Bridging AI & Cognitive Science
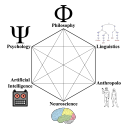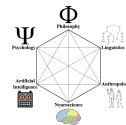
Jessica Hamrick, Aida Nematzadeh, Kaylee Burns, Alison Gopnik, Josh Tenenbaum & Emmanuel Dupoux

CENTER FOR
Brains
Minds+
Machines

DeepMind

facebook

Thanks to our sponsors, the ICLR organizers, SlidesLive, and to everyone who submitted, reviewed, and prepared content!
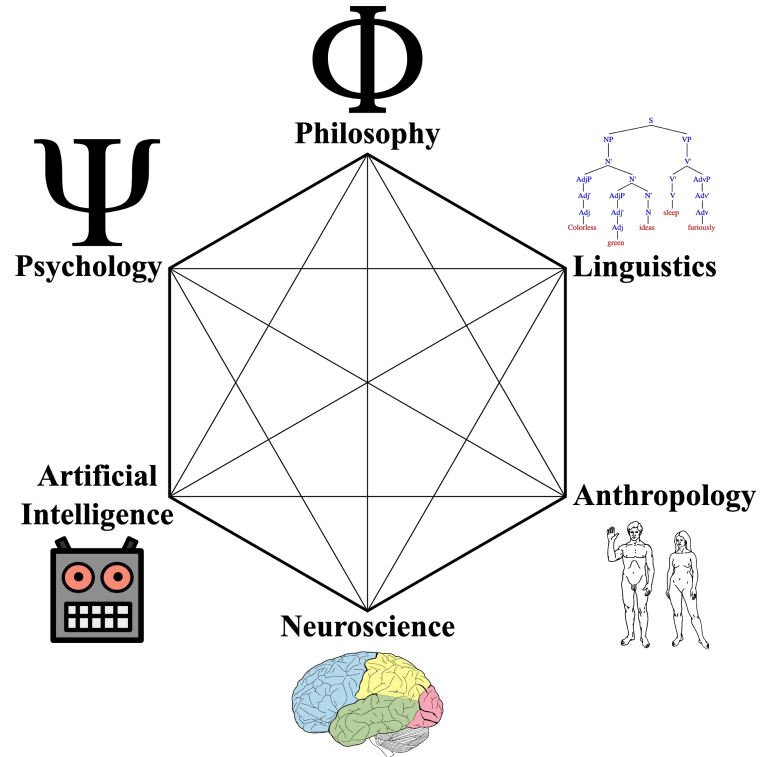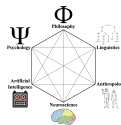
The 5 Minute History of AI and Cognitive Science
Jessica Hamrick
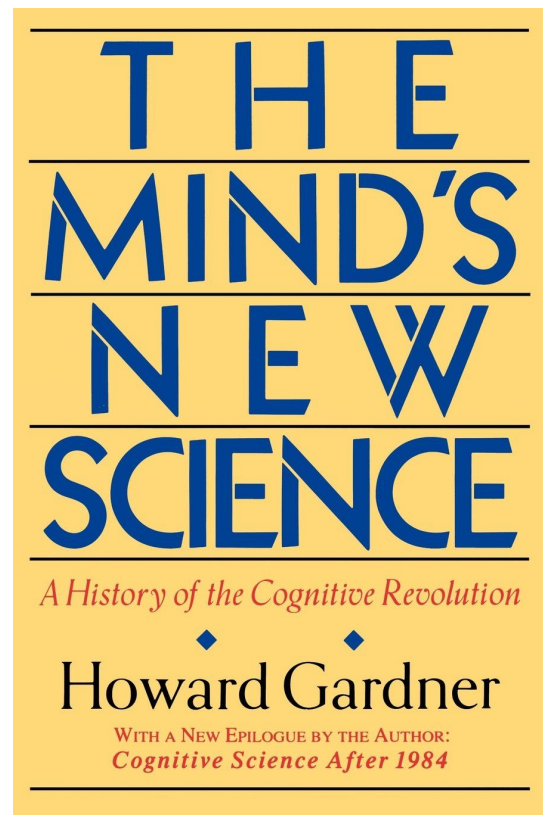
What is Cognitive Science?

