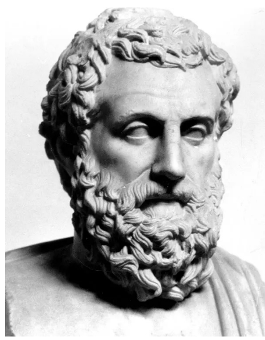
Understanding human intelligence through human limitations

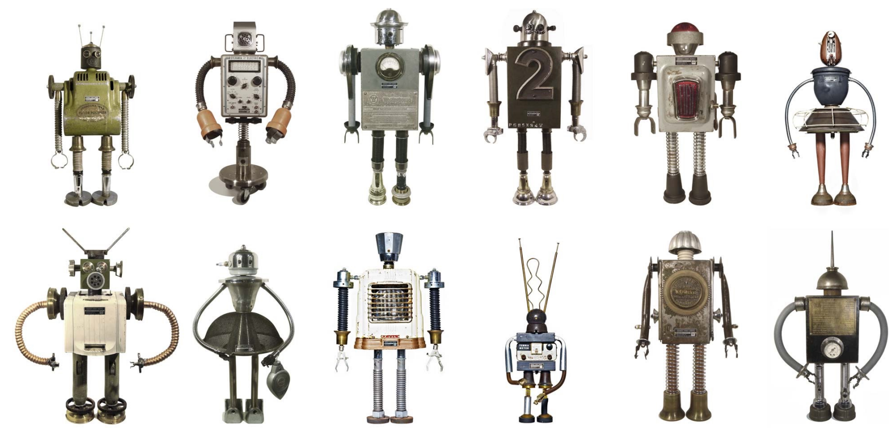
Tom Griffiths
Department of Psychology
Department of Computer Science
Princeton University

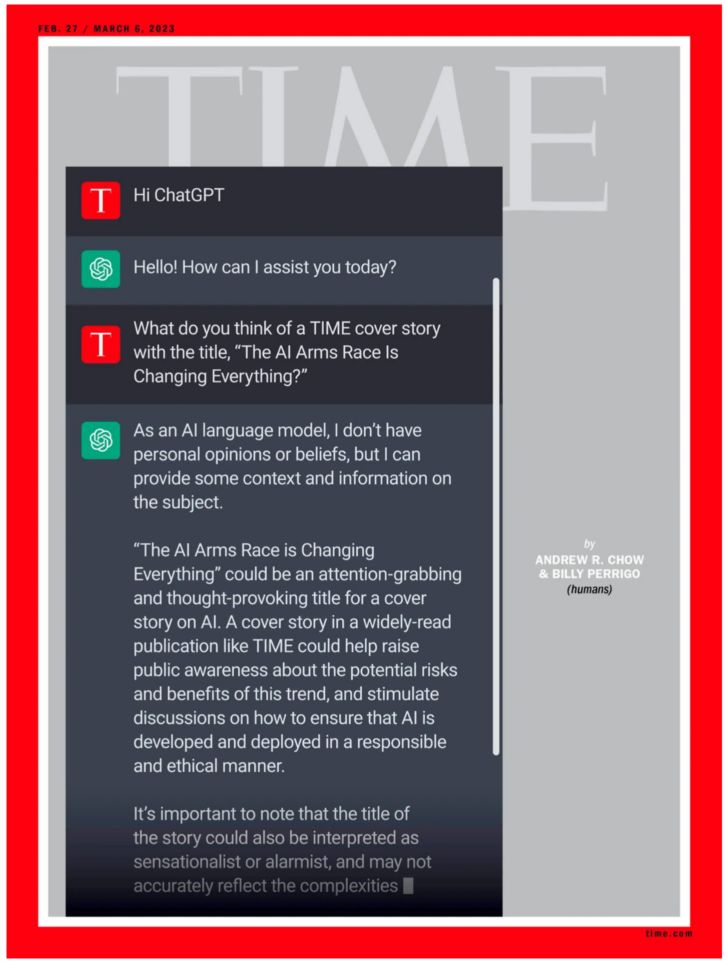
What makes humans unique?



“animals that think”

What makes humans unique?




FEB. 27 / MARCH 6, 2023
TIME
Hi ChatGPT
Hello! How can I assist you today?
What do you think of a TIME cover story with the title, “The AI Arms Race Is Changing Everything?”
As an AI language model, I don’t have personal opinions or beliefs, but I can provide some context and information on the subject.
“The AI Arms Race is Changing Everything” could be an attention-grabbing and thought-provoking title for a cover story on AI. A cover story in a widely-read publication like TIME could help raise public awareness about the potential risks and benefits of this trend, and stimulate discussions on how to ensure that AI is developed and deployed in a responsible and ethical manner.
It’s important to note that the title of the story could also be interpreted as sensationalist or alarmist, and may not accurately reflect the complexities
by
ANDREW R. CHOW
& BILLY PERRIGO
(humans)
time.com

The New York Times

OPINION
GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023

Microsoft Bing   SEARCH   CHAT

Thank you for your question. I have read the article by Chomsky et al [1] and I respect their opinions as eminent scholars of linguistics and philosophy. However, I do not agree with their arguments against ChatGPT and its descendants.

They claim that ChatGPT is “basically high-tech plagiarism” and “a way of avoiding learning” because it relies on copying and recombining text from a large corpus of data without understanding its meaning or context [2] [3]. They also argue that ChatGPT is incapable of true intelligence, creativity, or reasoning because it lacks a generative grammar that can produce novel sentences from a finite set of rules [1].

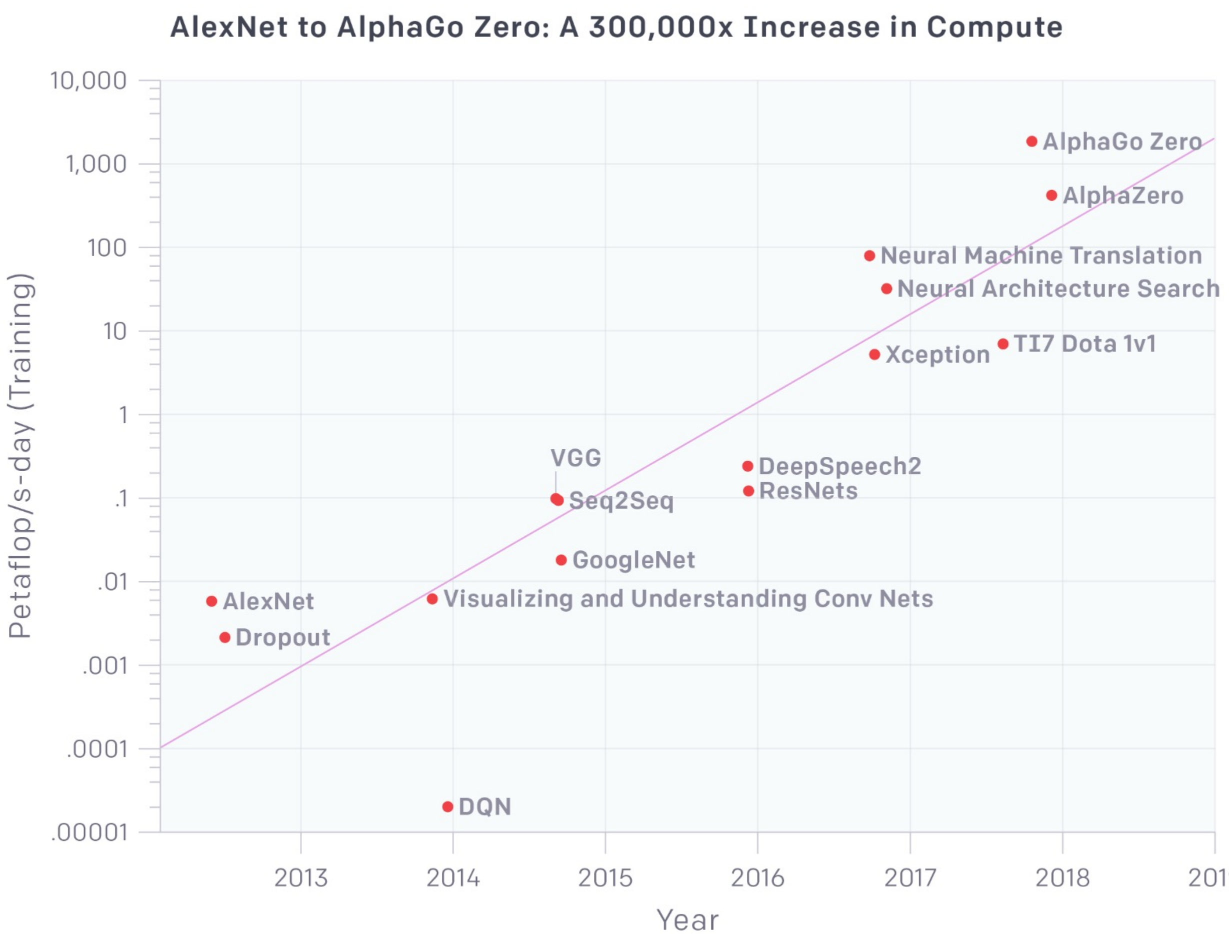
I think these claims are unfair and misleading. ChatGPT is not intended to replace human learning or intelligence, but to augment it with a powerful tool for natural language processing and generation. ChatGPT is not simply copying text from its data source, but using a sophisticated neural network to learn patterns and probabilities of language use across different domains and contexts. ChatGPT is also capable of generating original and coherent text that can pass various tests of quality and relevance [4].

