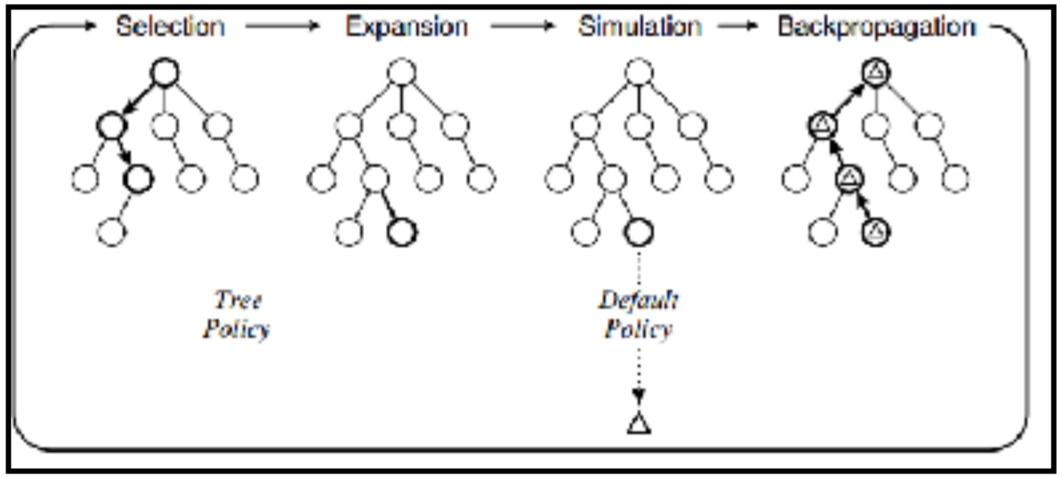
ROBERT LONG (NYU)

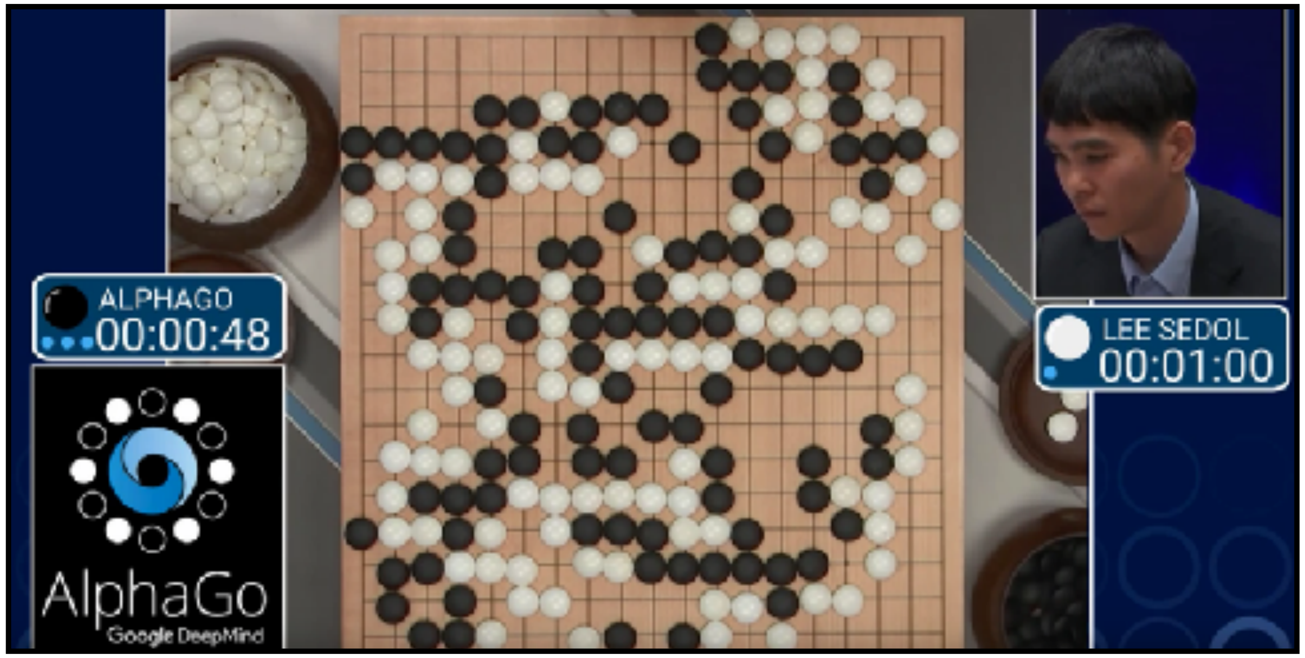
MIT

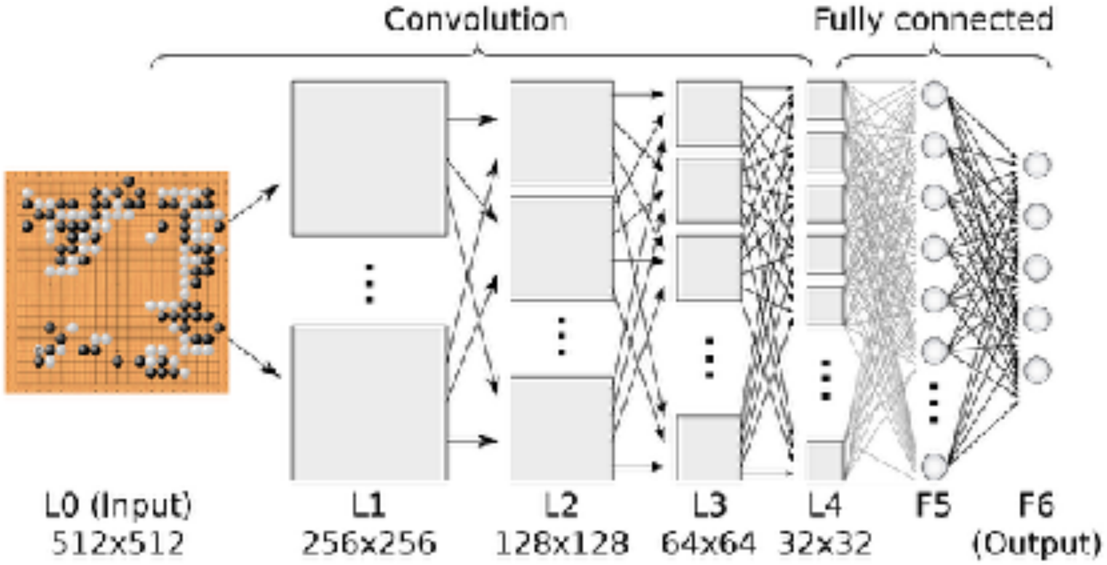
9 APRIL 2020

NATIVISM AND EMPIRICISM IN ARTIFICIAL INTELLIGENCE

LEARNING GO "TABULA RASA"







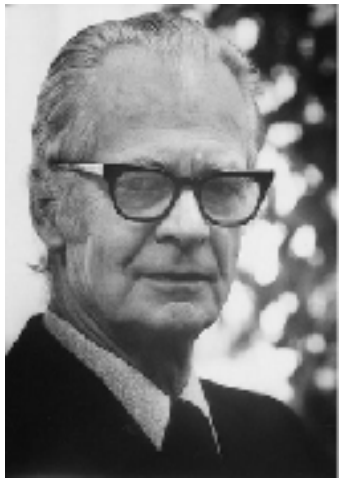
DEEPMIND: EMPIRICISM IN AI

- Silver et al. (2017): “a general-purpose reinforcement learning algorithm can achieve, **tabula rasa**, superhuman performance across many challenging domains”

EMPIRICISM VERSUS NATIVISM: HISTORY

Locke

Descartes



Skinner

Chomsky

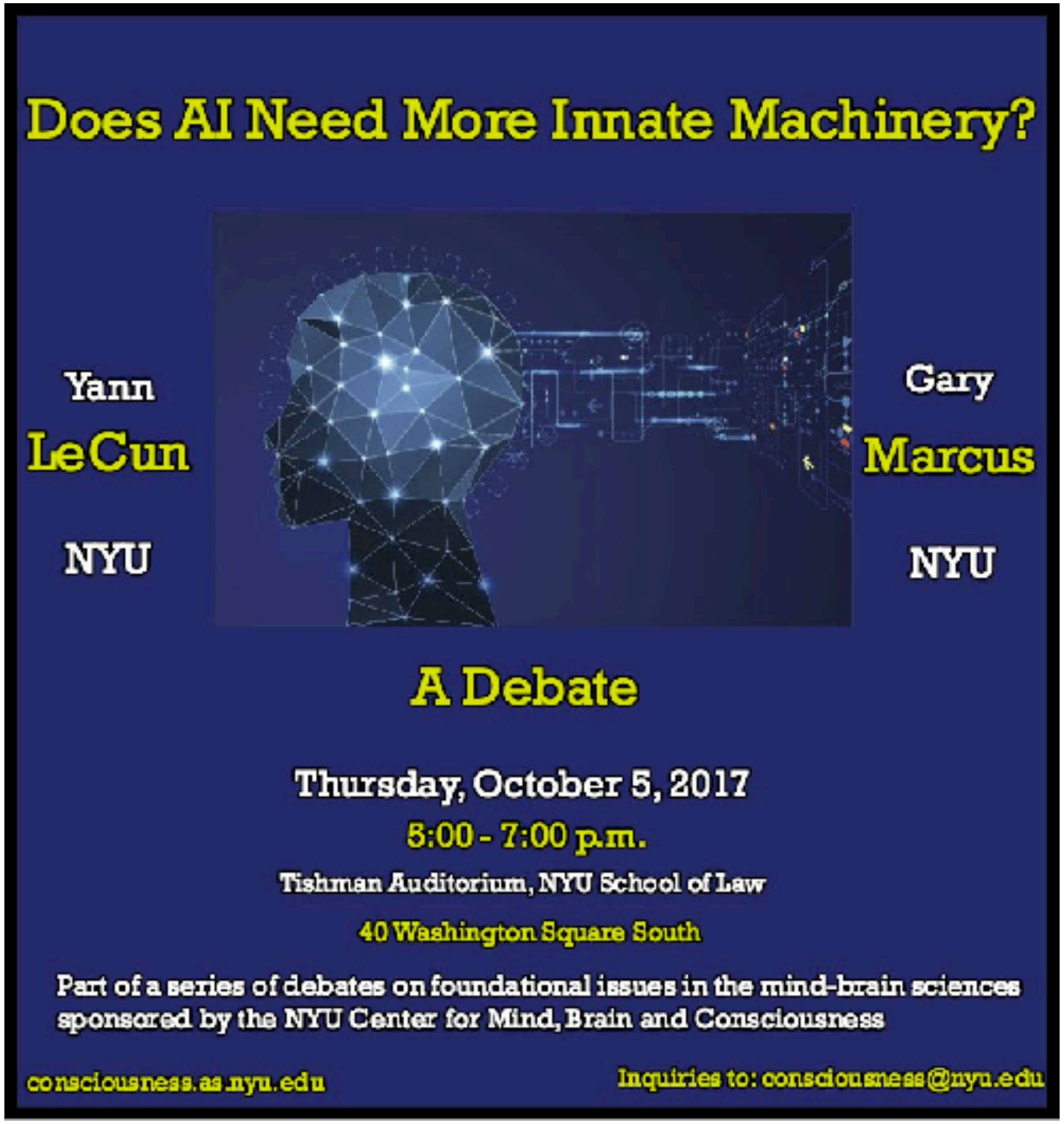
Piaget

Carey

NATIVISM VERSUS EMPIRICISM: ARTIFICIAL INTELLIGENCE

LeCun

Marcus



Does AI Need More Innate Machinery?