*The **study of intelligent systems** and how they produce behavior, rooted in the assumption that those systems follow **principles of computation**.*

1956: The Birth of AI and Cognitive Science

- Pre-1956: Lots of new ideas and inspiration
  - Turing, von Neumann, McCulloch, Pitts, Shannon, Tolman, Bartlett, Craik, Brunswick, etc.

- Summer 1956: Dartmouth Summer Research Project on AI
  - Considered to be the founding of **AI**

- September 1956: IEEE Symposium on Information Theory
  - Considered to be the founding of **cognitive science**
  - Many of the same participants as at Dartmouth

THE MIND'S NEW SCIENCE

A History of the Cognitive Revolution

Howard Gardner

WITH A NEW EPILOGUE BY THE AUTHOR:
Cognitive Science After 1984

COGNITIVE SCIENCE

Neural computation + signal processing

Efficient Coding

Behaviorism

Thinking as information processing

Logic

1956
Dartmouth
IEEE

"Thinking Machines"

Information theory

AI + ML

Pre-1950 | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s

COGNITIVE SCIENCE

Neural computation + signal processing

Efficient Coding

Mental models

Cognitive architectures

Behaviorism

Thinking as information processing

Language of Thought

Logic

1956
Dartmouth
IEEE

"Thinking Machines"

Search

LISP

Physical Symbol Systems

Expert Systems

Information theory

AI + ML

| Pre-1950 | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s |

COGNITIVE SCIENCE

Neural computation + signal processing

Efficient Coding

Mental models

Bounded rationality

Cognitive architectures

PDP + Connectionism

Behaviorism

Thinking as information processing

Language of Thought

Logic

1956 Dartmouth IEEE

"Thinking Machines"

Search

LISP

Physical Symbol Systems

Expert Systems

Reinforcement learning

Shallow neural networks

Information theory

AI + ML

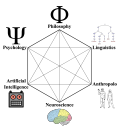| Pre-1950 | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s |

Two Conferences in 1987...

Conference #1

- Centric Models of the Orientation Map in Primary Visual Cortex
- Simulations Suggest Information Processing Roles for the Diverse Currents in Hippocampal Neurons
- Optimal Neural Spike Classification
- Neural Networks for Template Matching: Application to Real-Time Classification of the Action Potentials of Real Neurons
- A Computer Simulation of Olfactory Cortex with Functional Implications for Storage and Retrieval of Olfactory Information
- Schema for Motor Control Utilizing a Network Model of the Cerebellum
- A Computer Simulation of Cerebral Neocortex
- Discovering Structure from Motion in Monkey, Man and Machine