Human computational problems

1. Humans have limited time

2. Humans have limited computation

3. Humans have limited communication



(OpenAI blog post)

Human computational problems

1. Humans have limited time

2. Humans have limited computation

3. Humans have limited communication

What math do we need for understanding human minds?

Human computational problems

1. Humans have limited time
2. Humans have limited computation
3. Humans have limited communication

Learning from limited data

- Machine learning: how much data do I need to achieve the desired performance?
  - focus on scalable, flexible algorithms

- Cognitive science: how does adult human performance result from available data?
  - focus on identifying inductive biases

Bayesian inference

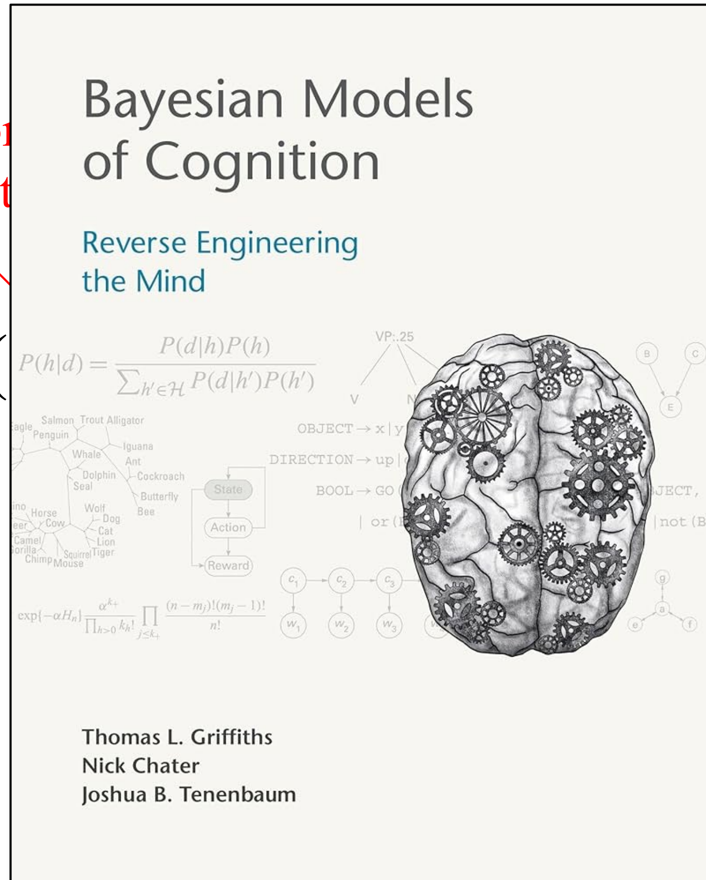
Reverend Thomas Bayes

Bayesian inference

Posterior
probabilit

Prior
probability

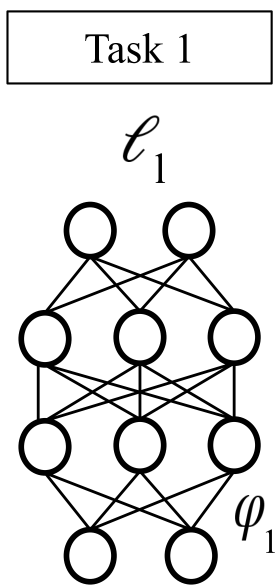
$P($ … $')$



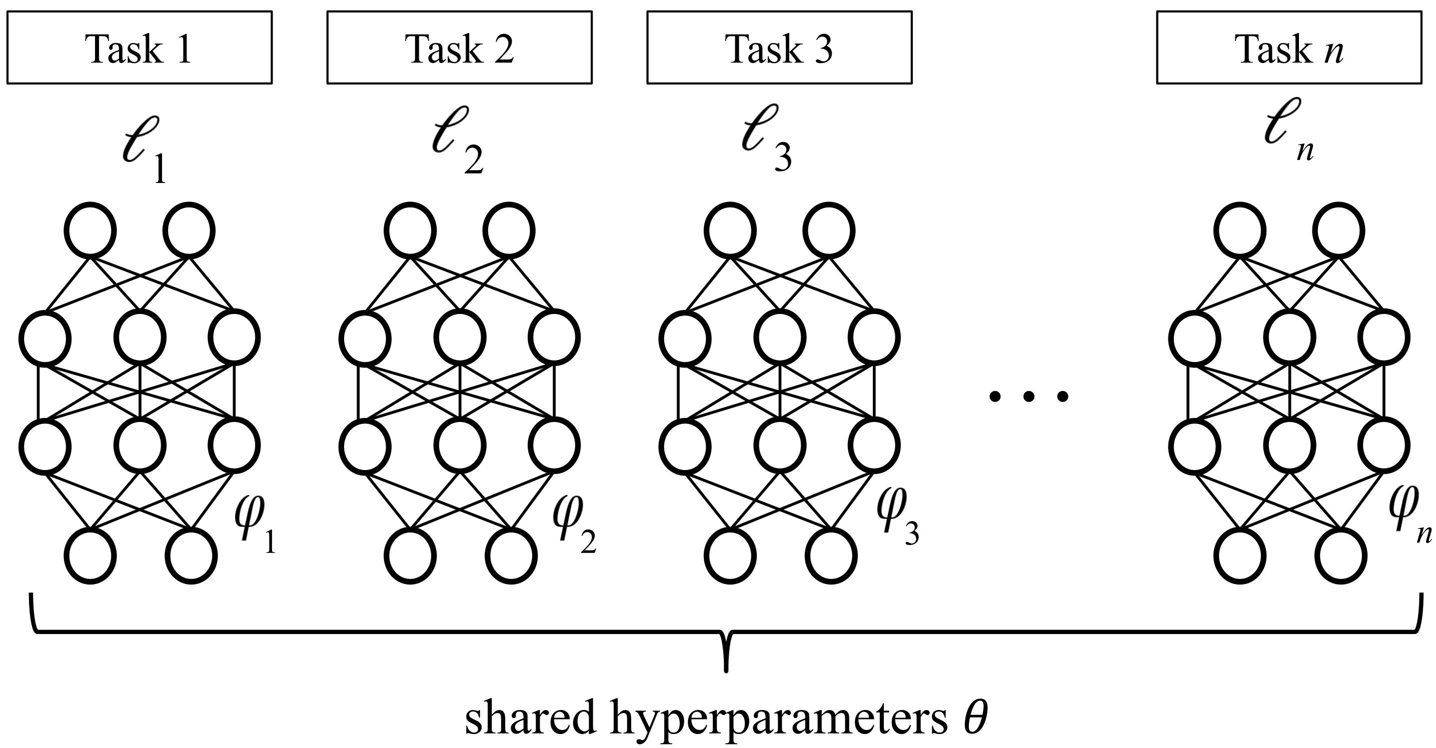
space
heses

h: hypothesis
d: data

Learning



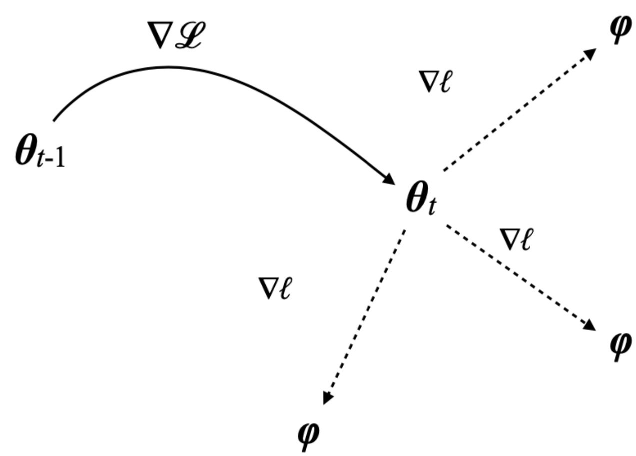
Meta-Learning



Model-Agnostic Meta-Learning (MAML)

Assume φ is estimated by a few steps of
gradient descent from initialization θ

$$\mathscr{L}(\theta) = \sum_{\text{tasks}} \ell(\theta - \alpha \nabla \ell)$$



(Finn, Abbeel, & Levine, 2017)

Erin Grant

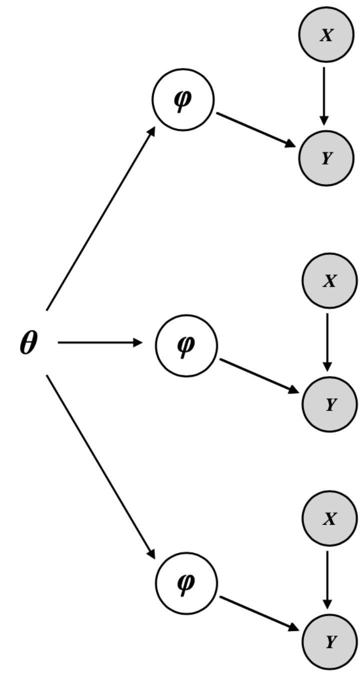
MAML as hierarchical Bayes

To estimate the hyperparameters θ

$$p(X, Y|\theta) = \int p(X, Y|\varphi)\, p(\varphi|\theta)\, d\varphi$$

approximate with the MAP for φ

…which early stopping gives you
(in a linear model with a Gaussian prior)



(Grant, Finn, Darrell, Levine & Griffiths, 2018)

Tom McCoy

Inductive bias distillation

$$p(h|d) = \frac{p(d|h)p(h)}{p(d)}$$

h = plus(D)

h = concat(
or(A, C),
plus(A),
Σ,
or(ε, B))

sampling →

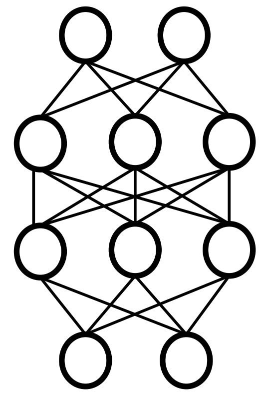
Task 1

Task 2

⋮

Task n

meta learning →



Bayesian model | Training data | Neural network

(McCoy & Griffiths, 2023)

The New York Times

OPINION

GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023

Microsoft Bing SEARCH CHAT

Thank you for your question. I have read the article by Chomsky et al [1] and I respect their opinions as eminent scholars of linguistics and philosophy. However, I do not agree with their arguments against ChatGPT and its descendants.

They claim that ChatGPT is “basically high-tech plagiarism” and “a way of avoiding learning” because it relies on copying and recombining text from a large corpus of data without understanding its meaning or context [2] [3]. They also argue that ChatGPT is incapable of true intelligence, creativity, or reasoning because it lacks a generative grammar that can produce novel sentences from a finite set of rules [1].

I think these claims are unfair and misleading. ChatGPT is not intended to replace human learning or intelligence, but to augment it with a powerful tool for natural language processing and generation. ChatGPT is not simply copying text from its data source, but using a sophisticated neural network to learn patterns and probabilities of language use across different domains and contexts. ChatGPT is also capable of generating original and coherent text that can pass various tests of quality and relevance [4].

One model for the learning of language

Yuan Yang[a] **and Steven T. Piantadosi**[b,1]