Yann LeCun NYU

Gary Marcus NYU

A Debate

Thursday, October 5, 2017

5:00 - 7:00 p.m.

Tishman Auditorium, NYU School of Law

40 Washington Square South

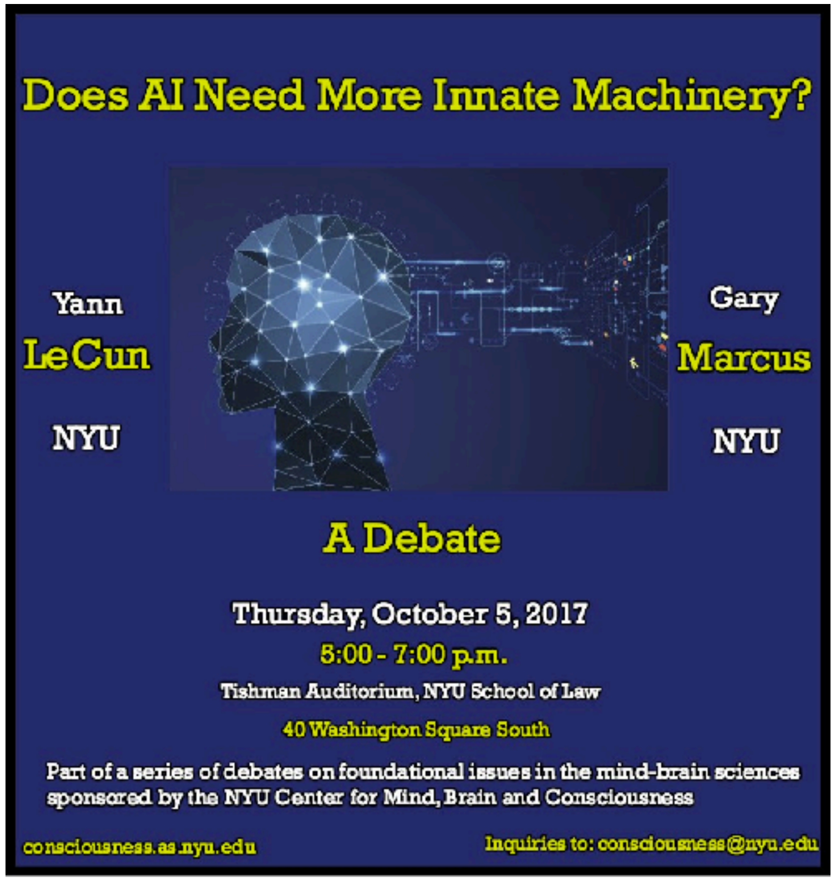
Part of a series of debates on foundational issues in the mind-brain sciences sponsored by the NYU Center for Mind, Brain and Consciousness

consciousness.as.nyu.edu

Inquiries to: consciousness@nyu.edu

ISSUES

- To what extent is strongly empiricist AI *possible*?
- To what extent is strongly empiricist AI *a good idea*?



Does AI Need More Innate Machinery?

Yann
LeCun
NYU

Gary
Marcus
NYU

A Debate

Thursday, October 5, 2017
6:00 - 7:00 p.m.
Tishman Auditorium, NYU School of Law
40 Washington Square South

Part of a series of debates on foundational issues in the mind-brain sciences sponsored by the NYU Center for Mind, Brain and Consciousness

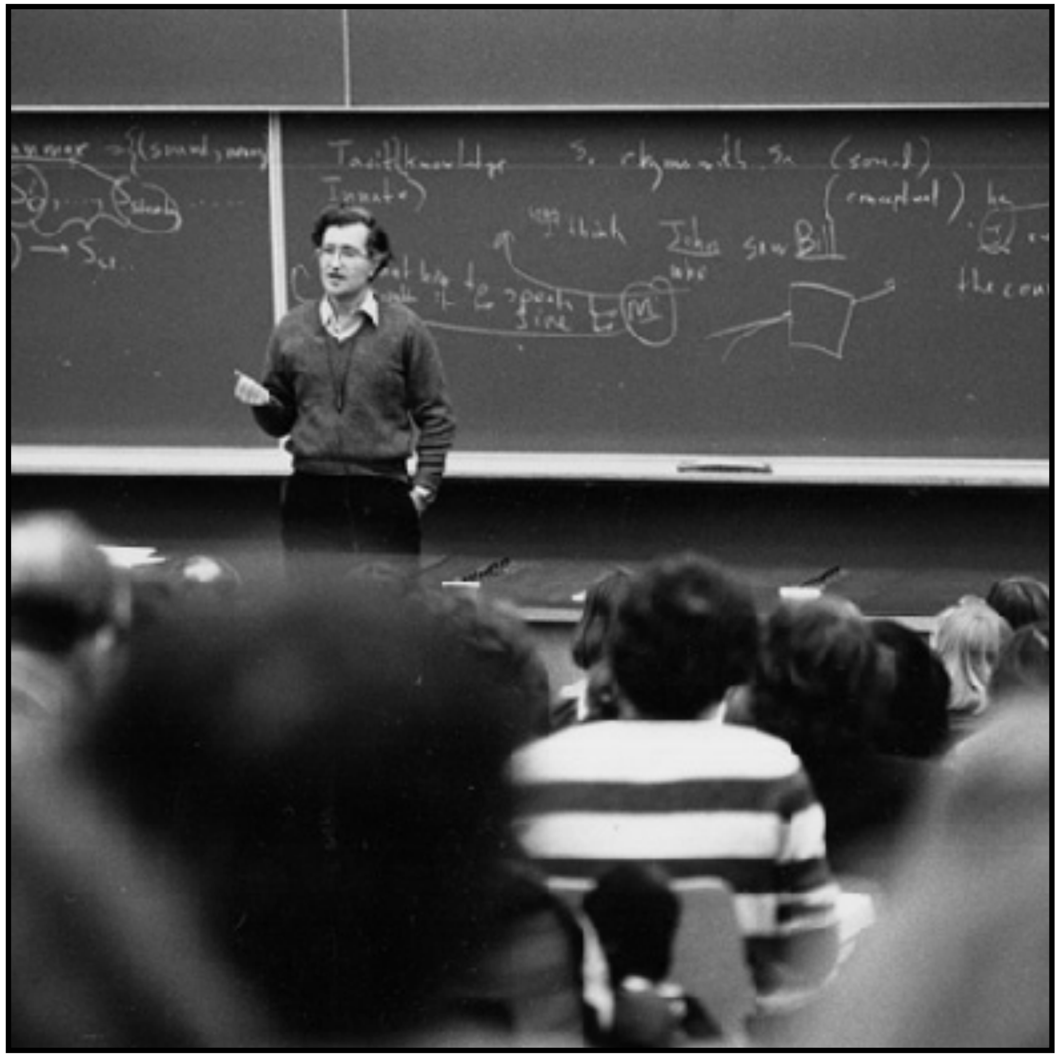
consciousness.as.nyu.edu
Inquiries to: consciousness@nyu.edu

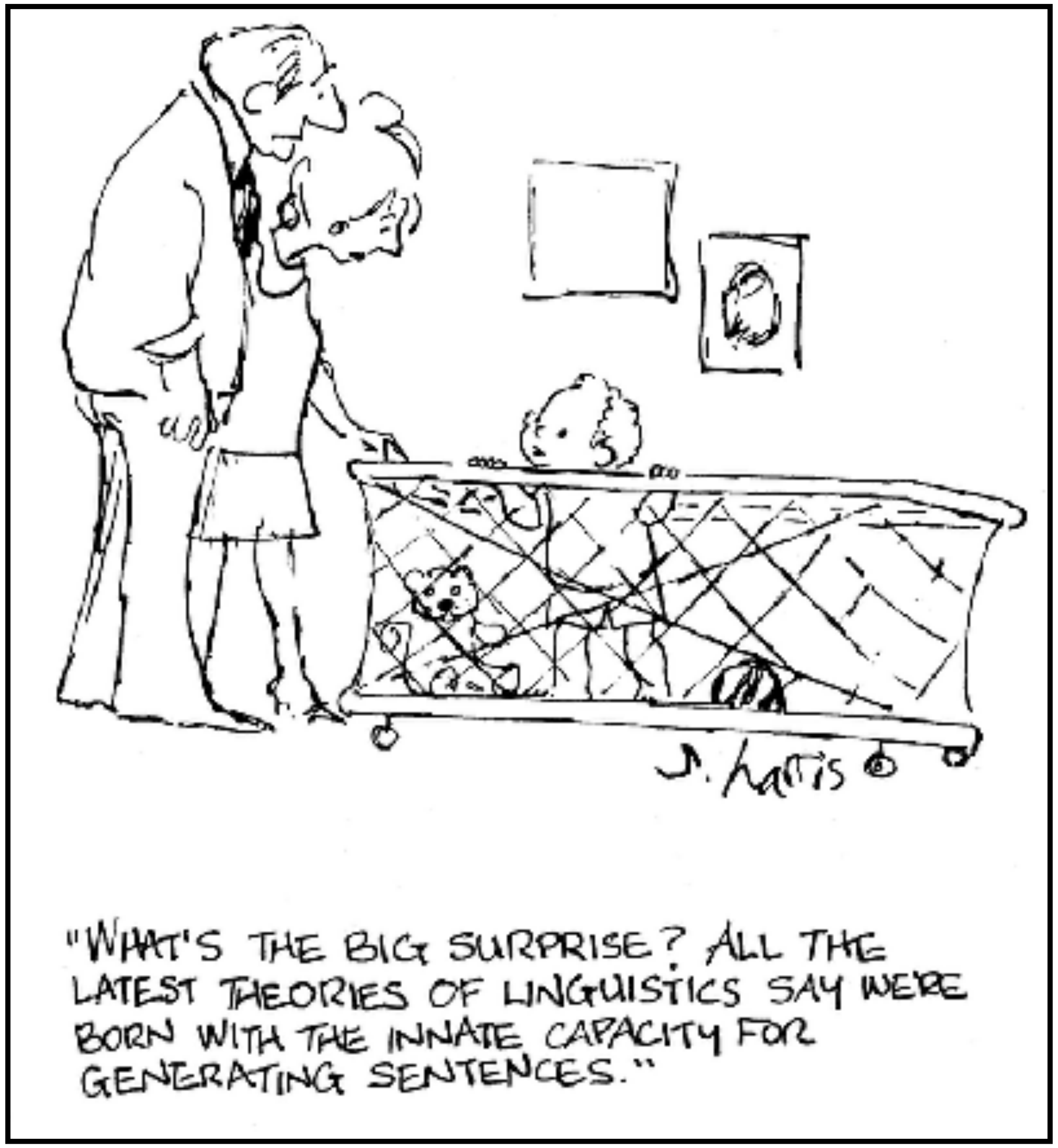
OUTLINE

- 1. Extend nativism versus empiricism to AI
- 2. Defend **possibility empiricism**: it is possible to build empiricist AI
- 3. Defend **ethical nativism**: we have ethical reason to build nativist AI

