Minimally Naturalistic AI

Steven Hansen

Outline

1. The Allegory of the Play-doh
2. No Free Lunch
3. Meta-Learning
4. Imitation Learning
5. Moving Forward
6. Suggestions for COMM-AI

Allegory Time!

Imagine you have a ball of play-doh

Allegory Time!

Make a door-stop

Allegory Time!

Make a paper-weight

Allegory Time!

Make a spear

Allegory Time!

Make an hamburger

Allegory Time!

Make a computer

Allegory Time!

Works fine:

- Door-stop
- Paper-weight

Needs to be harder:

- spear

Needs to be more edible:

- hamburger

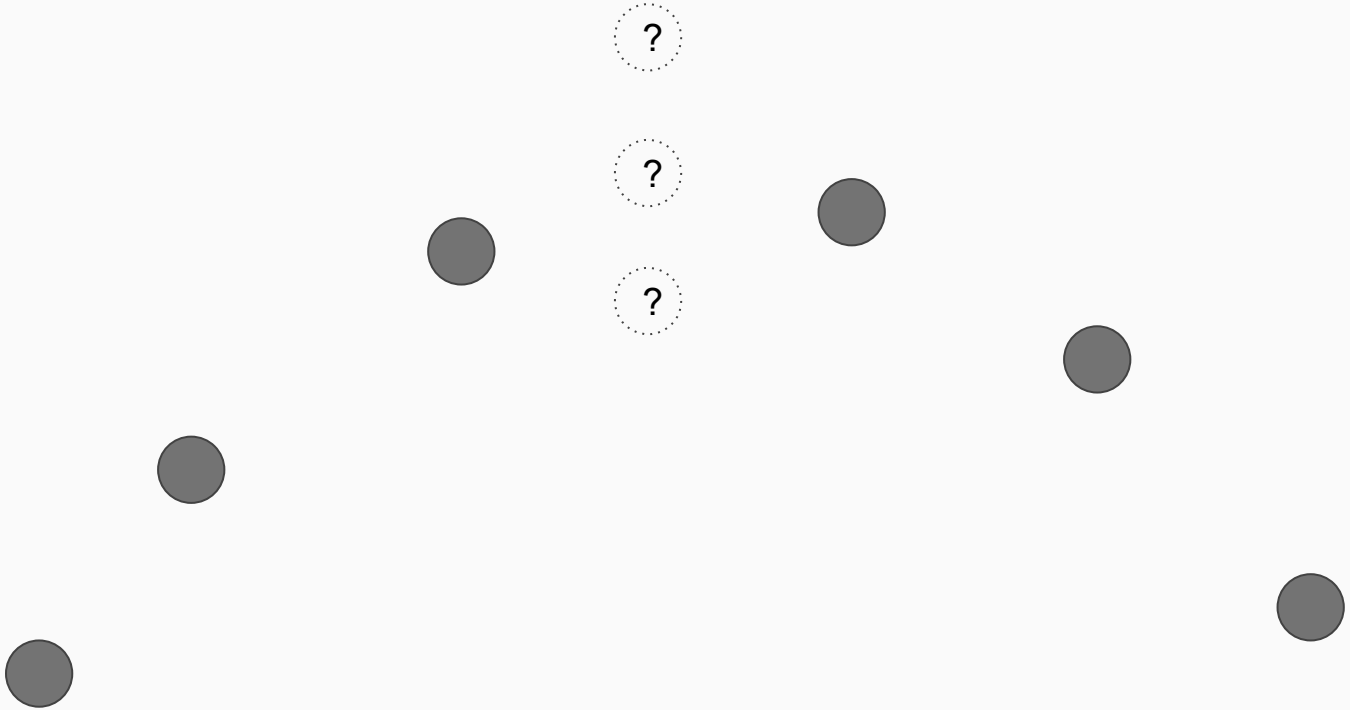
Needs to be more of a superconductor:

- computer

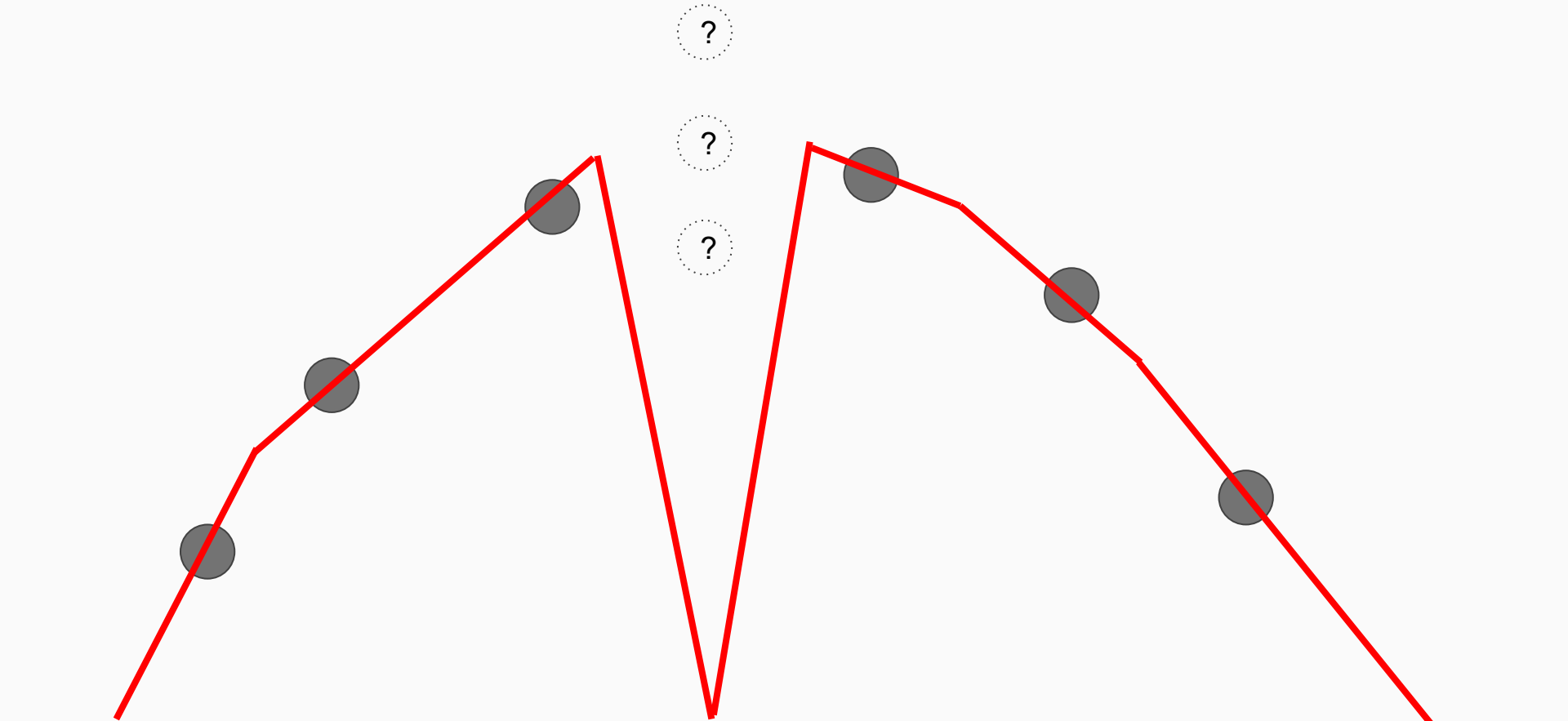
Moral

- Inductive biases must sharpen as task complexity rises
- The closer we get to human-level AI, the more naturalistic the tasks we must train on

No Free Lunch



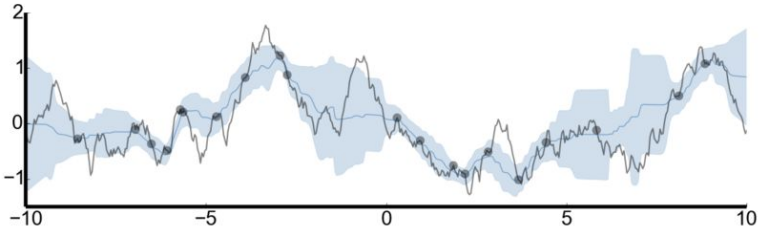
No Free Lunch



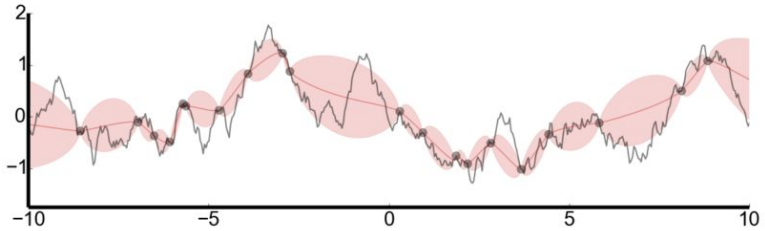
But deep learning just works...