Conference #2

- A Connectionist Network that Learns Natural Language Grammar
- A Connectionist Architecture for Representing and Reasoning about Structured Knowledge
- A Connectionist Encoding of Semantic Networks
- A Dual Back-Propagation Scheme for Scalar Reward Learning
- Using Fast Weights to Deblur Old Memories
- On the Connectionist Reduction of Conscious Rule Interpretation
- Cascaded Back-Propagation on Dynamic Connectionist Networks
- A Neural Network for the Detection and Representation of Oriented Edges
- Learning Internal Representations from Gray-Scale Images
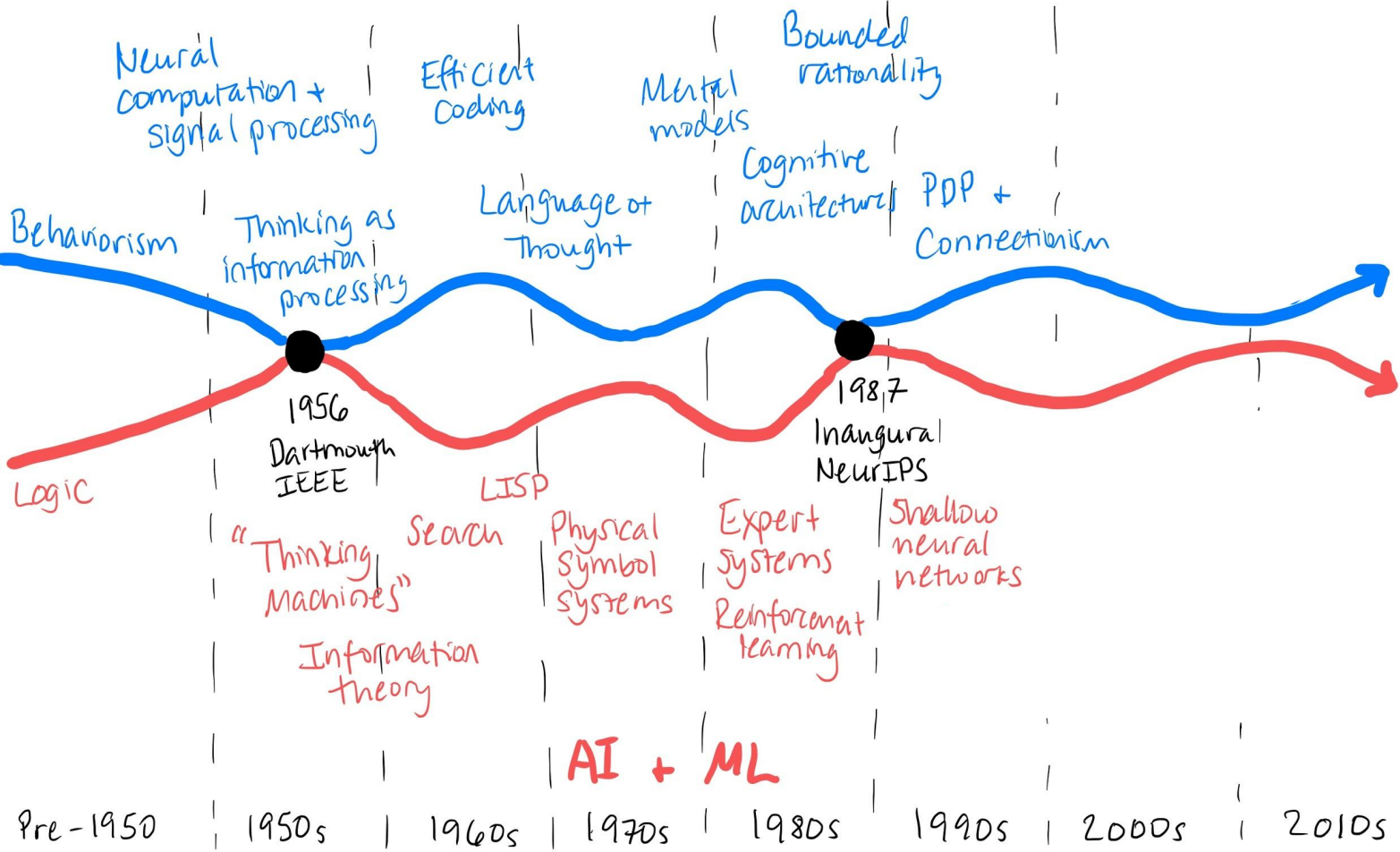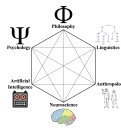
Two Conferences in 1987...
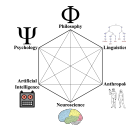
1st NeurIPS