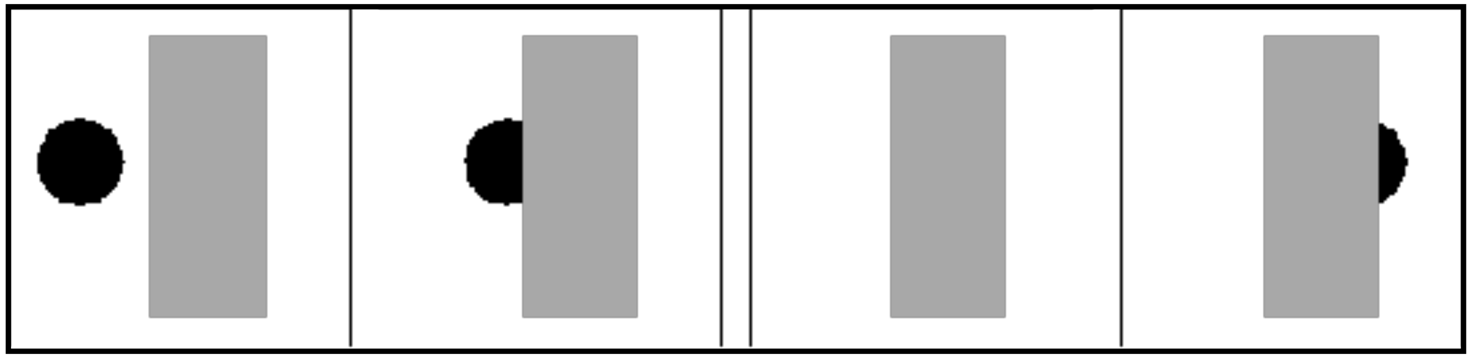
FRAMEWORK

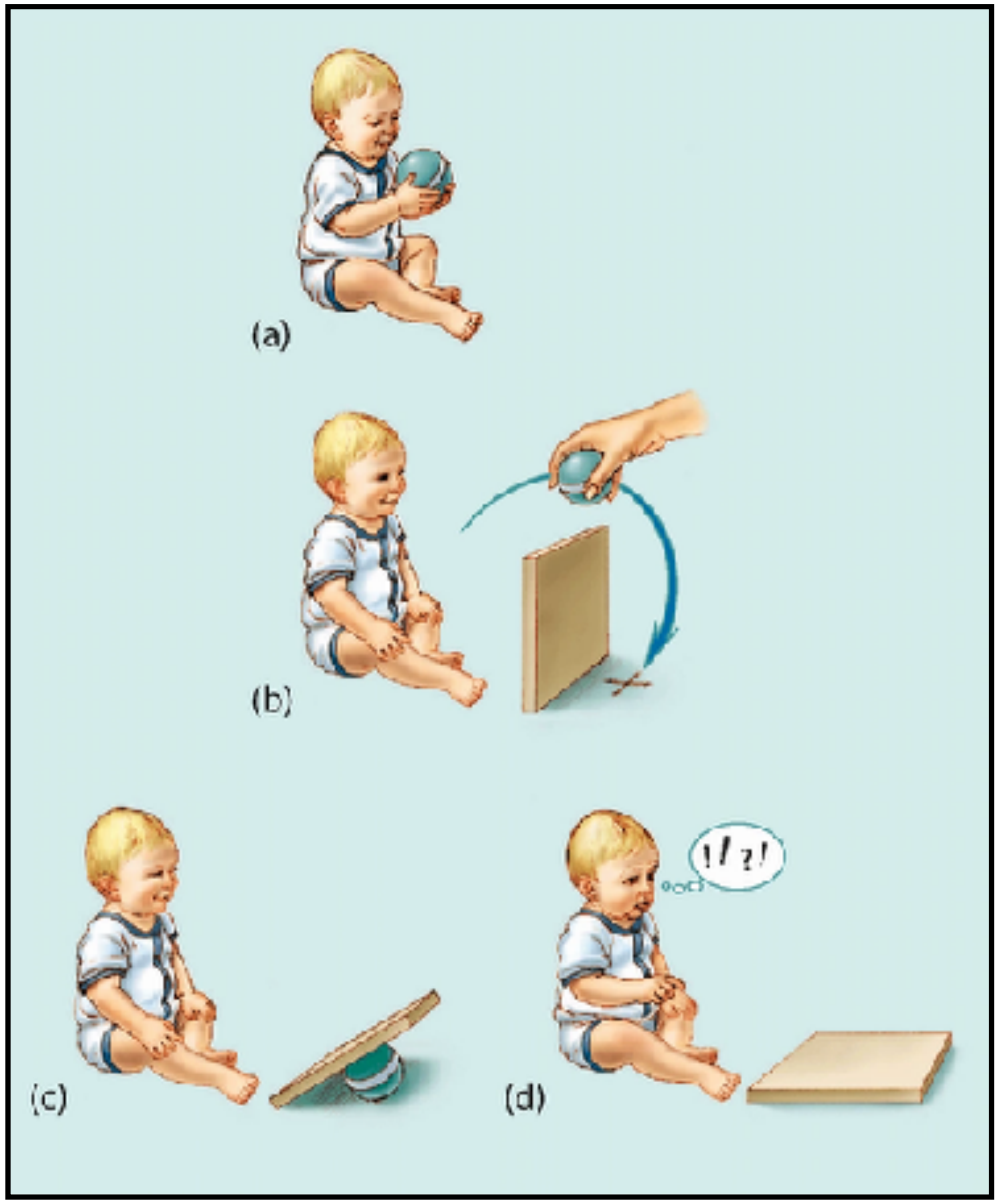
NATIVISM IN COGNITIVE SCIENCE: LANGUAGE



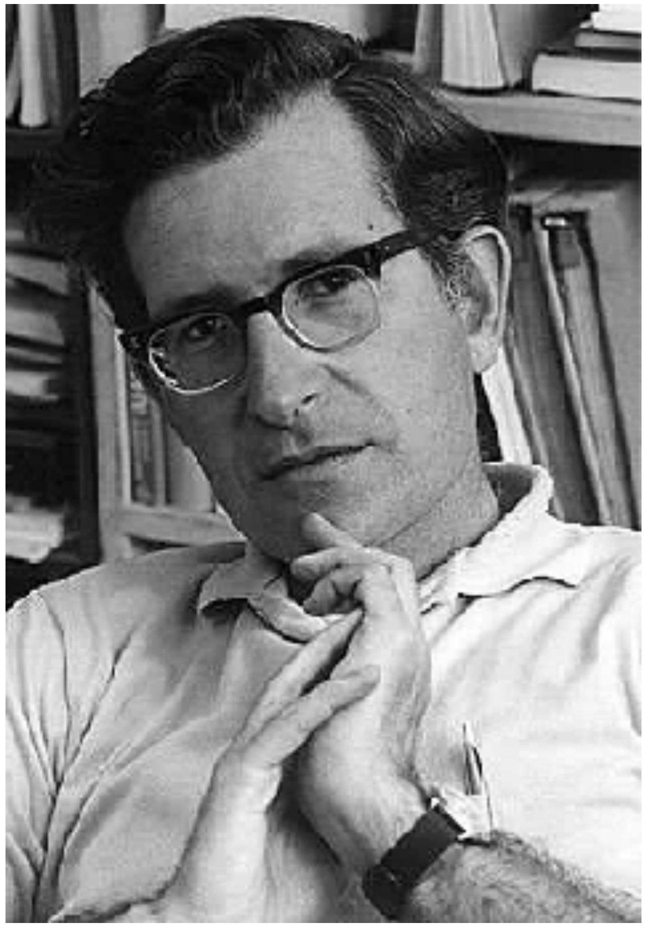

"WHAT'S THE BIG SURPRISE? ALL THE LATEST THEORIES OF LINGUISTICS SAY WE'RE BORN WITH THE INNATE CAPACITY FOR GENERATING SENTENCES."