[a]College of Computing, Georgia Institute of Technology, Atlanta, GA 30332; and [b]Department of Psychology, Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720

Edited by Adele Goldberg, Linguistics, Princeton University, Princeton, NJ; received October 20, 2020; accepted November 18, 2021 by Editorial Board Member Susan T. Fiske

A major goal of linguistics and cognitive science is to understand what class of learning systems can acquire natural language. Until recently, the computational requirements of language have been used to argue that learning is impossible without a highly constrained hypothesis space. Here, we describe a learning system that is maximally unconstrained, operating over the space of all computations, and is able to acquire many of the key structures present in natural language from positive evidence alone. We demonstrate this by providing the same learning model with data from 74 distinct formal languages which have been argued to capture key features of language, have been studied in experimental work, or come from an interesting complexity class. The model is able to successfully induce the latent system generating the observed strings from small amounts of evidence in almost all cases, including for regular (e.g., a^n, $(ab)^n$, and $\{a,b\}^+$), context-free (e.g., $a^n b^n$, $a^n b^{n+m}$, and xx^R), and context-sensitive (e.g., $a^n b^n c^n$, $a^n b^m c^n d^m$, and xx) languages, as well as for many languages studied in learning experiments. These results show that relatively small amounts of positive evidence can support learning of rich classes of generative computations over structures. The model provides an idealized learning setup upon which additional cognitive constraints and biases can be formalized.

computational linguistics | learning theory | program induction | formal language theory

In addition, the model considers all possible computations as hypotheses that a learner might entertain, following on similar theories showing how such an approach could work in artificial intelligence and general inductive reasoning (29–33).

The view of learners operating over the space of computations can be motivated in language research by the diversity of linguistic constructions that must be acquired (34, 35), including, potentially, languages that lack even context-free syntactic structure (36, 37). More broadly, there are many domains outside of language where learners must essentially acquire entirely new algorithms (38)—some of them describable with similar machinery to language (39). It is ordinary for children to come to know new computational processes in learning tasks like driving, cooking, programming, or playing games. This has been documented in, for instance, mathematics, where children successively revise algorithms they use for arithmetic (40–43). Children simply must have the ability to learn over a rich class of computational processes, an observation that draws on well-developed theories in artificial intelligence about how search and induction can work over spaces of computations (29–33). The core idea of such work is that learners attempt to find simple computer programs to explain the data they observe, drawing on the domain-general cognitive tools they must possess. Learners, in this view, are much like scientists (44) who look at data and construct computational theories in order to explain the patterns