- Centric Models of the Orientation Map in Primary Visual Cortex
- Simulations Suggest Information Processing Roles for the Diverse Currents in Hippocampal Neurons
- Optimal Neural Spike Classification
- Neural Networks for Template Matching: Application to Real-Time Classification of the Action Potentials of Real Neurons
- A Computer Simulation of Olfactory Cortex with Functional Implications for Storage and Retrieval of Olfactory Information
- Schema for Motor Control Utilizing a Network Model of the Cerebellum
- A Computer Simulation of Cerebral Neocortex
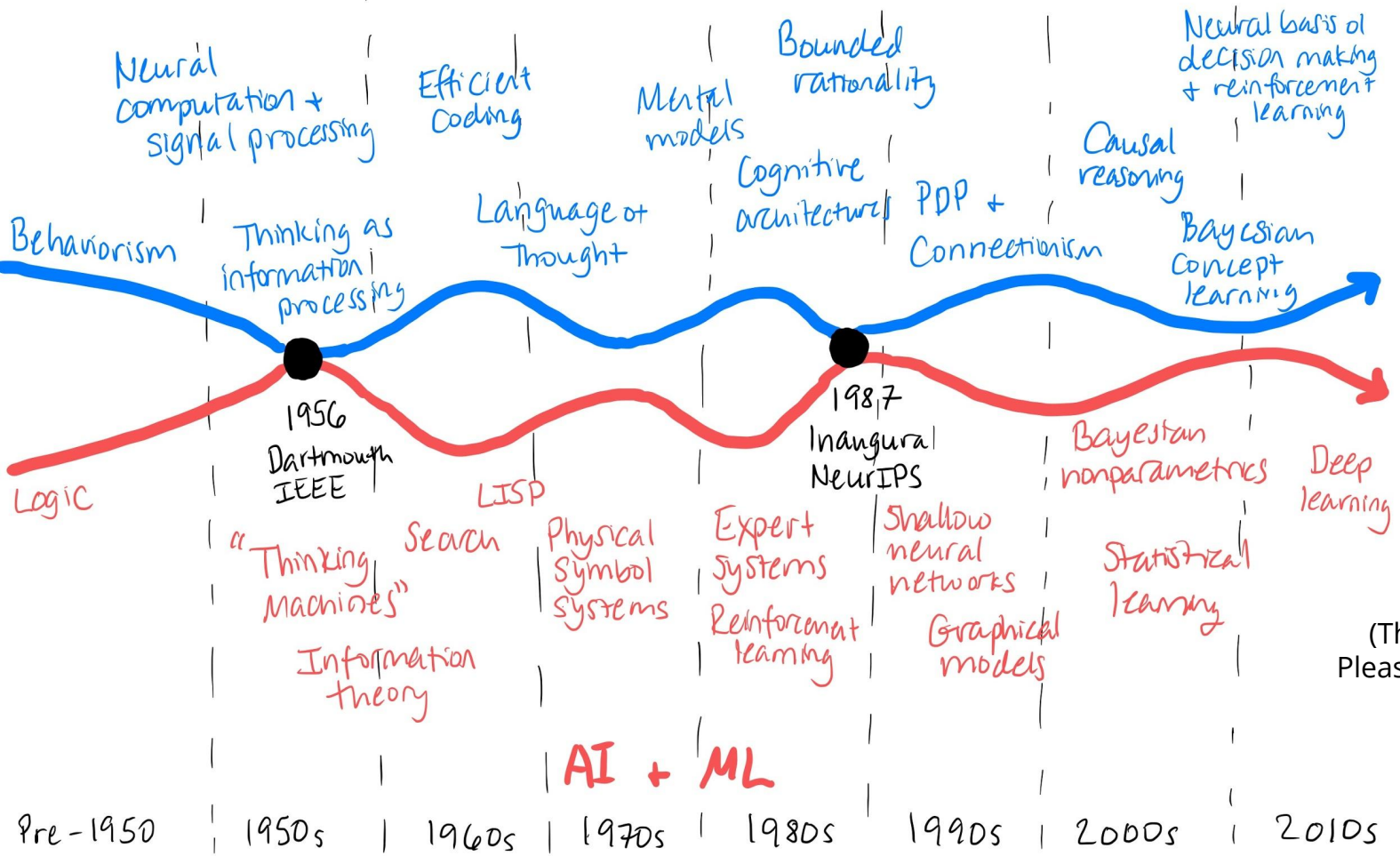- Discovering Structure from Motion in Monkey, Man and Machine

9th CogSci

- A Connectionist Network that Learns Natural Language Grammar
- A Connectionist Architecture for Representing and Reasoning about Structured Knowledge
- A Connectionist Encoding of Semantic Networks
- A Dual Back-Propagation Scheme for Scalar Reward Learning
- Using Fast Weights to Deblur Old Memories
- On the Connectionist Reduction of Conscious Rule Interpretation
- Cascaded Back-Propagation on Dynamic Connectionist Networks
- A Neural Network for the Detection and Representation of Oriented Edges
- Learning Internal Representations from Gray-Scale Images

COGNITIVE SCIENCE

Neural computation + signal processing

Efficient Coding

Mental models

Bounded rationality

Cognitive architectures

PDP + Connectionism

Behaviorism

Thinking as information processing

Language of Thought

Logic

1956 Dartmouth IEEE

"Thinking Machines"

Search

LISP

Physical Symbol Systems

Expert Systems

Reinforcement learning

1987 Inaugural NeurIPS

Shallow neural networks

Information theory

AI + ML

Pre-1950 | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s

COGNITIVE SCIENCE

Neural computation + signal processing

Efficient Coding

Mental models

Bounded rationality

Neural basis of decision making + reinforcement learning

Causal reasoning

Behaviorism

Thinking as information processing

Language of Thought

Cognitive architectures

PDP + Connectionism

Bayesian concept learning

BAICS Workshop!