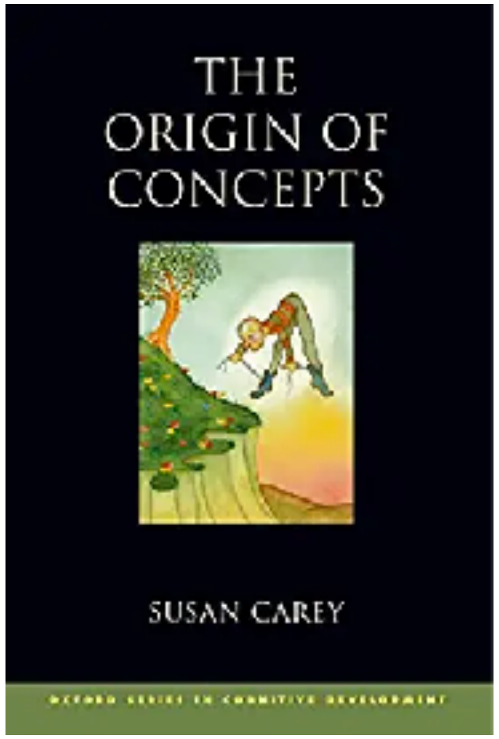
NATIVISM IN COGNITIVE SCIENCE: OBJECTS





NATIVISM IN COGNITIVE SCIENCE

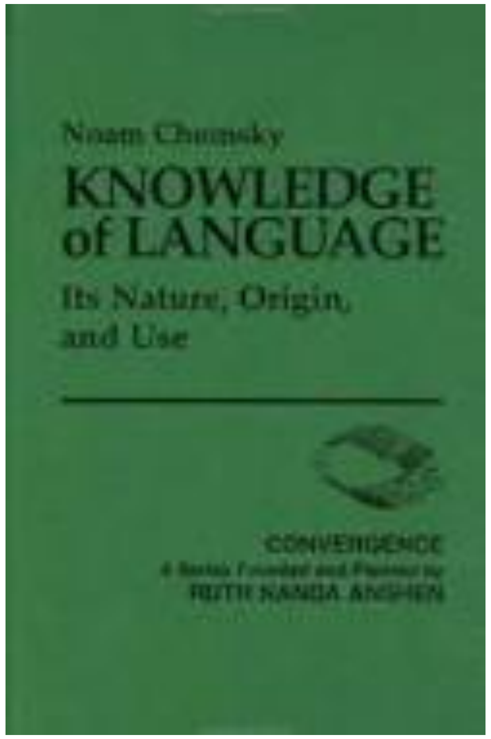




THE ORIGIN OF CONCEPTS

SUSAN CAREY

OXFORD SERIES IN COGNITIVE DEVELOPMENT



Noam Chomsky

KNOWLEDGE of LANGUAGE

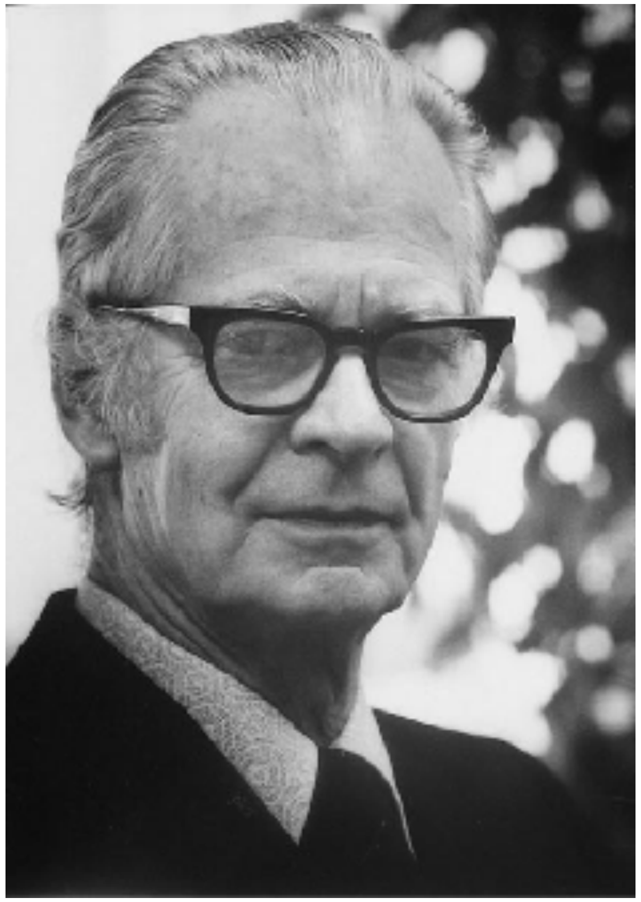
Its Nature, Origin, and Use

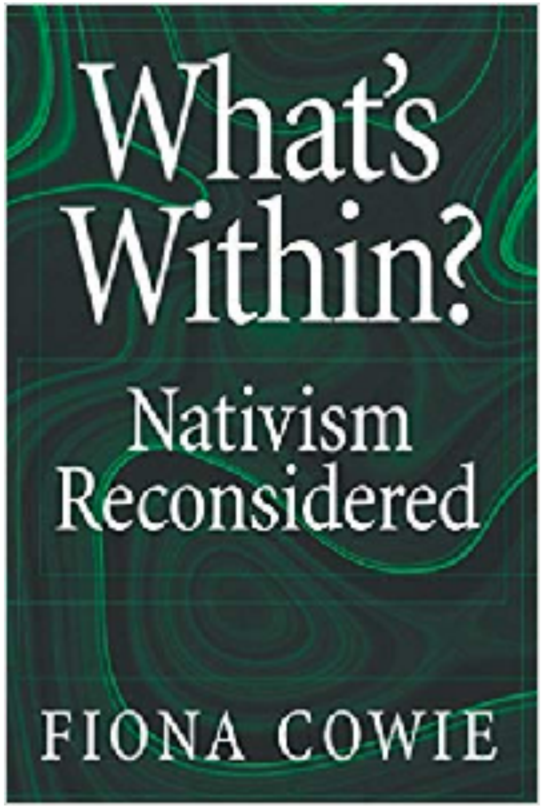
CONVERGENCE

A Series Founded and Planned by

RUTH NANDA ANSHEN

EMPIRICISM IN COGNITIVE SCIENCE

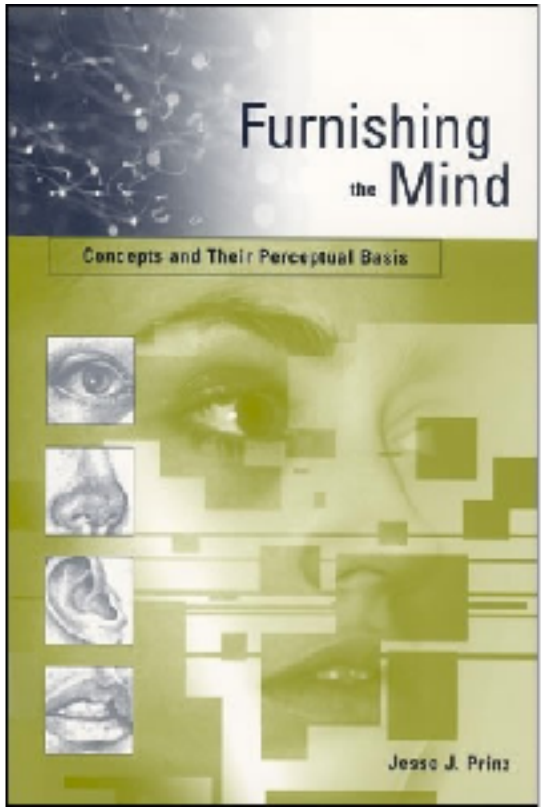




What's Within?

Nativism Reconsidered

FIONA COWIE



Furnishing the Mind

Concepts and Their Perceptual Basis

Jesse J. Prinz

WHAT IS THE DISAGREEMENT ABOUT?

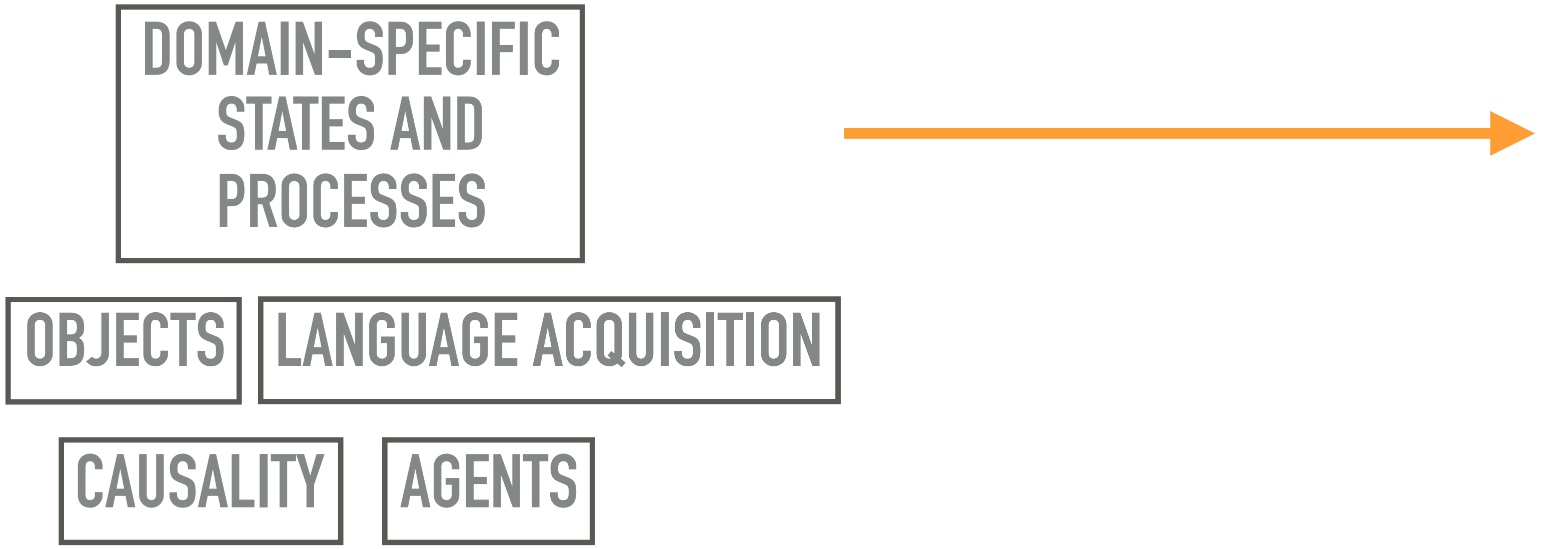
- *Something* has to be built in
  - Samet (1987): "Everyone agrees that learning requires that *something* be innate"
  - Quine (1969): "the behaviorist is knowingly and cheerfully up to his neck in innate mechanisms"

NATIVISM AND EMPIRICISM IN GENERAL

- **Nativist system:** a system whose initial state contains domain-specific mechanisms, states, and processes
- **Empiricist system:** a system whose initial state contains only domain-general mechanisms, states, and processes

HUMAN NATIVISM

- **Human nativism:** human beings are nativist systems, i.e. humans’ initial state contains domain-specific mechanisms, states, and processes

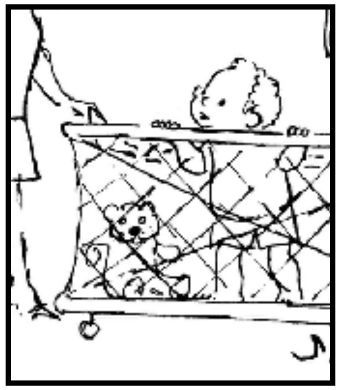

DOMAIN-SPECIFIC STATES AND PROCESSES

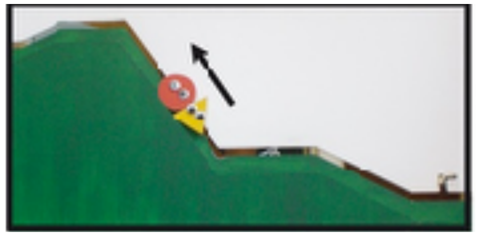
OBJECTS

LANGUAGE ACQUISITION

CAUSALITY

AGENTS

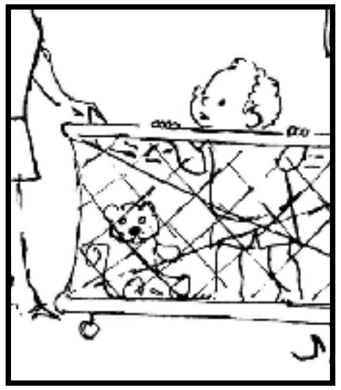


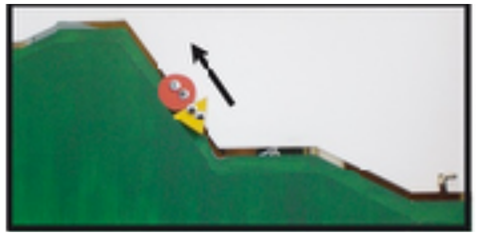


HUMAN EMPIRICISM

- **Human empiricism:** human beings are empiricist systems, i.e. humans’ initial state contains only domain-general mechanisms, states, and processes

DOMAIN-GENERAL STATES AND PROCESSES





NATIVISM AND EMPIRICISM IN GENERAL

- Human nativism and empiricism: about the character of **existing** systems
- Nativism and empiricism in AI: about the design of **possible** systems
  - what kinds of systems it is *possible* to build
  - what kinds of systems it is *practical* to build
  - what kinds of systems we *ought* to build