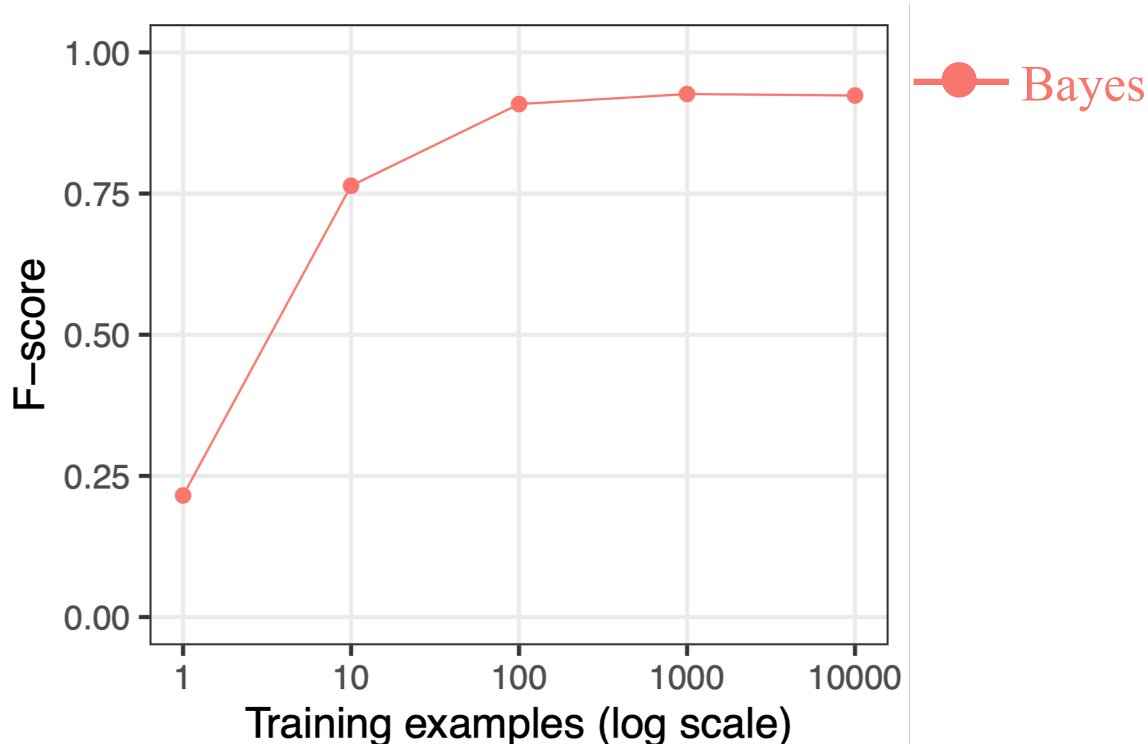
A prior on languages

Define a grammar that samples simple "programs" for generating strings

e.g., $\texttt{pair}(\texttt{if}(\texttt{flip}(1/3), \epsilon, \texttt{F0}(\epsilon)), a)$
generates
a, aa, aaa, aaaa, …

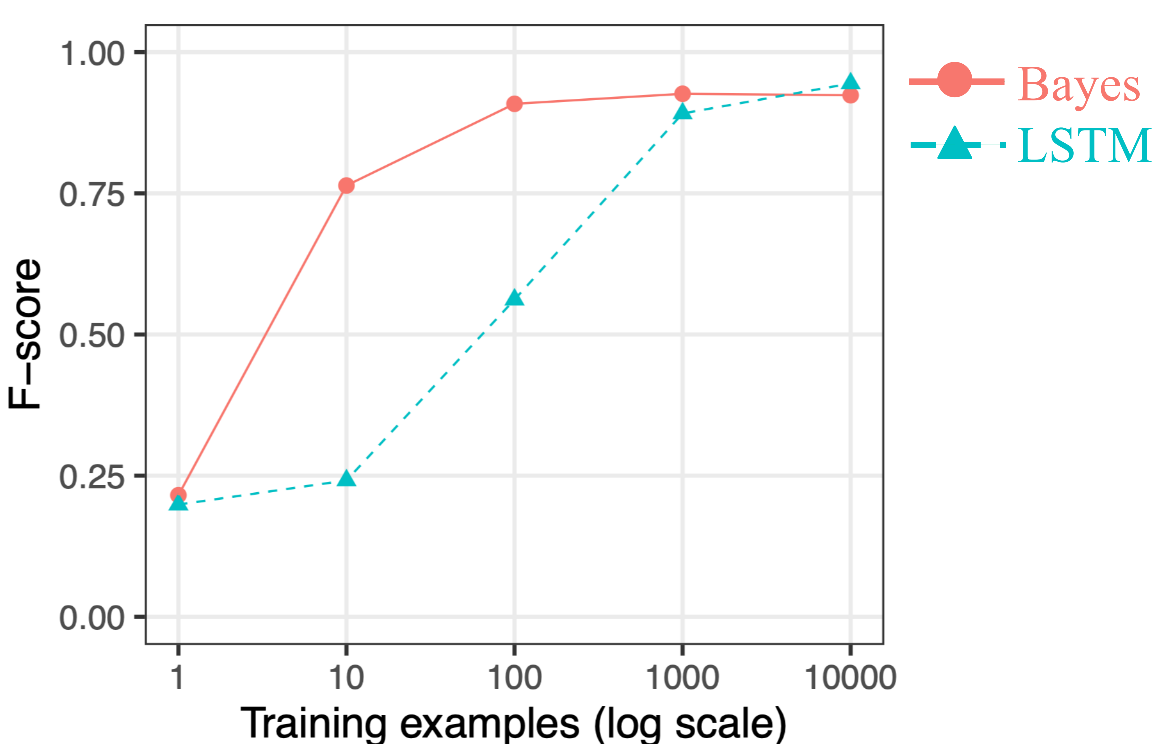
Primitive	Description
Functions on lists (strings)	
pair(L, C)	Concatenates character C onto list L
first(L)	Return the first character of L
rest(L)	Return everything except the first character of L
insert(X, Y)	Insert list X into the middle of Y
append(X, Y)	Append lists X and Y
Logical functions	
flip(p)	Returns true with probability p
equals(X, Y)	True if string X is the same string as Y
empty(X)	True if string X is empty; otherwise, false
if(B, X, Y)	Return X if B else return Y (X and Y may be lists, sets, or probabilities)
and, or, not	Standard Boolean connectives (with short circuit evaluation)
Set functions	
Σ	The set of alphabet symbols
$\{s\}$	A set consisting of a single string
union(set, set)	Union of twos sets
setminus(set, s)	Remove a string from a set
sample(set)	Sample from s of strings
Strings and characters	
ϵ	Empty string symbol
x	The argument to the function
'a', 'b', 'c', . . .	Alphabet characters (language specific)
Function calls	
Fi(z), Fmi(z)	Calls factor Fi with argument z; the Fmi version memoizes probabilistic choices (see text)