- **Explicit priors aren’t the only way we shape the inductive bias**
- Convolutions and 2D equivariance
- RNNs and repeated computation
- Clockwork-RNNs and periodicity
- NTMs and... turing machines

Meta-learning / Learning-to-learn



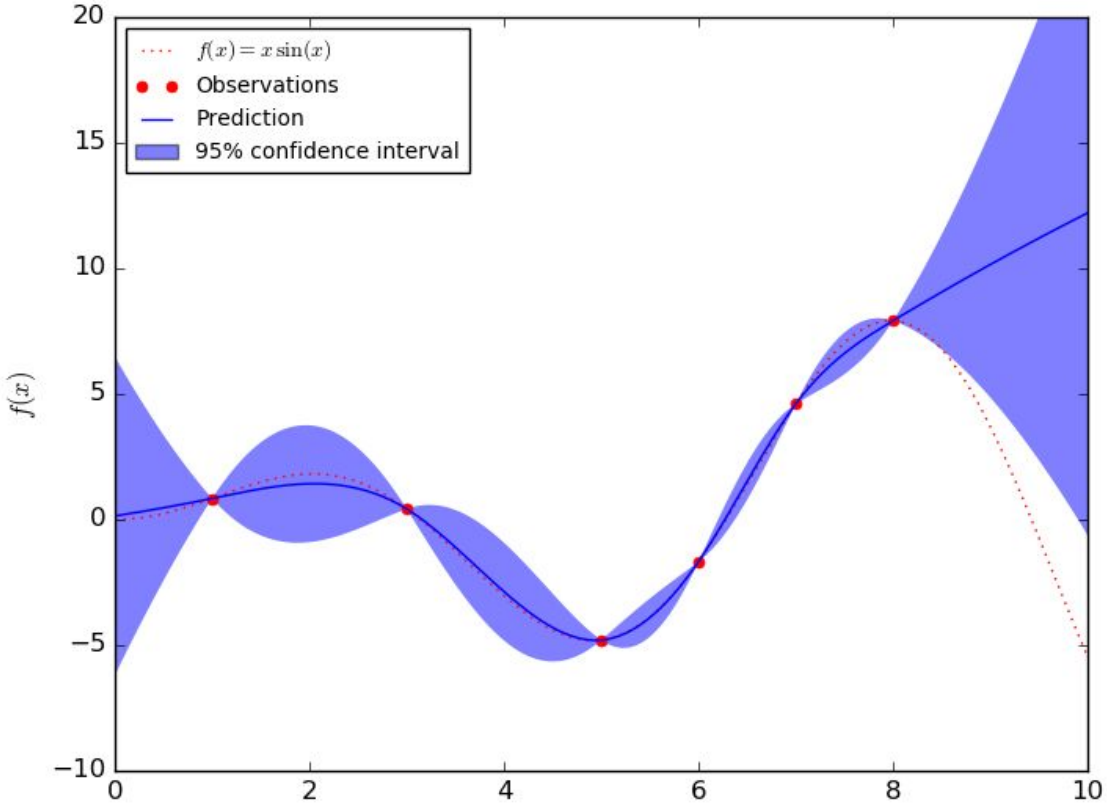
(a) MANN predictions along all x-inputs after 20 samples