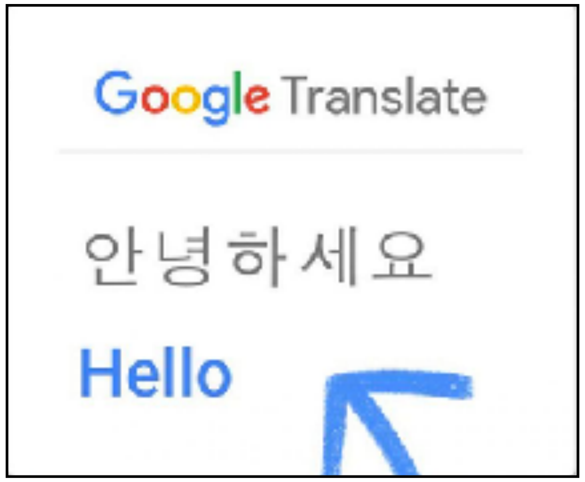
NATIVISM AND EMPIRICISM IN AI

- 1. Is empiricist AI *possible*?
  - Answer: yes. I'll argue for **possibility empiricism.**
- 2. What kinds of systems, ethically speaking, *ought* we to build?
  - Answer: we have ethical reasons to build nativist systems. I'll argue for **ethical nativism.**

POSSIBILITY EMPIRICISM

TASK AI

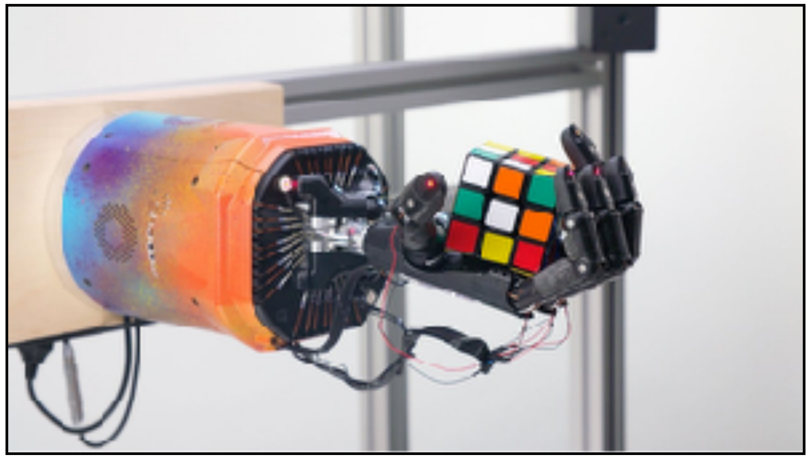
- Task AI: AI that achieves high levels of performance at a narrow range of tasks

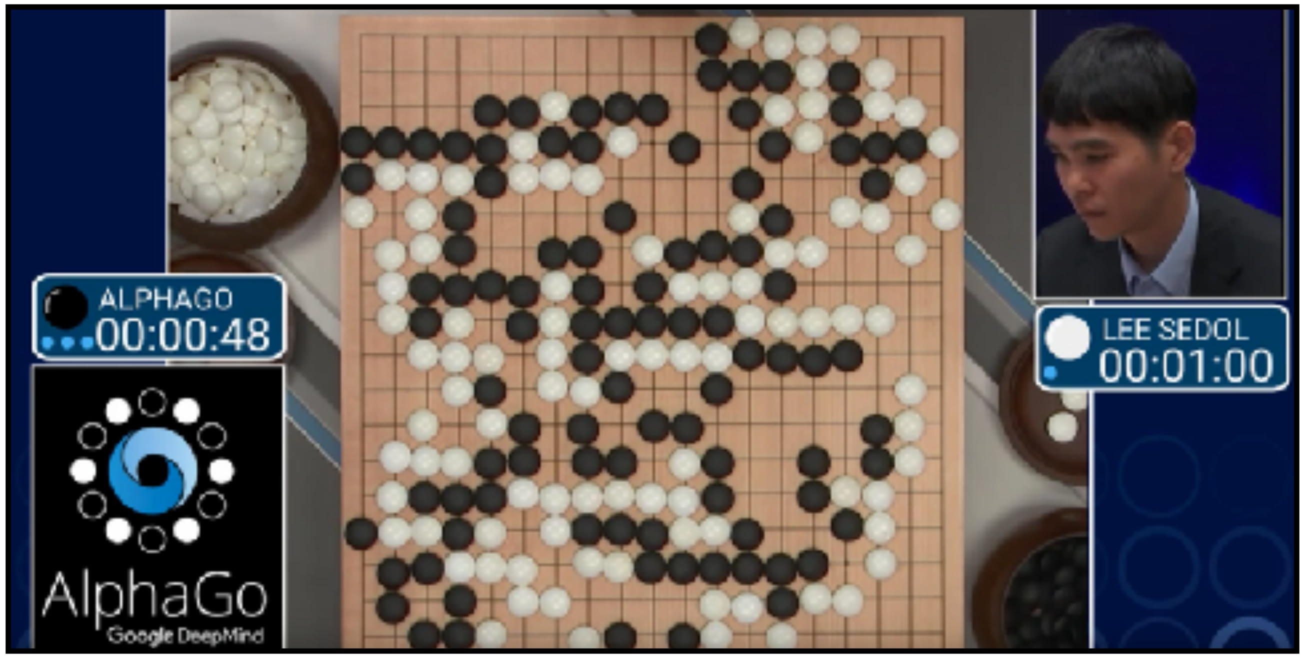


Google Translate

안녕하세요

Hello





ARTIFICIAL GENERAL INTELLIGENCE (AGI)

- AGI: achieves human-level performance on a wide variety of tasks, not just a few
- Ability to achieve goals in a *wide range* of environments
- Nativism vs. empiricism about AGI:
    - Will empiricist methods lead to continued progress?
    - Is empiricist AGI possible?

POSSIBILITY EMPIRICISM ABOUT AGI

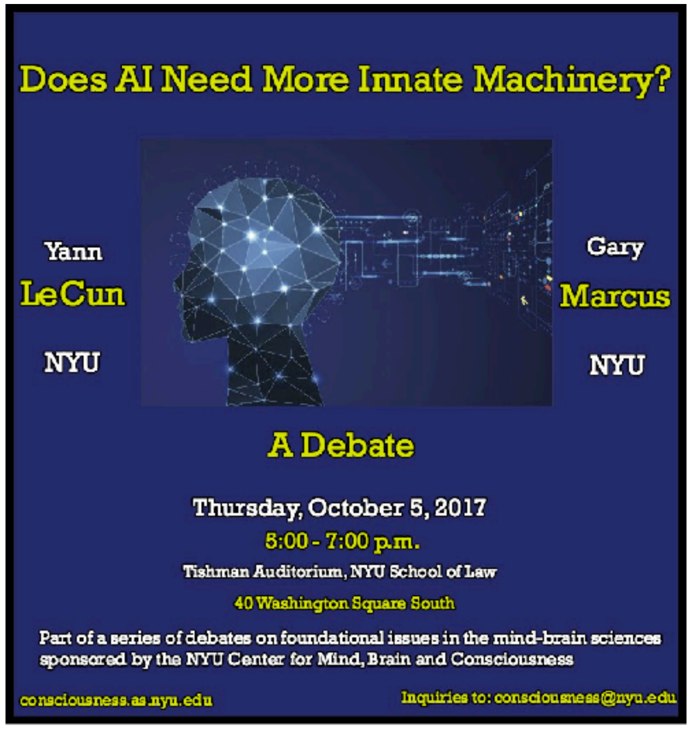
- **Possibility empiricism:** it is possible for an AGI to be an empiricist system, i.e. a system whose initial state contains only domain-general mechanisms, states, and processes
- **Necessity nativism:** necessarily, an AGI will be a nativist system, i.e. a system whose initial state contains domain-specific mechanisms, states, and processes

AI NATIVISM

- Marcus (2018): AI needs “the sort of things that strong nativists like myself, Noam Chomsky, Elizabeth Spelke, Steve Pinker and the late Jerry Fodor have envisioned.”

POSSIBILITY EMPIRICISM

- Yann LeCun (Facebook AI): none of the innate machinery proposed by Marcus is necessary for artificial general intelligence



Does AI Need More Innate Machinery?

Yann
LeCun
NYU

Gary
Marcus
NYU

A Debate

Thursday, October 5, 2017
5:00 - 7:00 p.m.
Tishman Auditorium, NYU School of Law
40 Washington Square South

Part of a series of debates on foundational issues in the mind-brain sciences sponsored by the NYU Center for Mind, Brain and Consciousness

consciousness.as.nyu.edu

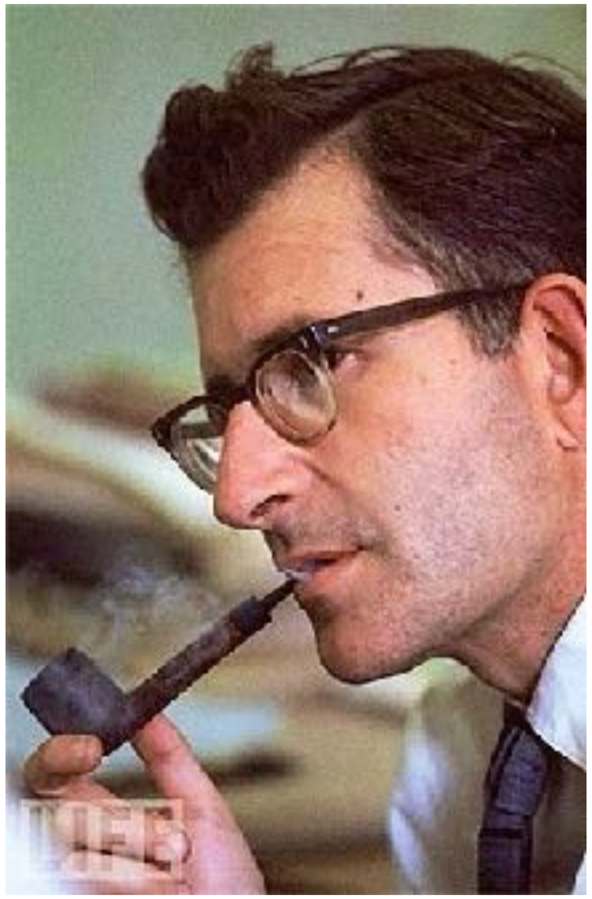
Inquiries to: consciousness@nyu.edu

AI EMPIRICISM

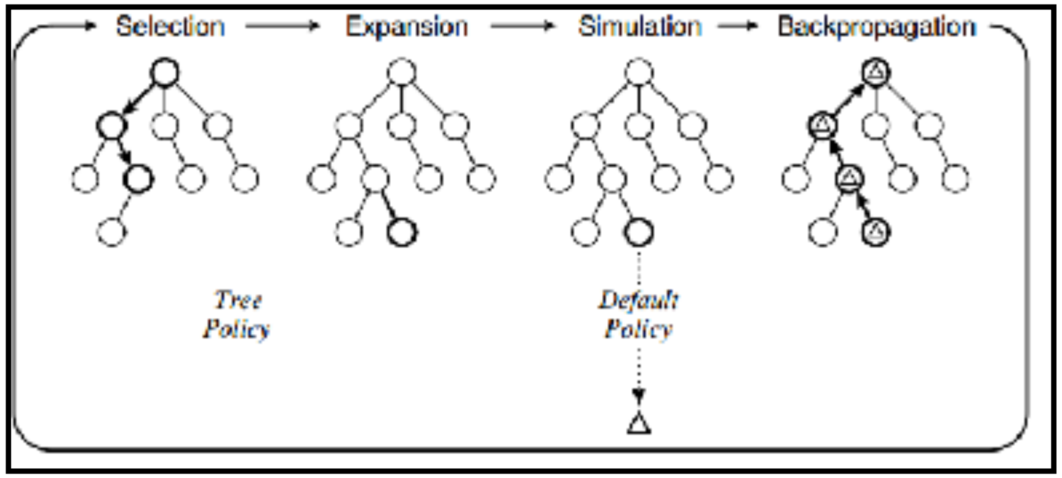
- Botvinick et al. (2017): DeepMind wants to leave “wide scope for learning to absorb domain-specific structure… **avoiding a dependence on detailed, domain-specific prior information**”