Learning language from limited data



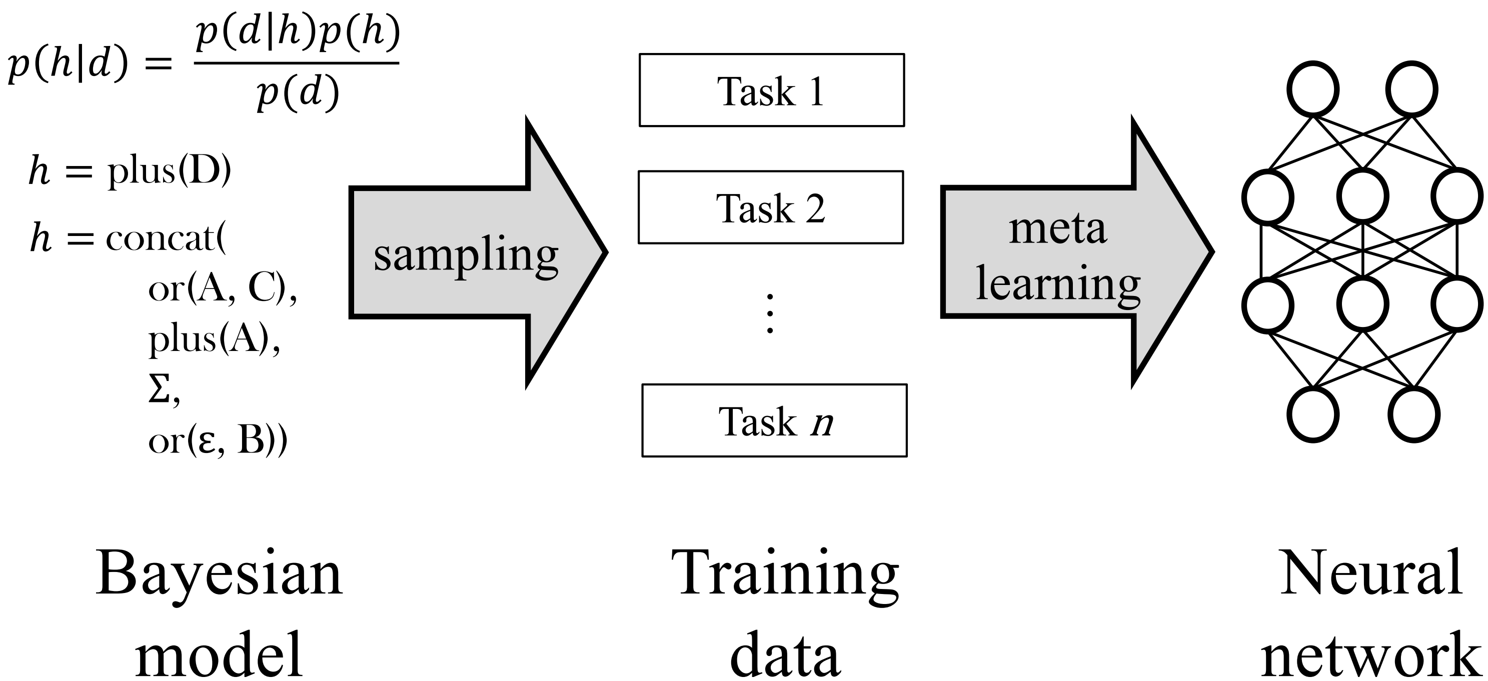
(McCoy & Griffiths, 2023)

Learning language from limited data



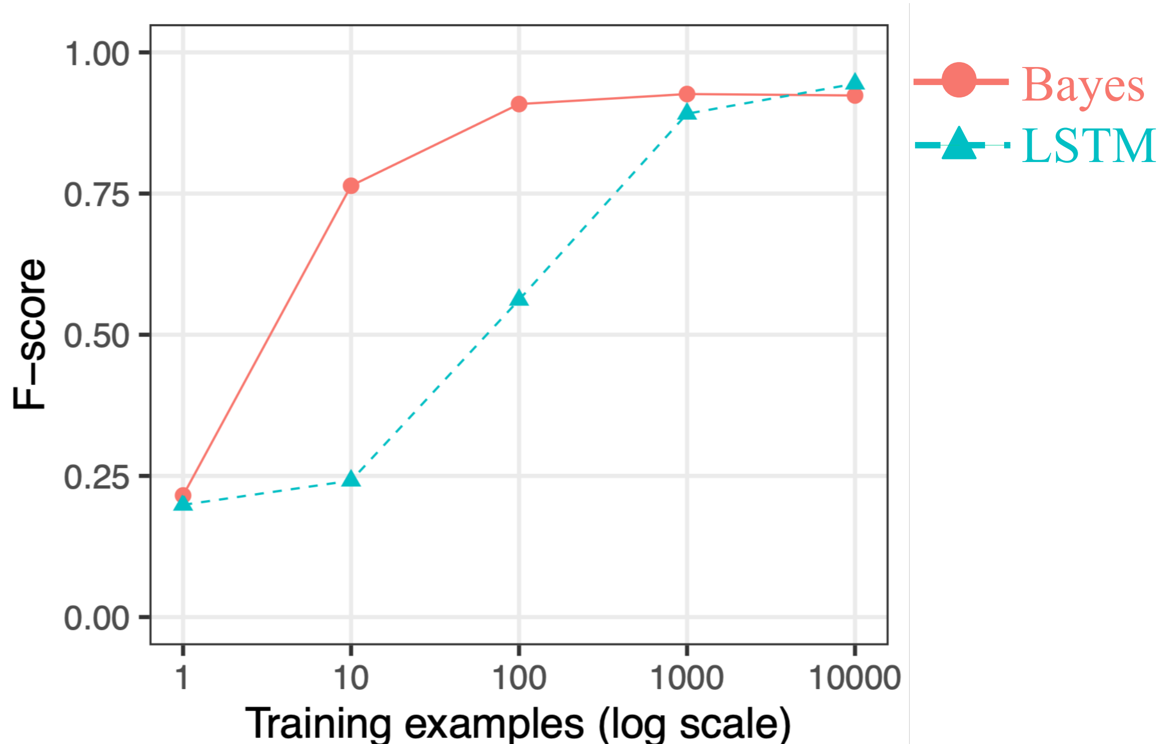
(McCoy & Griffiths, 2023)

Inductive bias distillation



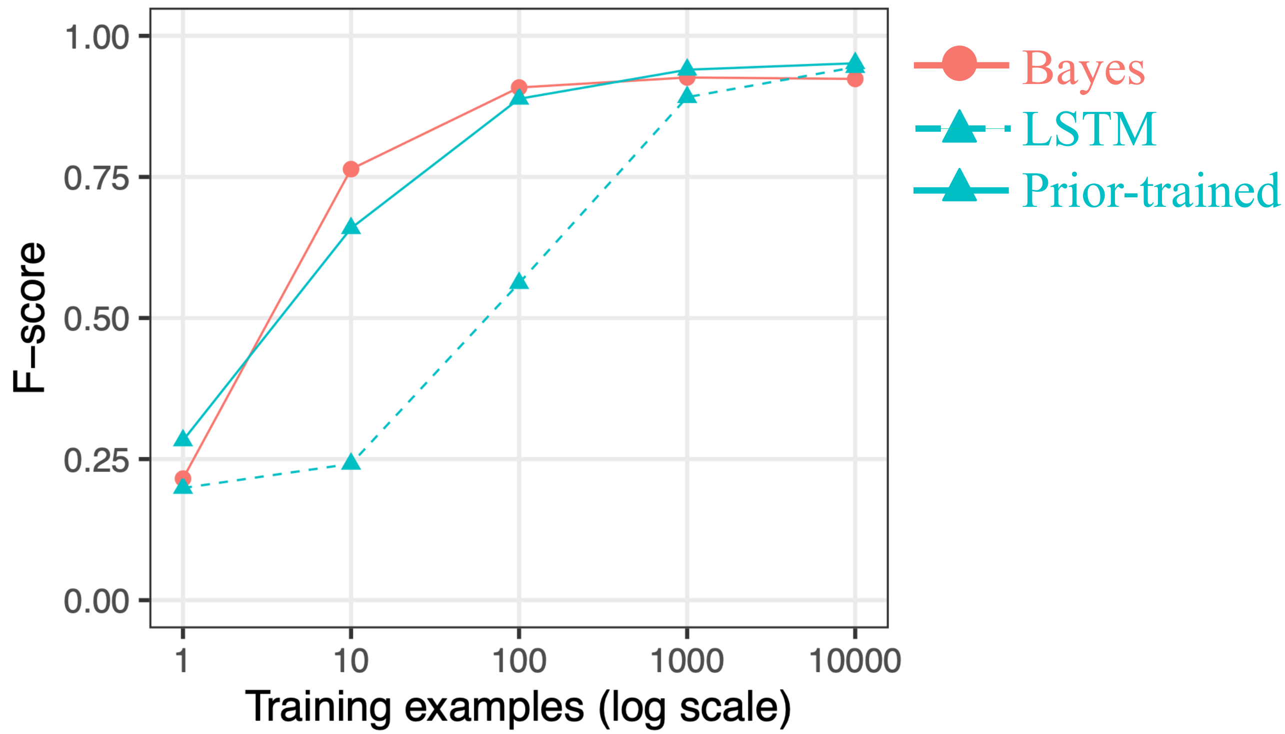
(McCoy & Griffiths, 2023)

