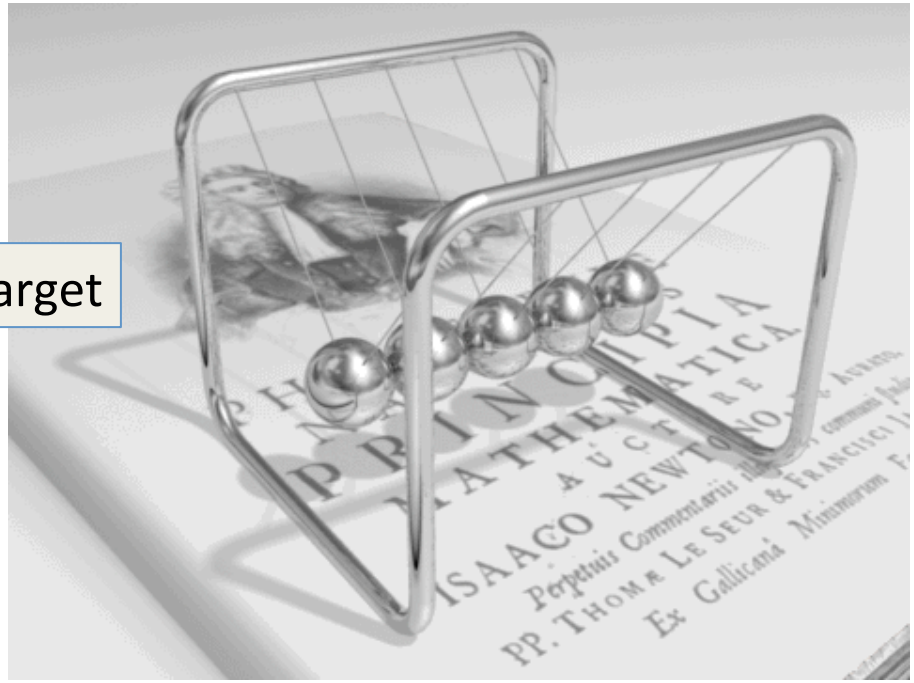
Explainable Mechanical Systems (XME)

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

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

The model

The target

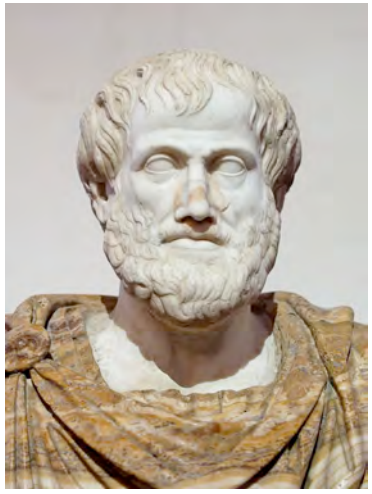


In this example, the *modeling universe* is calculus and Newton's laws in a time and space continuum.

Do We Understand Time?

We understand time, right?