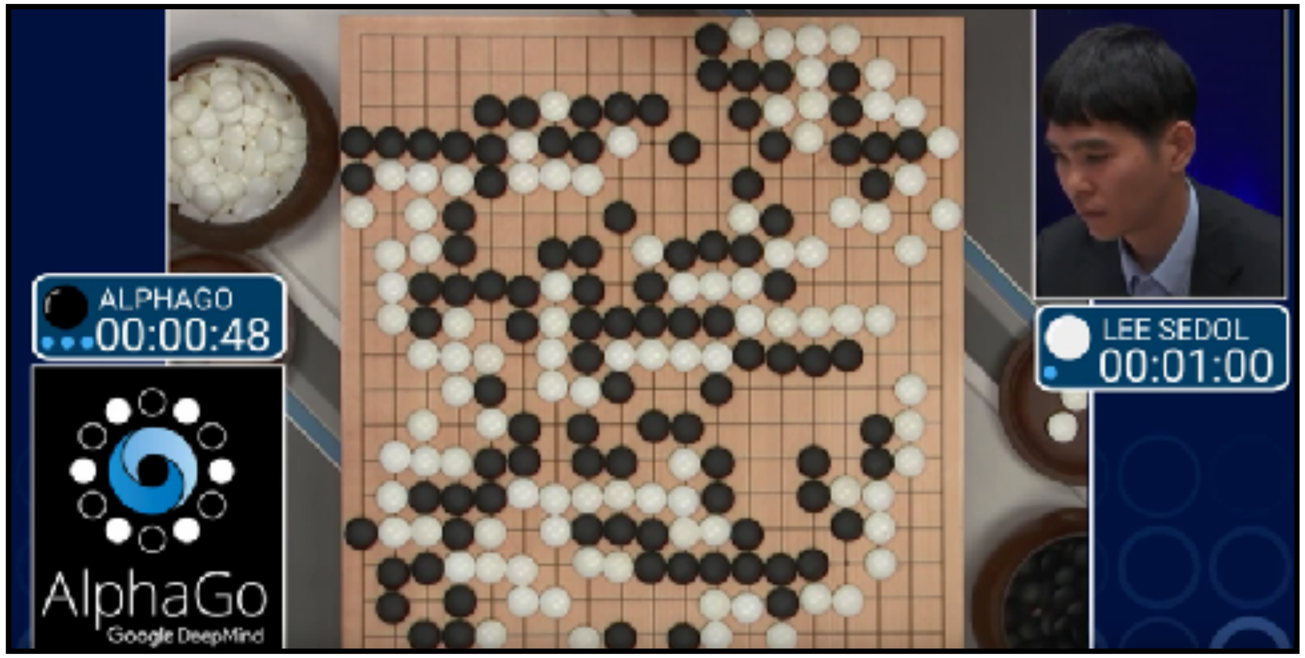
THE POVERTY OF THE STIMULUS

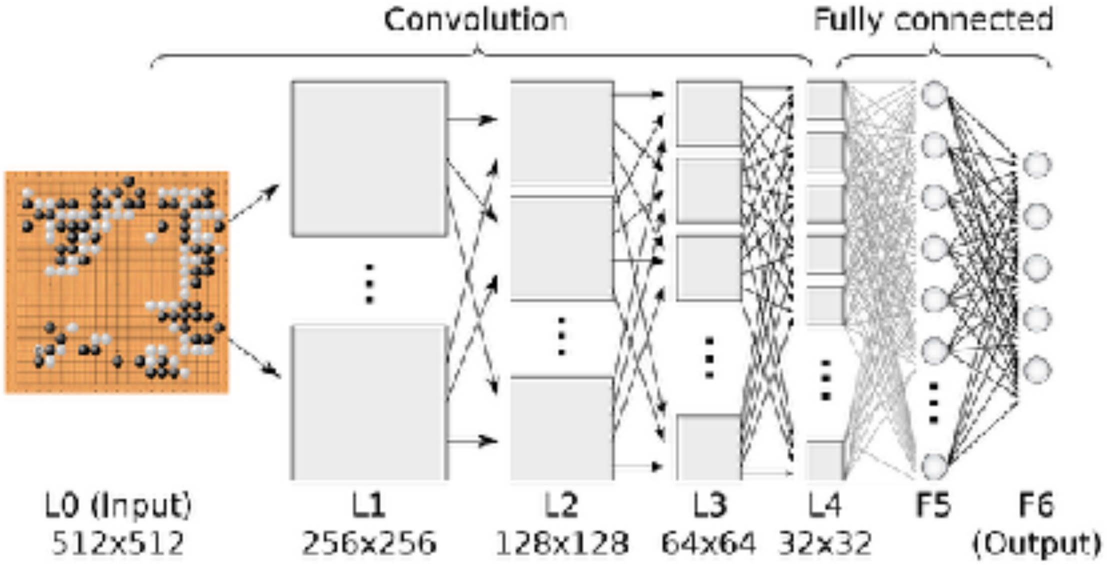
- 1. If children were empiricist learners, the data available to them would be too impoverished - they would not reliably arrive at the correct grammar for their language.
- 2. Children do reliably arrive at the correct grammar for their language.
- 3. Children are not empiricist learners.



BIG DATA EMPIRICISM: PROSPERITY OF THE STIMULUS







IMPOSSIBILITY ARGUMENTS FOR NECESSITY NATIVISM

- 1. Capacity C is required for AGI
- 2. Capacity C cannot be learned from data using domain-general empiricist learning mechanisms, and instead requires innate machinery N
- 3. Therefore, AGI requires innate machinery N

REQUIREMENTS

- Marcus (2018): "**many different types of tasks will have their own innate requirements**: …syntactic tree manipulation operations for language understanding, geometric primitives for 3-D scene understanding, theory of mind for problems demanding social coalitions, and so forth."

LEARNABILITY

- 1. Capacity C is required for AGI
- **2. Capacity C cannot be learned from data using domain-general empiricist learning mechanisms, and instead requires innate machinery N**
- 3. Therefore, AGI requires innate machinery N

LEARNABILITY

- Spelke (1996): "It is far from clear how children could learn anything about the entities in a domain, however, if they could not single out those entities."

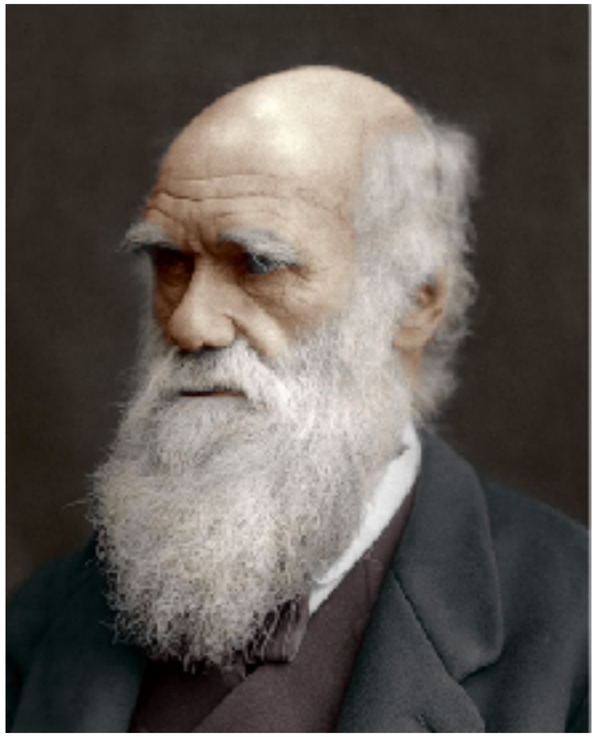
- Carey (2009): "Learnability considerations also argue that the representations in core cognition are the output of innate input analyzers."

TROUBLES FOR LEARNABILITY ARGUMENTS

- Learnability arguments to date bear on what *children* could learn, but do not bear on what AI systems can learn.
    - Carey: "There is no proposal I know for a learning mechanism *available to non-linguistic creatures*…"
- So these arguments are inconclusive as arguments for necessity nativism

EVOLUTIONARY ARGUMENTS FOR POSSIBILITY EMPIRICISM

- Samet and Zaitchik (2012): "the range of '**learning from experience**', the Empiricist's core commitment, would simply be **extended** to cover…species-based learning as well."



- Turing (1950): "We cannot expect to find a good child-machine at the first attempt…There is an obvious connection between this process and evolution."

EVOLUTIONARY EMPIRICISM: EVOLUTION AS LEARNING

1. Evolution plus learning leads from a domain-general initial state to general intelligence.

2. Evolution can be recapitulated as learning.

3. If (1) and (2), then learning from a domain-general starting point can achieve general intelligence.

4. Learning from a domain-general starting point can achieve general intelligence.

CHEAP LEARNING?