Logic

1956 Dartmouth IEEE

1987 Inaugural NeurIPS

Bayesian nonparametrics

Deep learning

"Thinking Machines"

Search

Physical Symbol Systems

Expert Systems

Shallow neural networks

Statistical learning

Information theory

Reinforcement learning

Graphical models

(This is incomplete! Please send me ideas of things to add!)

AI + ML

Pre-1950 | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s
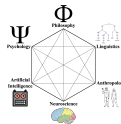
AI→CogSci: More powerful tools yield more powerful models
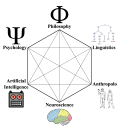
- Algorithms for language acquisition (Anderson, 1975)
- The Human Semantic Potential: Spatial Language and Constrained Connectionism (Regier, 1996)
- Bayesian statistics to model inductive reasoning (Tenenbaum et al., 2006)
- Probabilistic Topic Models (Steyvers & Griffiths, 2007)
- How to Grow a Mind: Statistics, Structure, and Abstraction (Tenenbaum et al., 2011)
- Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner (Dupoux, 2018)

AI→CogSci: More powerful tools yield more powerful models

- Algorithms for language acquisition (Anderson, 1975)
- The Human Semantic Potential: Spatial Language and Constrained Connectionism (Regier, 1996)
- Bayesian statistics to model inductive reasoning (Tenenbaum et al., 2006)
- Probabilistic Topic Models (Steyvers & Griffiths, 2007)
- **How to Grow a Mind: Statistics, Structure, and Abstraction (Tenenbaum et al., 2011)**
- Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner (Dupoux, 2018)

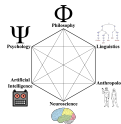# AI→CogSci: Computation influences theories of intelligence

- The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information (Miller, 1956)
- Syntactic Structures (Chomsky, 1957)
- Distributed Representations, Simple Recurrent: Networks, and Grammatical Structure (Elman, 1991)
- Re-thinking innateness: Development in a connectionist perspective (Elman et. al., 1996)
- Goal-driven deep learning to understand the sensory cortex (Yamins & DiCarlo, 2016)

AI→CogSci: Computation influences theories of intelligence

- **The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information (Miller, 1956)**
- Syntactic Structures (Chomsky, 1957)
- Distributed Representations, Simple Recurrent: Networks, and Grammatical Structure (Elman, 1991)
- Re-thinking innateness: Development in a connectionist perspective (Elman et. al., 1996)
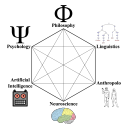- Goal-driven deep learning to understand the sensory cortex (Yamins & DiCarlo, 2016)

CogSci→AI: Algorithms and architectures draw inspiration

- Simulation of self-organizing systems by digital computer (Farley & Clark, 1954)
- Neocognitron (Fukushima, 1980)
- Physical Symbol Systems (Newell, 1980)
- Learning internal representations by error propagation (Rumelhart et. al., 1985)
- Parallel Distributed Processing (Rumelhart & McClelland, 1986)
- Finding Structure in Time (Elman, 1990)
- DQN (Mnih et al., 2015); Neuroscience-Inspired Artificial Intelligence (Hassabis et al., 2017)
- A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs (George et al., 2017)

CogSci→AI: Algorithms and architectures draw inspiration

- **Simulation of self-organizing systems by digital computer (Farley & Clark, 1954)**
- Neocognitron (Fukushima, 1980)
- Physical Symbol Systems (Newell, 1980)
- Learning internal representations by error propagation (Rumelhart et. al., 1985)
- Parallel Distributed Processing (Rumelhart & McClelland, 1986)
- Finding Structure in Time (Elman, 1990)
- DQN (Mnih et al., 2015); Neuroscience-Inspired Artificial Intelligence (Hassabis et al., 2017)
- A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs (George et al., 2017)
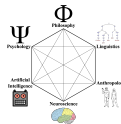
CogSci→AI: Human behavior helps us calibrate AI