Change

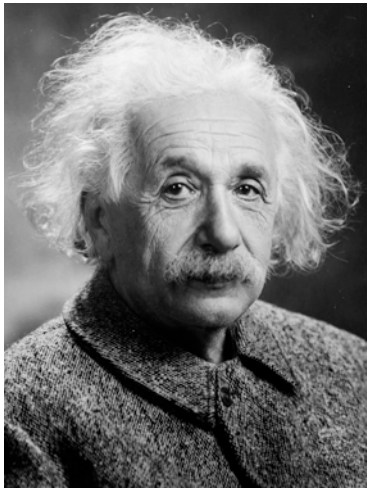


Aristotle

Smooth

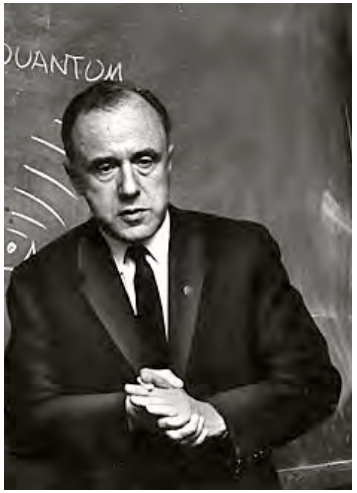
Newton

Relative



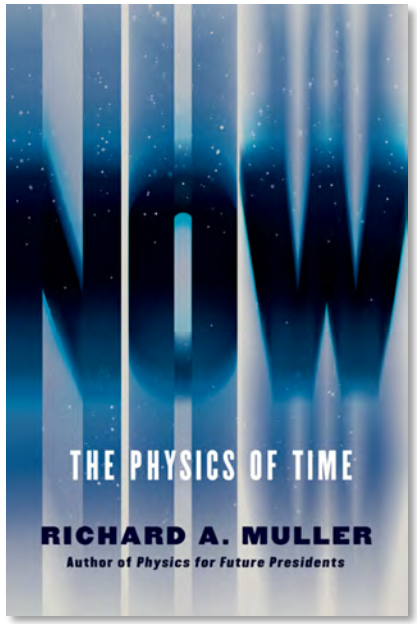
Einstein

Discrete



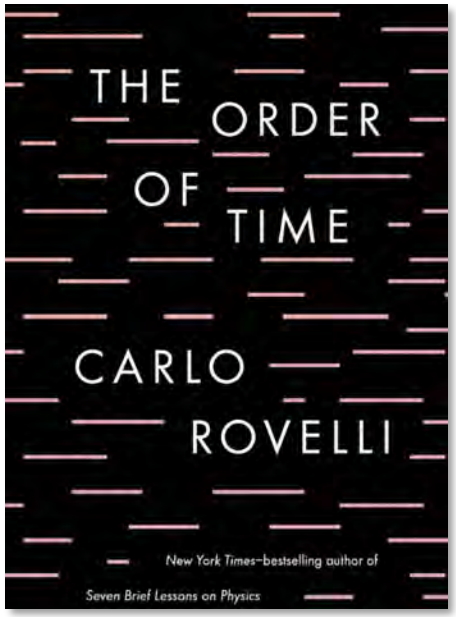
Wheeler

Do We Understand Time?



2016

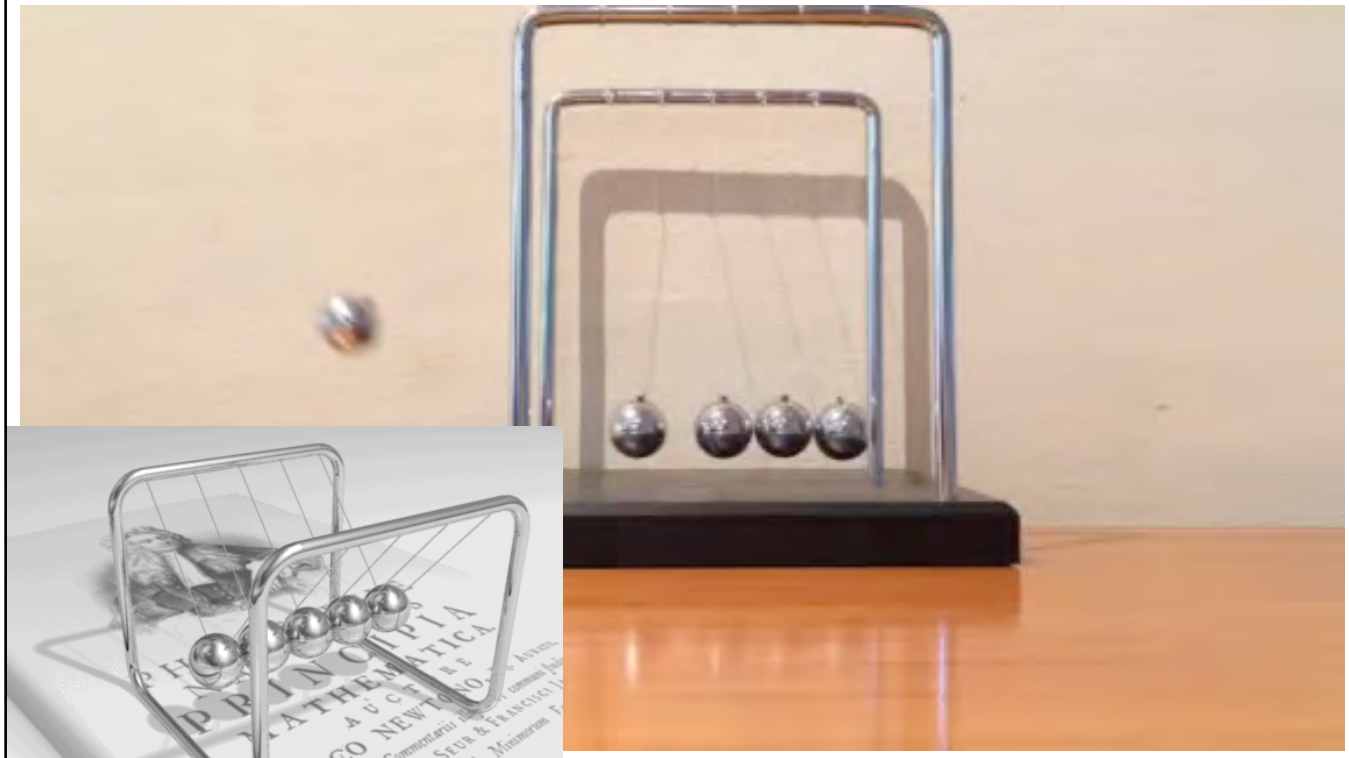
Muller: Gives a theory of time that requires big black holes to collide somewhere near us to test it.