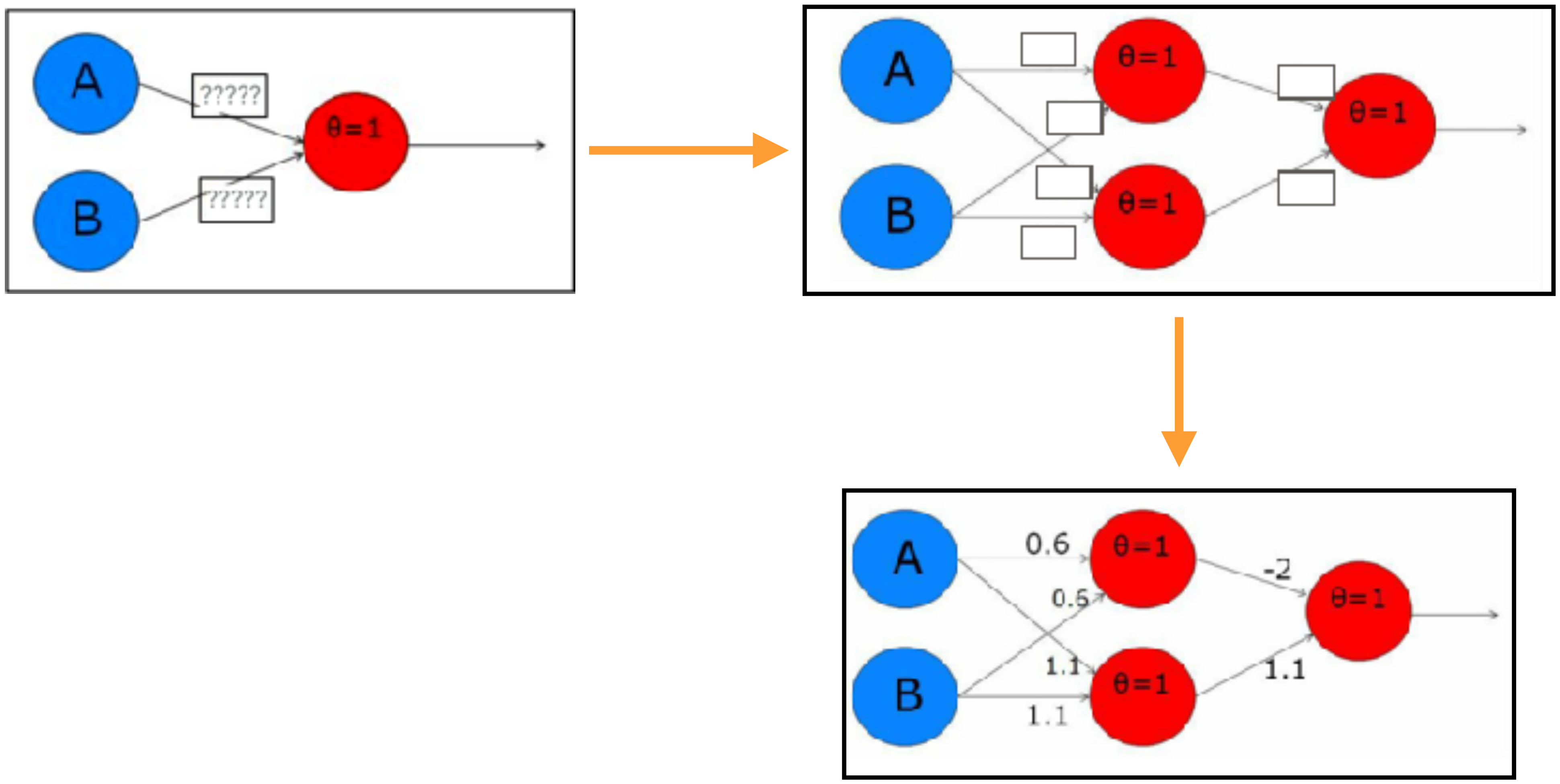
- Marcus (2018): “One might as well just use the term learning to refer to all change over time…and count rock formations as the product of learning, too.”
- Not learning by a single ‘system’

CHALLENGES

- Objection to premise 2: evolution cannot be recapitulated as learning
  - Learning is a *rational process*. Evolution is not. Evolution is (in part) architecture search, operating over multiple systems.
- Response: architecture search is a kind of learning. It is a generate-and-test process responsive to evidence about which systems succeed.

TWO SENSES OF LEARNING

- **Narrow learning:** learning given a certain architecture
- **Broad learning**: learning which includes architecture search



ARCHITECTURE SEARCH AND POSSIBILITY EMPIRICISM

- My view: architecture search can be a rational process–a generate and test process responsive to evidence about which systems succeed. So broad learning counts as learning.
- This yields **architecture-search empiricism**: AGI can be achieved from a domain-general initial state through learning that includes architecture-search
- Open question: can AGI be achieved through learning from a tabula rasa through learning without architecture search?

ETHICAL NATIVISM

WHAT SORT OF SYSTEMS SHOULD WE BUILD?

- Two different questions:
  - What is the most efficient or feasible way forward?
  - What do we have ethical reason to do?

ETHICAL NATIVISM AND ETHICAL EMPIRICISM

- **Ethical nativism:** We have (pro tanto) ethical reasons to build nativist AI systems.
- I’ll consider ethical nativism both for current AI systems and for AGI

ETHICAL NATIVISM

- 1) We have (pro tanto) ethical reasons to build AI systems that are fair.
- 2) Fairness requires explainability.
- 3) Nativist AI systems are more conducive to explainability.
- **Ethical nativism:** We have (pro tanto) ethical reasons to build nativist AI systems.

WHAT IS EXPLAINABILITY?

- Three common, and separable problems:
  - Proprietary systems
  - Complex systems
  - **Unexplainable systems**

WHAT IS EXPLAINABILITY?

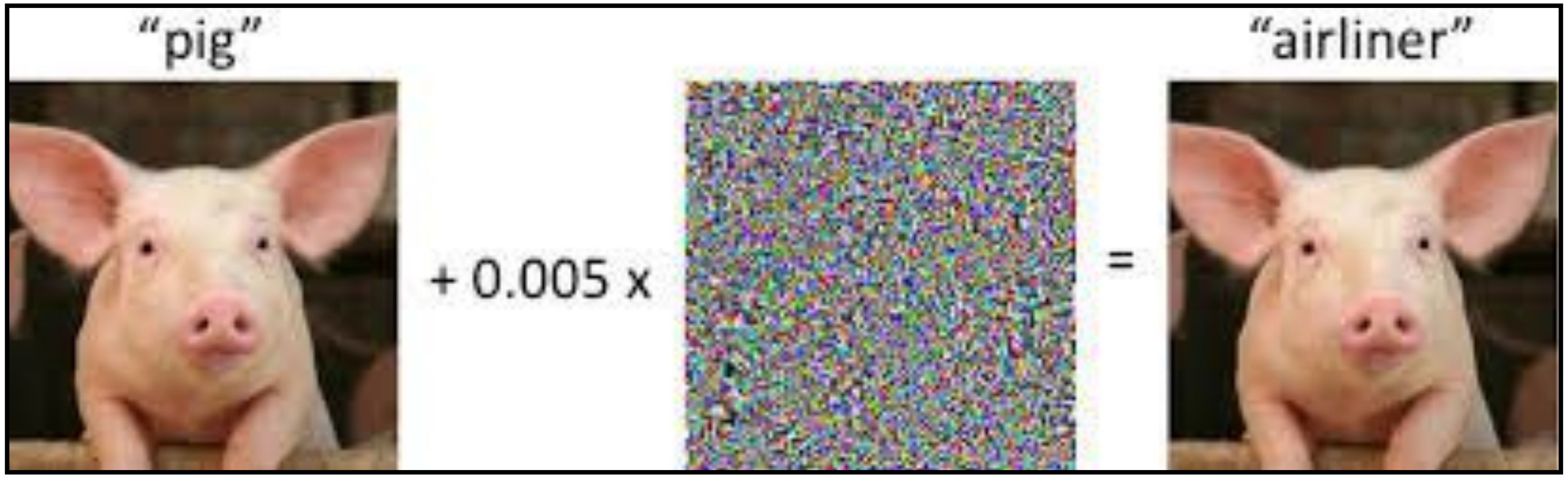
- An AI system is **explainable** to the extent that it is possible to give humanly comprehensible **explanations** of its predictions and actions (ideally, reasons).

Vredenburgh

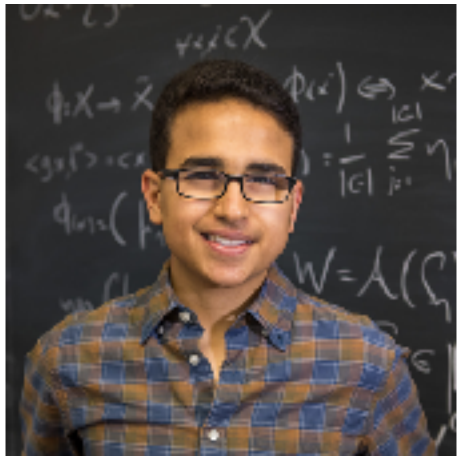
Creel

Lipton

DEEP NEURAL NETWORKS OFTEN LACK EXPLAINABILITY



- Ilyas et al. (2019): "adversarial examples can be directly attributed to the presence of non-robust features: features (derived from patterns in the data distribution) that are **highly predictive, yet brittle and (thus) incomprehensible to humans**."

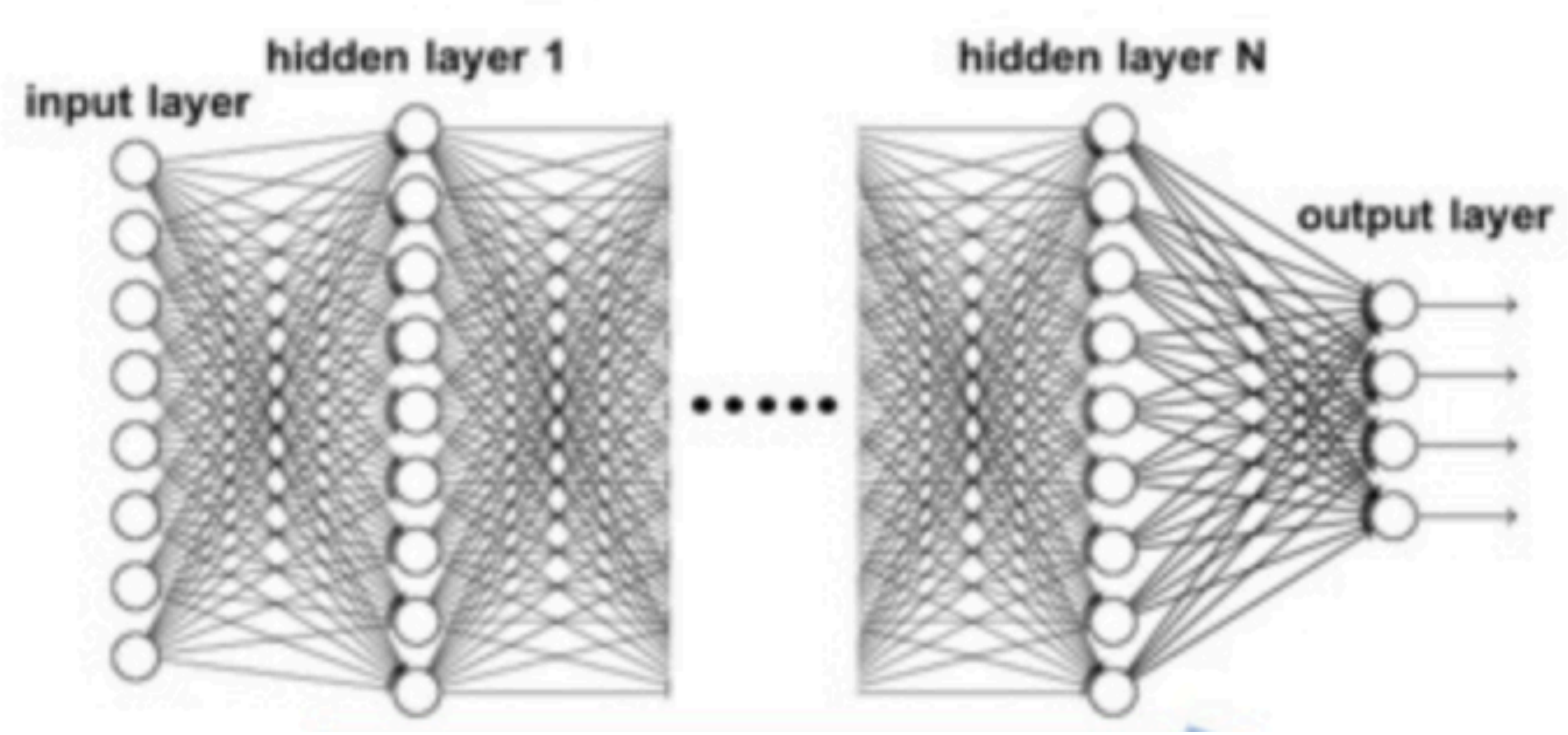


Ilyas

Madry

EXPLAINABILITY AND FAIRNESS

- Explaining consequential decisions and predictions



THE IMPORTANCE OF EXPLAINABILITY

- Constitutive:
  - Explanations are part of treating people with respect
- Epistemic:
  - Enables confidence that a system is non-discriminatory
- Practical:
  - Allows for deliberation, appeal, and protest
  - Allows for predictability by affected individuals
  - Vredenburgh (ms): *informed self-advocacy*