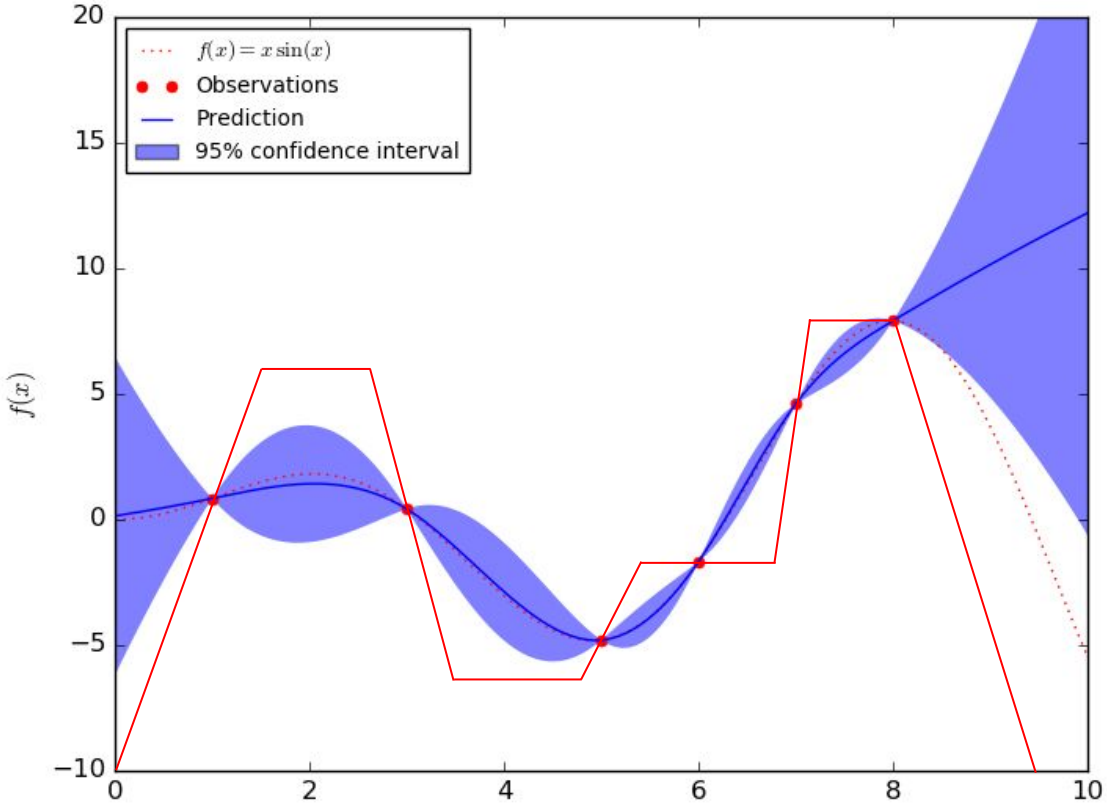
(b) GP predictions along all x-inputs after 20 samples

- From tasks to task distributions
- Learn an algorithm that can generalize from few samples
- “One-shot learning with Memory-Augmented Neural Networks”
  - Santoro et al 2016

An Even Less Free Lunch



An Even Less Free Lunch



Imitation Learning

- Inverse reinforcement learning, apprenticeship learning, goal inference
  - Supervised learning++
- Learn to copy your mentor by inferring their values/goal
  - Generalize better than copying behavior
- Who is the mentor? A human or a program written by one.
- Are we worth copying in artificial environments?
- Would our goals in such environments have the same structure as in natural environments?
- These questions bound the naturalism required

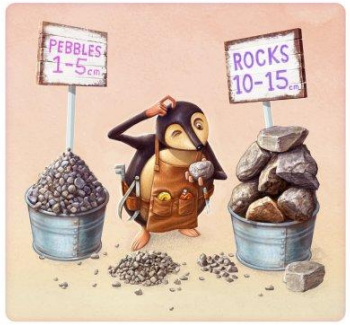
Moving Forward

1. Identify the next milestone where humans outperform AI
2. Look for the regularities in that environment and in human performance
3. Create artificial environments that still contain those regularities
4. Look again if the AI fails to scale the real thing
5. Remember that the regularities needed might include any previous encountered environments!

Some Regularities for COMM-AI

- Communication tasks tend to be encountered in a structured way
  - The participants take into account each other's intelligence
  - Tasks tend to be somewhat periodic
- Communication is grounded in sensory modalities
  - Visual structure
  - Auditory emotion cues
- Communication allows for rich feedback
  - Observations of coherent episodes
  - Occasional corrections

Fuzzy



Scaleless

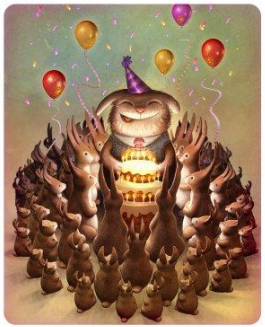
Granular

Emergent

Resources

Energy

Reproduction



Multi-agent

Diversity

Credit: Drew Purves

Thanks for listening!

Questions?