2018

Rovelli: "The nature of time is perhaps the greatest remaining mystery."

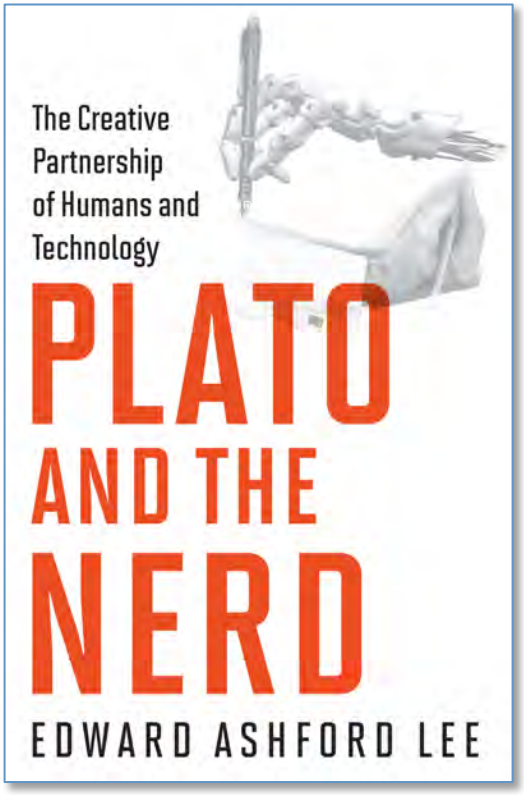
Unexplainable Mechanical Systems



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

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

An Epiphany



Lee, Berkeley

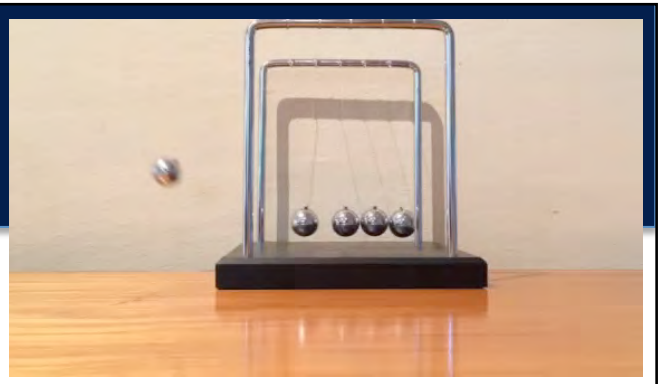
The Value of Models

- 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

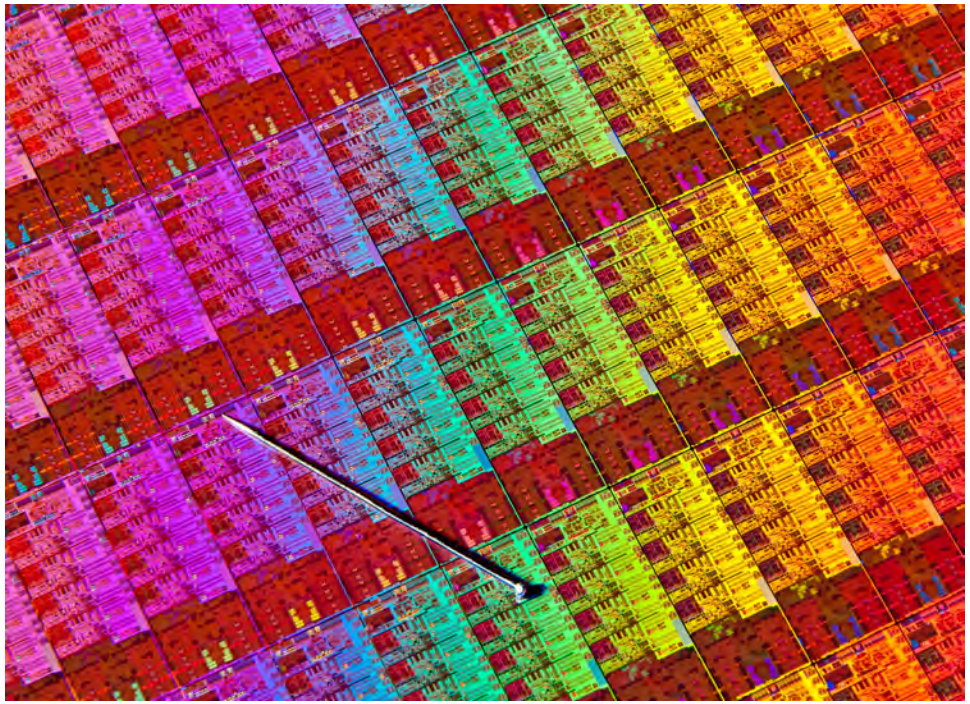


- To a *scientist*, the model is flawed.
- To an *engineer*, the realization is flawed.