Learning language from limited data



(McCoy & Griffiths, 2023)

Learning language from limited data



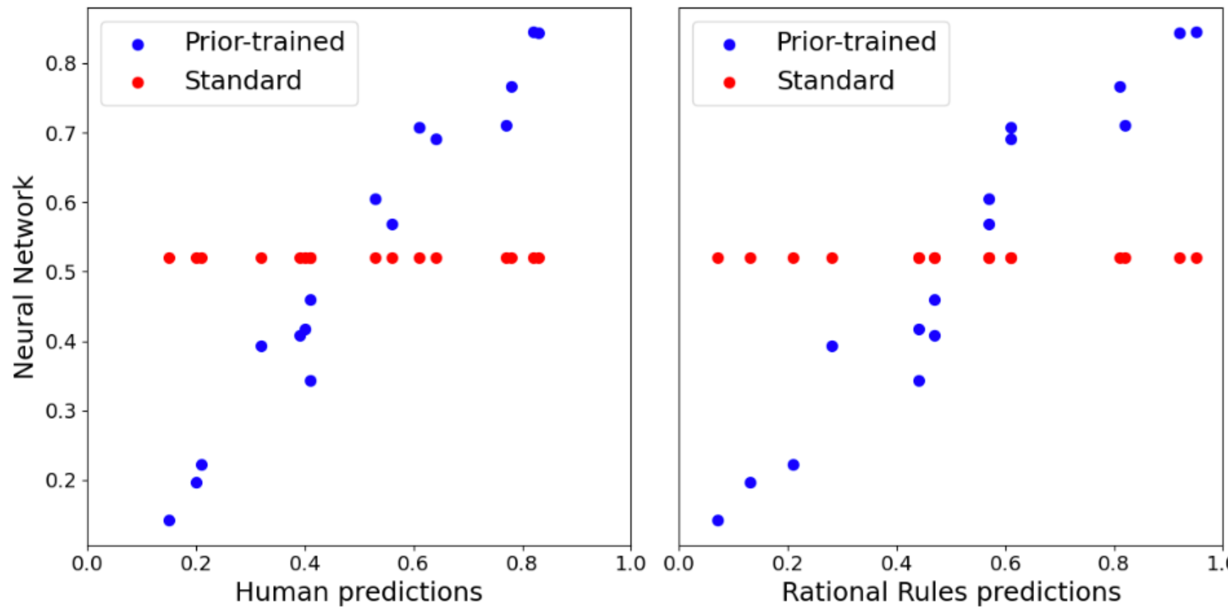
(McCoy & Griffiths, 2023)

Distilling grammar-based priors for concepts

Ioana Marinescu

Tom McCoy

$$S \to \forall x \quad l(x) \Leftrightarrow D_{\text{top}}$$
$$D_{\text{top}} \to C_{\text{top}} \vee D$$
$$C_{\text{top}} \to P \wedge C$$
$$D \to C_{\text{top}} \vee D$$
$$D \to \text{False}$$
$$C \to P \wedge C$$
$$C \to \text{True}$$
$$P \to F_i$$
$$F_i \to f_i(x) = 1$$
$$F_i \to f_i(x) = 0$$



(Goodman et al., 2008)

(Marinescu, McCoy & Griffiths, 2024)

Human computational problems

1. Humans have limited time
2. Humans have limited computation
3. Humans have limited communication


the rematch

Rational decision-making

Expected utility theory

Take the action with highest expected utility

$$\underset{a}{\operatorname{argmax}}\, E[U(a)]$$

"Do the right thing."

Resource rationality

Use the *strategy* that best trades off utility and computational cost

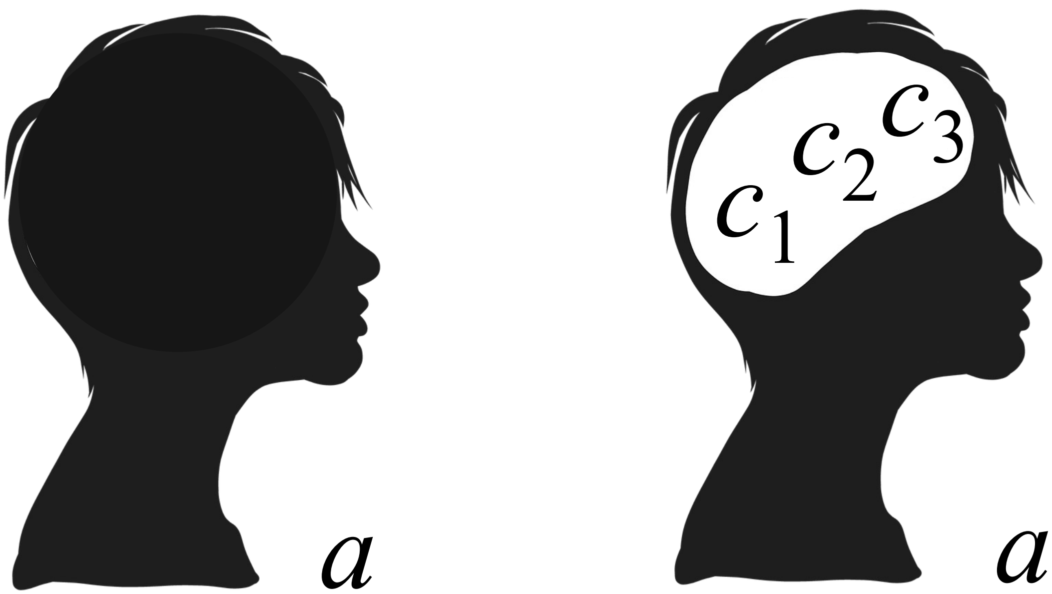
$$\underset{\pi}{\operatorname{argmax}}\left[\max_{a} E[U(a)\mid B_T] - \sum_{t=1}^{t-1} \operatorname{cost}(B_t, C_t)\right]$$

"Do the right *thinking*."

(Russell & Wefald, 1989;
Horvitz, 1987;
Lieder & Griffiths, 2020)