- Computing Machinery and Intelligence (Turing, 1950)
- Human Problem Solving (Newell & Simon, 1972)
- The Winograd Schema Challenge (Levesque, 2012)
- Generating Legible Motion (Dragan & Srinivasa, 2013)
- Assessing the ability of LSTMs to learn syntax-sensitive dependencies (Linzen et al., 2016)
- Building Machines That Learn and Think Like People (Lake et. al., 2017)
- Analogues of Mental Simulation and Imagination in Deep Learning (Hamrick, 2019)
- Making AI more human (Gopnik, 2017)
- Evaluating theory of mind in question answering (Nematzadeh et al., 2018)

CogSci→AI: Human behavior helps us calibrate AI

- Computing Machinery and Intelligence (Turing, 1950)
- Human Problem Solving (Newell & Simon, 1972)
- **The Winograd Schema Challenge (Levesque, 2012)**
- Generating Legible Motion (Dragan & Srinivasa, 2013)
- Assessing the ability of LSTMs to learn syntax-sensitive dependencies (Linzen et al., 2016)
- Building Machines That Learn and Think Like People (Lake et. al., 2017)
- Analogues of Mental Simulation and Imagination in Deep Learning (Hamrick, 2019)
- Making AI more human (Gopnik, 2017)
- Evaluating theory of mind in question answering (Nematzadeh et al., 2018)

Open Questions in AI and CogSci

The ability to generalize
What inductive biases support the rapid learning that humans exhibit?

Learning representations from complex, noisy, and unsupervised data
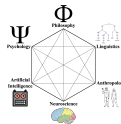How are concepts shared across multiple domains (e.g. language, movement, perception)?

Intelligence despite bounded cognition
How can models of the world be both approximate *and* useful?
How do memory limitations facilitate learning?

Interacting with other people
How should other people's goals and intentions be represented?

Open Questions in AI and CogSci

The ability to generalize
What inductive biases support the rapid learning that humans exhibit?

Learning representations from complex, noisy, and unsupervised data
How are concepts shared across multiple domains (e.g. language, movement, perception)?

Intelligence despite bounded cognition
How can models of the world be both approximate *and* useful?
How do memory limitations facilitate learning?

Interacting with other people
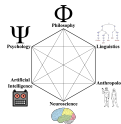How should other people's goals and intentions be represented?

A Non-Exhaustive List of Influential Cognitive Scientists

- **John Anderson:** rational analysis, ACT-R
- **Horace Barlow:** visual processing
- **Frederic Bartlett:** memory
- **Jerome Bruner**: "A Study of Thinking" (1956)
- **Susan Carey:** word and concept learning
- **Patricia Churchland:** philosophy of mind
- **Kenneth Craik:** mental models
- **Peter Dayan**: reinforcement learning
- **Jeff Elman:** recurrent neural networks
- **Jerry Fodor:** philosophy of mind
- **Rochel Gelman:** concept learning
- **Dedre Gentner:** mental models, analogy
- **Lila Gleitman:** language acquisition
- **Jacqueline Goodnow: "**A Study of Thinking" (1956)
- **Eleanor Gibson:** perceptual learning
- **James Gibson:** visual perception
- **Jane Goodall:** primatology
- **Geoffrey Hinton:** neural networks
- **Keith Holyoak:** analogy
- **Philip Johnson-Laird:** model theory
- **Daniel Kahneman:** decision making
- **Nancy Kanwisher:** visual perception

- **Stephen Kosslyn:** mental simulation
- **David Marr**: computational vision
- **James McClelland:** neural networks
- **Jacqueline Metzler:** mental representations
- **George Miller**: memory
- **Michael Posner**: attention
- **Molly (Mary) Potter:** word & image recognition
- **Eleanor Rosch:** categorization, concepts
- **David Rumelhart**: neural networks
- **Roger Shepard:** spatial reasoning
- **Rich Shiffrin:** attention and memory
- **Herb Simon**: bounded rationality
- **Linda Smith:** perceptual development
- **Paul Smolensky**: connectionist & symbolic reasoning
- **Elizabeth Spelke:** core knowledge
- **Michael Tanenhaus**: grounded language
- **Esther Thelen:** motor development
- **Edward Tolman:** cognitive maps
- **Michael Tomasello:** social cognition
- **Anne Treisman:** attention
- **Amos Tversky:** decision making
- **Shimon Ullman:** vision
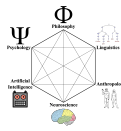
BAICS Program
Aida Nematzadeh

How Did BAICS Start?

An increased interest in multidisciplinary research in **AI** & **cognitive science**.