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.

[Freemont, 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 describing the behavior of intelligent systems?

Or

Is the question whether we can build systems whose behavior matches that of intelligent models?

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)

More Intelligent Systems May Not Resemble Humans at All

Self awareness:

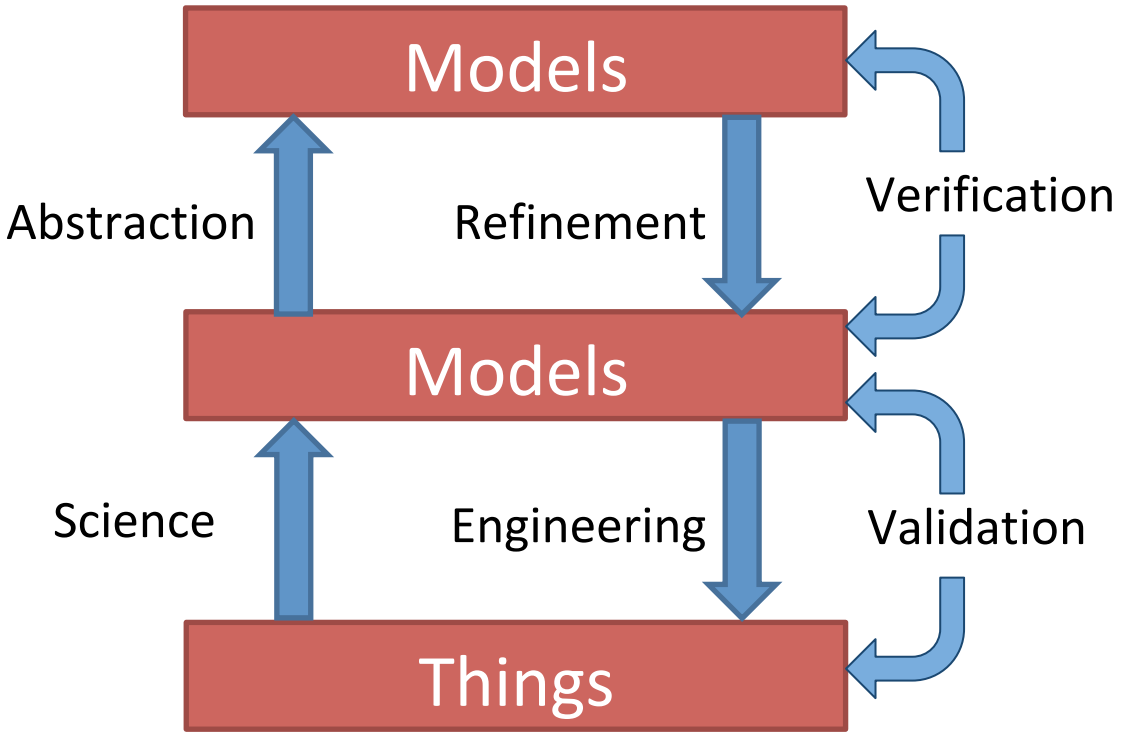
Consider a thermostat, miswired so that the heat control is connected to the AC and vice versa.

Which one will be more intelligent?

Towards Engineering-Model-Based Design of Intelligent Systems

Per Boehm:

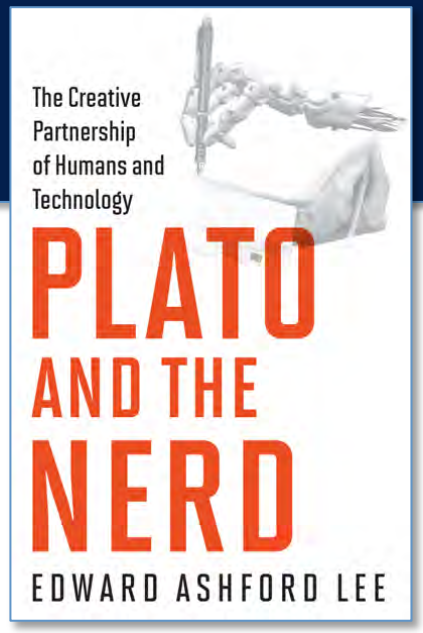
- Am I building the right product? (validation)
- Am I building the product right? (verification)



Conclusions

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

The pertinent question is not whether our models accurate reflect the physical world, but rather whether we can build physical artifacts that reflect our models.



MIT Press, 2017