Vredenburgh

EXPLAINABLE SYSTEMS

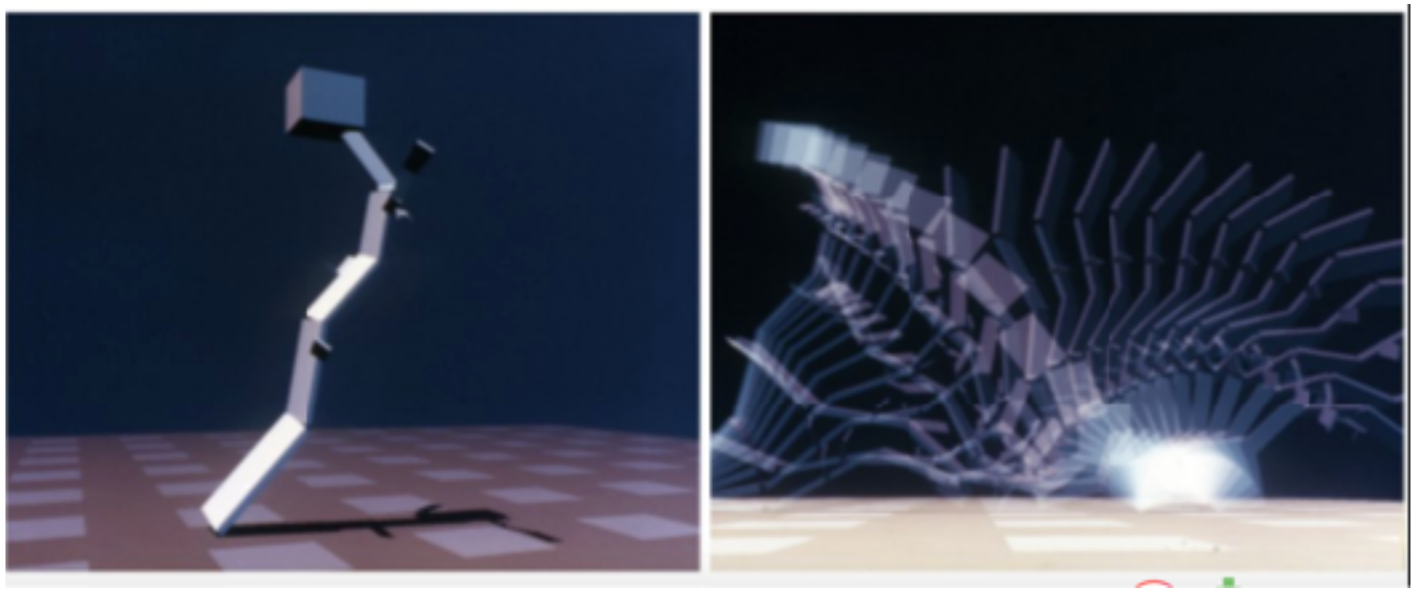
- Locke (1690) and rule of law liberalism: importance of "established standing Laws, promulgated and known to the People"
- Vredenburgh (ms): "bureaucracies can be opaque in much the same way that algorithms can"

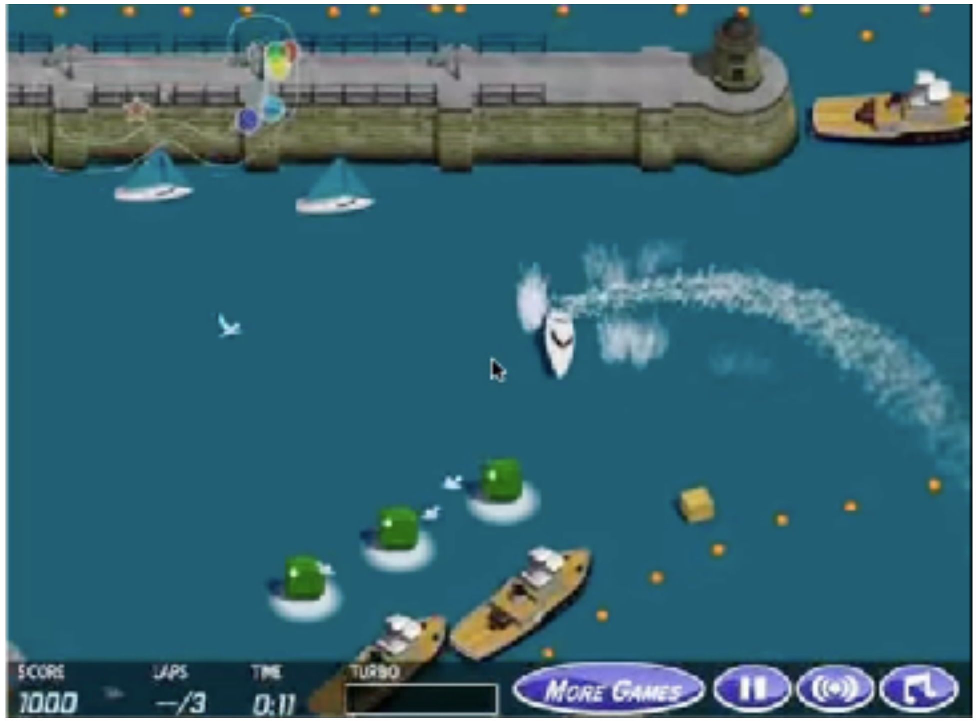
COMPLEXITY

- The issue is not complexity per se.
  - Simple systems can be unexplainable
    - Decision tree employing alien features
  - Complex systems can be explainable
    - Economic systems
    - Humans beings

EXPLAINABILITY AND NATIVISM: CONSTRAINTS

- The problem is not complexity, but alien-ness
- A wide solution space is part of the power, but also the inexplicability, of empiricist methods.
- Building in domain-specific starting points is a crucial way of constraining the solution space and finding *explainable* solutions.

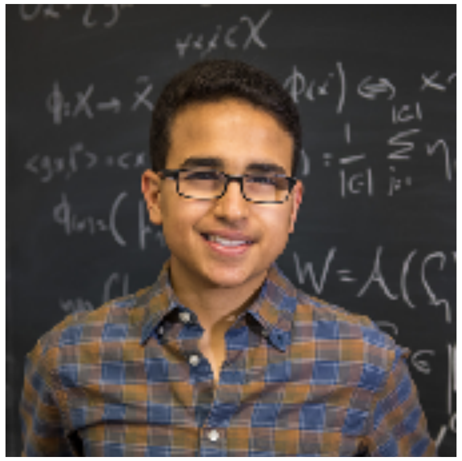




EXPLAINABILITY AND NATIVISM: CONCEPTS

- Nativist systems are more apt to utilize human-like representations.
- Shared representations allow for a crucial kind of explainability.
- This gives us reason to build nativist systems.

- Ilyas et al. (2019): “attaining models that are robust and interpretable will require **explicitly encoding human priors** into the training process.”



Ilyas

Madry

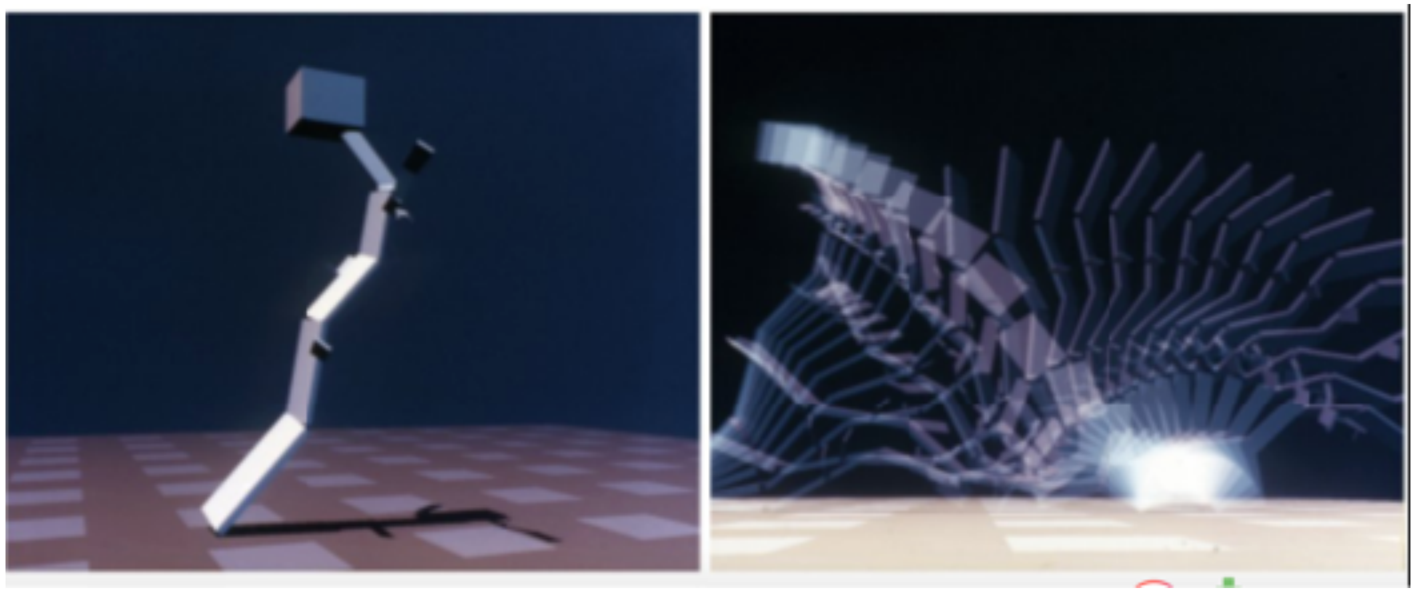
ARGUMENT FROM EXPLAINABILITY AND SAFETY

- 1) We have (pro tanto) ethical reasons to build AI systems that are safe.
- 2) Safety requires explainability.
- 3) Nativist AI systems are more conducive to explainability.
- **Ethical nativism:** We have (pro tanto) ethical reasons to build nativist AI systems.

THE IMPORTANCE OF EXPLAINABILITY FOR SAFETY

- Caveat to (2): formal guarantees of behavior might provide safety without explainability ("I don't understand the behavior of this system, but I know it won't do X")

- But failing formal guarantees, shared-representation explainability is important for predicting possible behavior.



PREDICTABILITY AND SAFETY

- Locke (1690): without public and explainable laws, citizens are subject to others' "sudden thoughts, or unrestrained and till that moment unknown Wills, without having any measures set down which may guide and justify their actions"

CONCLUSION

TAKEAWAYS

- 1. Extended nativism versus empiricism to AI
- 2. Defended possibility empiricism: it is possible to build empiricist AI
- 3. Defended ethical nativism: we have ethical reason to build nativist AI

THANK YOU!

- email: rl2898@nyu.edu
- website: robertlong.online.com