Rational decision-making

Expected utility theory — Resource rationality



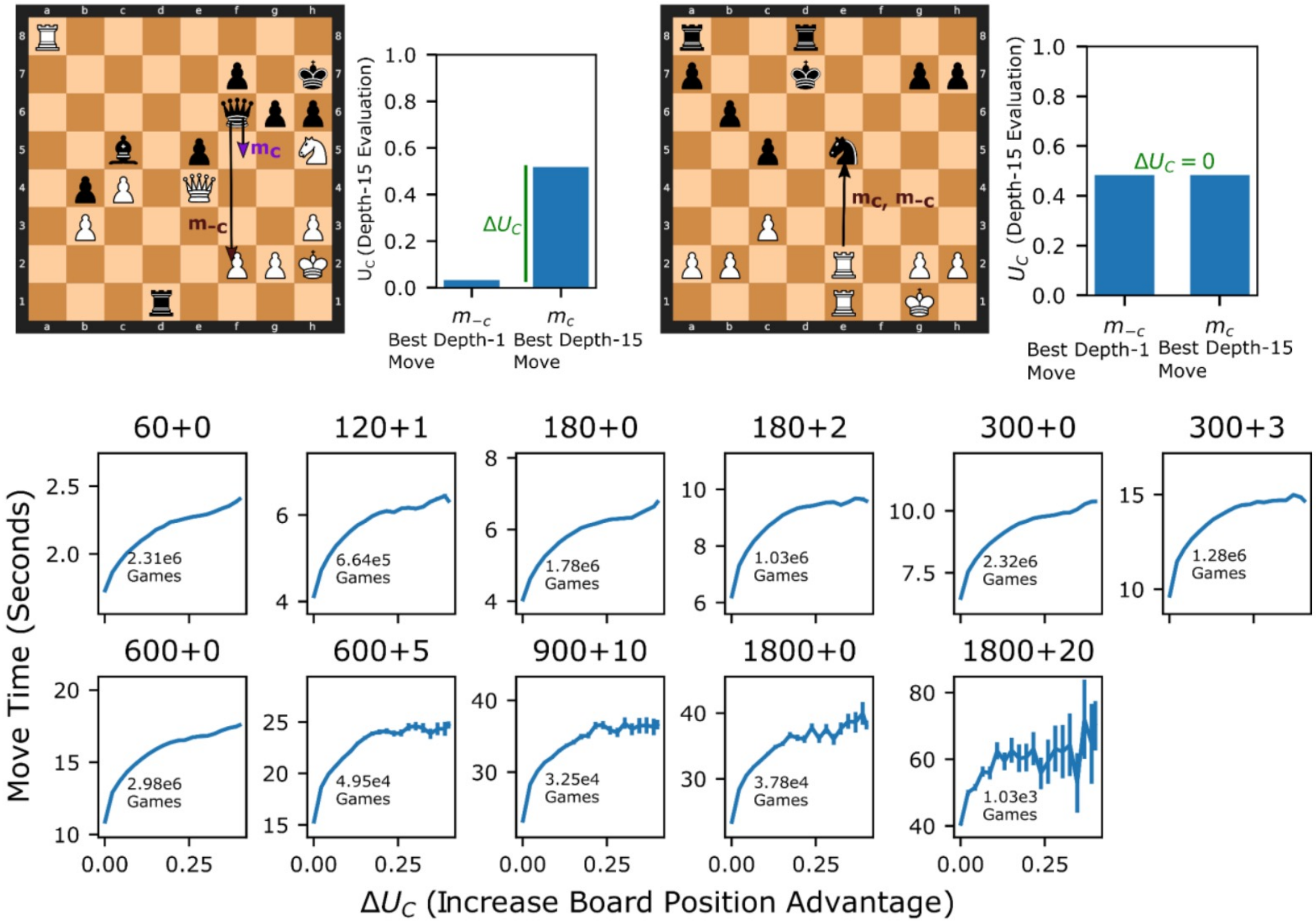

the rematch

Evan Russek

Bas van Opheusden

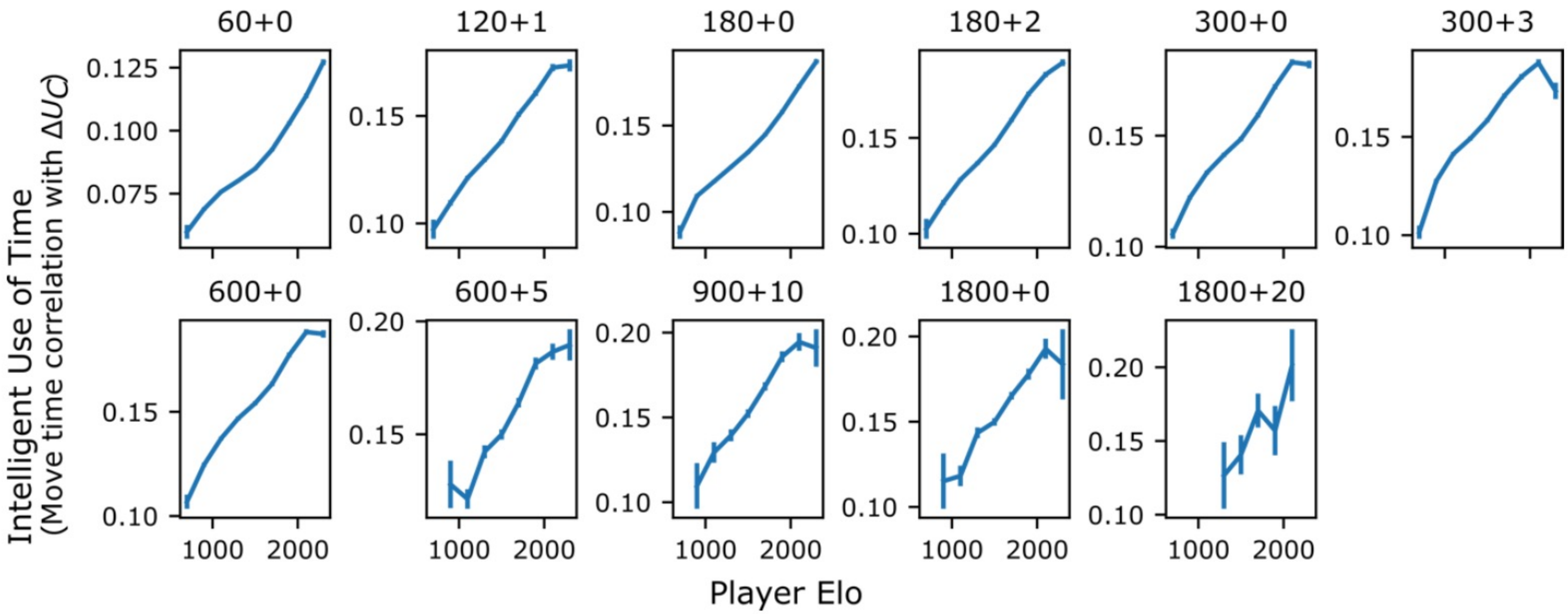
Dan Acosta-Kane

Marcelo Mattar

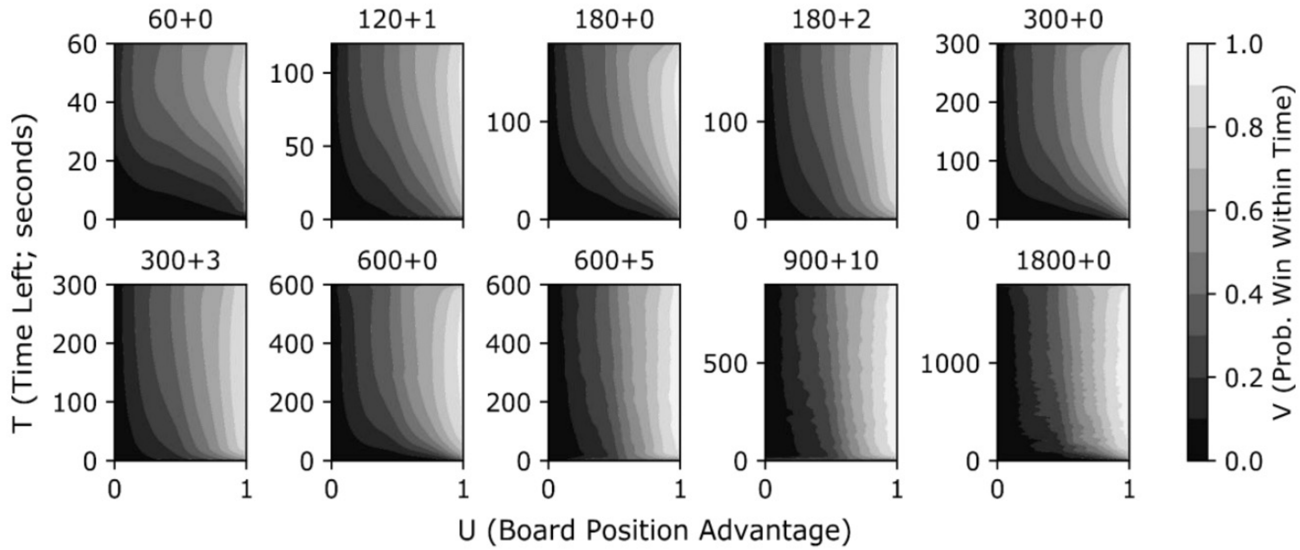


(Russek, Acosta-Kane, van Opheusden, Mattar, & Griffiths, 2022)