Create opportunities for discussions and collaborations between researchers from different fields by bringing them together in a smaller forum.

Inspired by the success of the *Cognitively Informed Artificial Intelligence* workshop at NeurIPS 2017.

Who Are The Organizers?
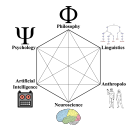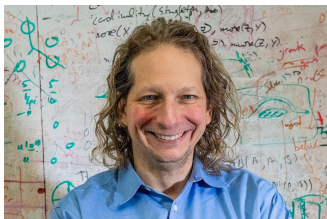

Kaylee Burns
Stanford
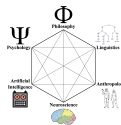
Jessica Hamrick
DeepMind

Aida Nematzadeh
DeepMind

Emmanuel Dupoux
EHESS/FAIR

Josh Tenenbaum
MIT

Alison Gopnik
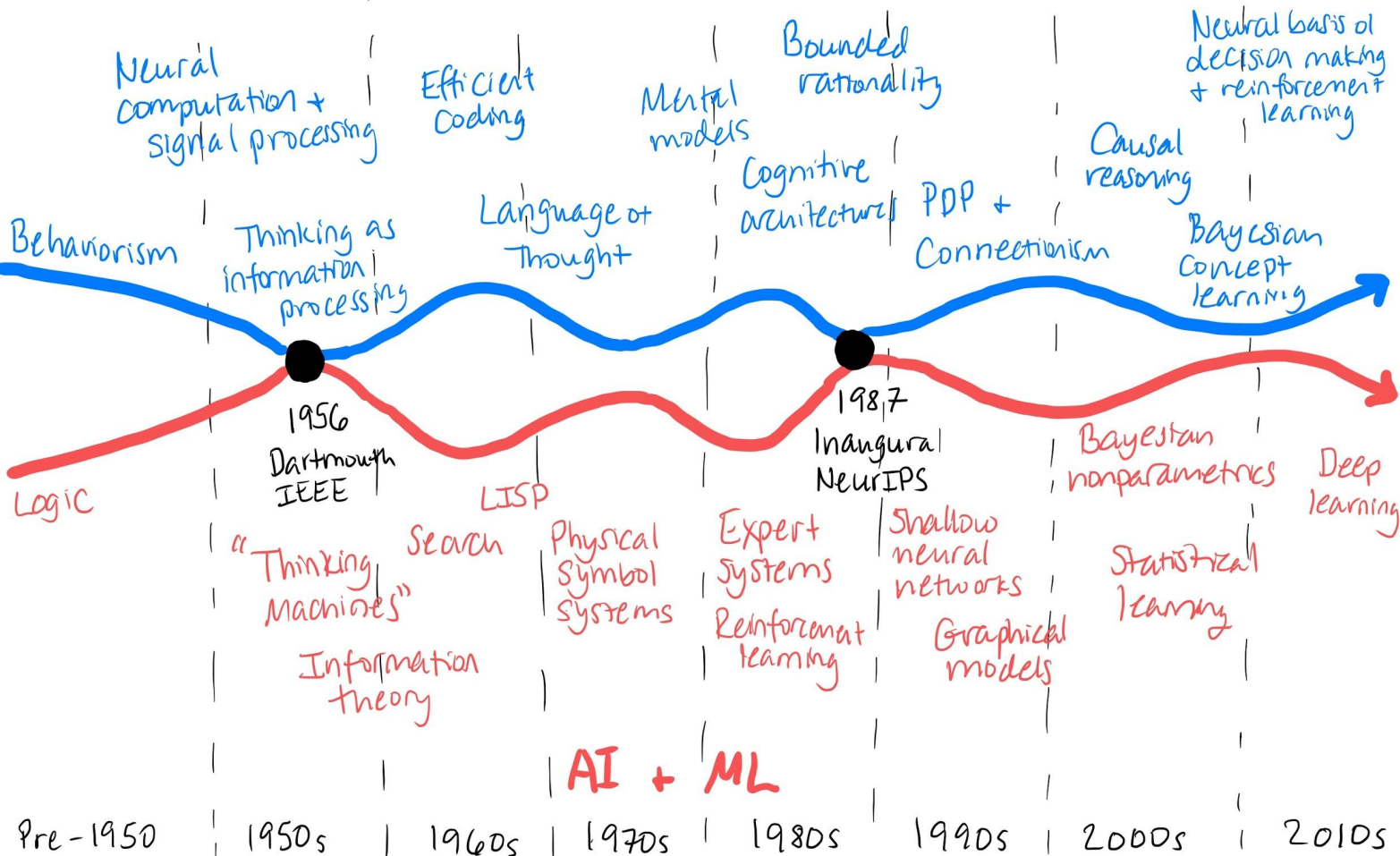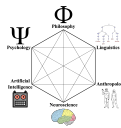UC Berkeley

The Rationale Behind Our Tracks

The **Research track** showcases work that combines data or methods from AI, cognitive science, and neuroscience.
- Designing evaluation tasks inspired by behavioral data.
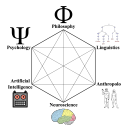- Computational models of cognition that leverage methods from AI.

The **Blue Sky Ideas track** encourages longer-term ideas and position papers.
- What aspects of cognition are under-explored in AI?
- How could AI research inform cognitive science?

COGNITIVE SCIENCE

BAICS Workshop!

Blue (Cognitive Science) timeline labels:
- Behaviorism
- Neural computation + signal processing
- Thinking as information processing
- Efficient Coding
- Language of Thought
- Mental models
- Cognitive architectures
- Bounded rationality
- PDP + Connectionism
- Causal reasoning
- Bayesian concept learning
- Neural basis of decision making + reinforcement learning

Red (AI + ML) timeline labels:
- Logic
- "Thinking Machines"
- Information theory
- 1956 Dartmouth IEEE
- Search
- LISP
- Physical Symbol Systems
- Expert Systems
- Reinforcement learning
- 1987 Inaugural NeurIPS
- Shallow neural networks
- Graphical models
- Bayesian nonparametrics
- Statistical learning
- Deep learning

AI + ML

Timeline: Pre-1950 | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s

The Review Process

We recruited a program committee of machine learning, cognitive science, and neuroscience researchers whose work span a wide range of domains (e.g., language, reinforcement learning, etc).

The program committee bid on the papers. Each paper received two reviews that include **quality score** and **presentation format**.

The organizers made decisions about acceptance and presentation format based on the reviews.

| submissions	accepted	posters	spotlight	contributed
63	43 (68%)	23 (36%)	16 (25%)	4 (6%)
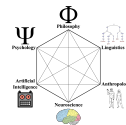
BAICS Program Committee

Adam Marblestone
Aishwarya Agrawal
Andrea Banino
Andrew Jaegle
Anselm Rothe
Ari Holtzman
Bas van Opheusden
Ben Peloquin
Bill Thompson
Charlie Nash
Danfei Xu
Emin Orhan

Erdem Biyik
Erin Grant
Jon Gauthier
Josh Merel
Joshua Peterson
Kelsey Allen
Kevin Ellis
Kevin McKee
Kevin Smith
Leila Wehbe
Lisa Anne Hendricks
Luis Piloto

Mark Ho
Marta Halina
Marta Kryven
Matthew Overlan
Max Kleiman-Weiner
Maxwell Forbes
Maxwell Nye
Michael Chang
Minae Kwon
Pedro Tsividis
Peter Battaglia
Qiong Zhang

Raphael Koster
Richard Futrell
Robert Hawkins
Sandy Huang
Stephan Meylan
Suraj Nair
Tal Linzen
Tina Zhu
Wai Keen Vong

Thanks for reviewing!
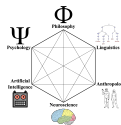
Plans for Today

Poster sessions are at the beginning and end of the day to increase remote participation from different time zones. You can chat with the authors during their poster session.

Panel discussion we be live! The panel consists of invited speakers and senior organizers.

Contributed and invited talks are streamed before and after the panel.

The schedule is available at the BAICS website:
https://baicsworkshop.github.io/

How to Participate Remotely

Attend the poster session: join Zoom/RocketChat and talk to the authors!
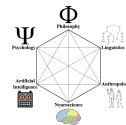
Engage with the panelists: vote and submit questions at Sli.do.

Ask questions from invited speakers: join Zoom/RocketChat and submit your questions to the invited speakers.

Get to know other participants.

Follow #BAICS2020 on twitter. We will be tweeting: **@kaylburns**, @**jhamrick**, and @**aidanematzadeh.**

For more details, see the BAICS website:
https://baicsworkshop.github.io/