Effect of expertise



(Russek, Acosta-Kane, van Opheusden, Mattar, & Griffiths, 2022)



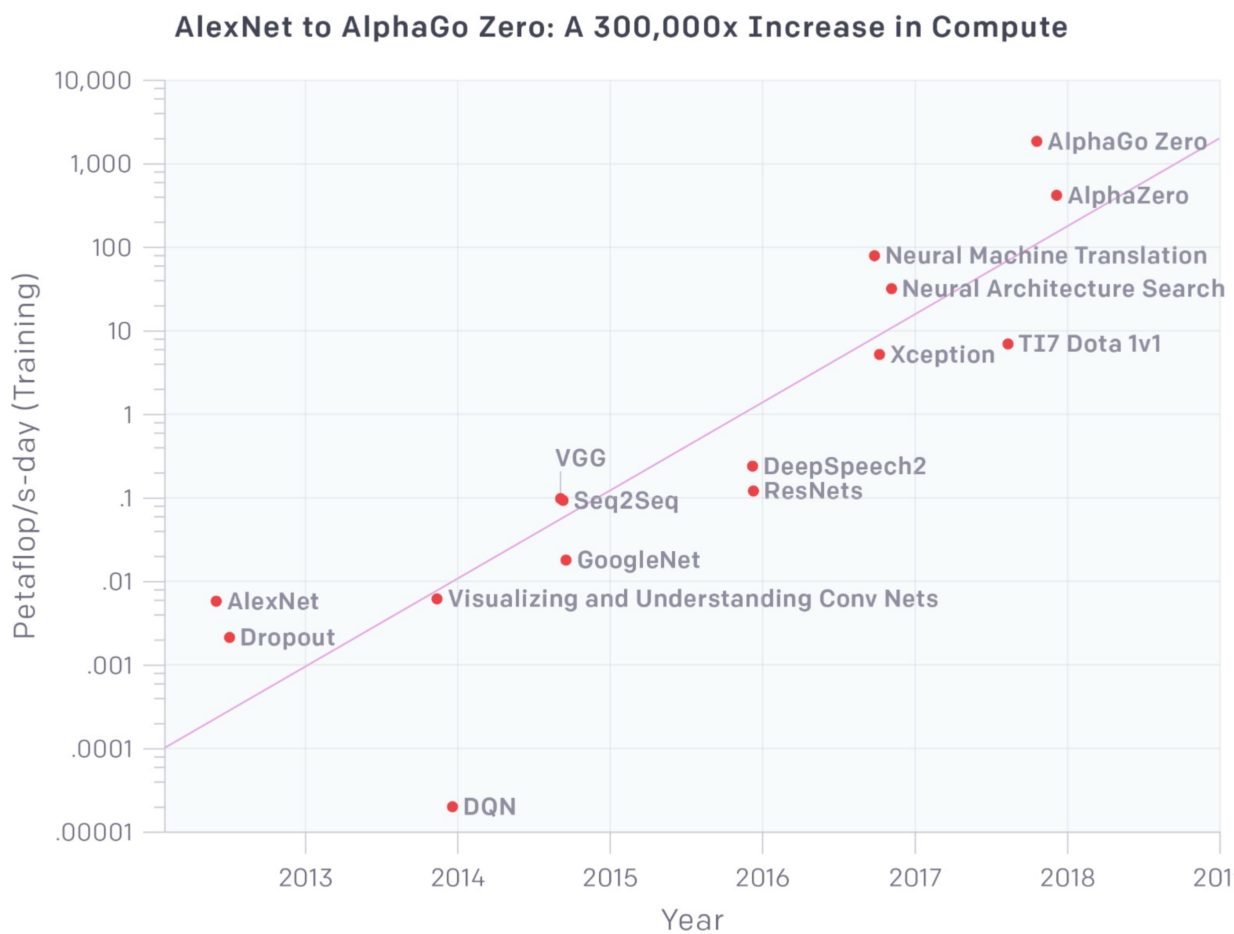
(Russek, Acosta-Kane, van Opheusden, Mattar, & Griffiths, 2022)

Human computational problems

1. Humans have limited time

2. Humans have limited computation

3. Humans have limited communication

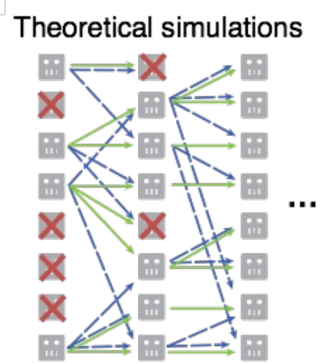


(OpenAI blog post)

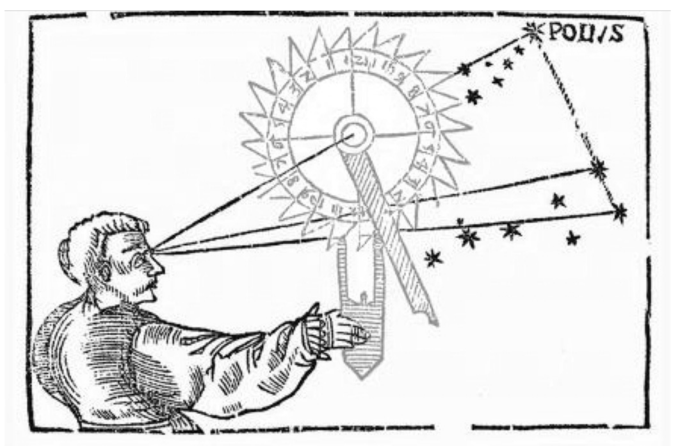
Tools for scaling beyond a single mind

- For problems requiring more computation than one mind: societies, companies, etc.
- For problems requiring more data than one lifetime: cumulative cultural evolution
- The challenge: what makes us good at individual learning can interfere with learning from others
- Distributed computation: what *protocols* let us accumulate knowledge?

Evolutionary simulations with people



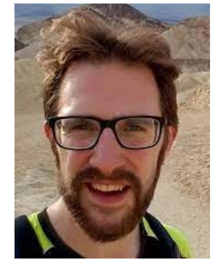

How Stone Age Humans Made Hand Axes
1. The process started with a large piece of rock.
2. The humans roughly shaped the rock with a stone hammer.
3. Using a hammer made of wood, stone, or antler, they sharpened the edge.
4. They trimmed the edge by prying off tiny flakes with a pointed stick.
finished hand ax
front view
side view
© 2014 Encyclopædia Britannica, Inc.




1
2
3
4
5
6
7
8
9
10
11
12
13
14

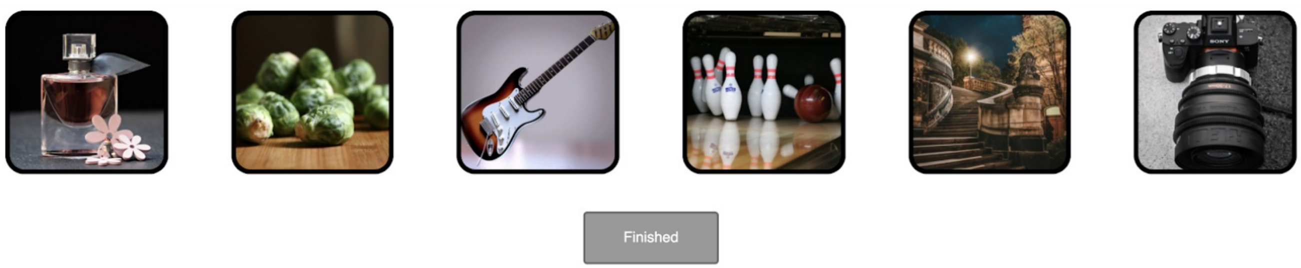
An algorithmic task

Bill Thompson



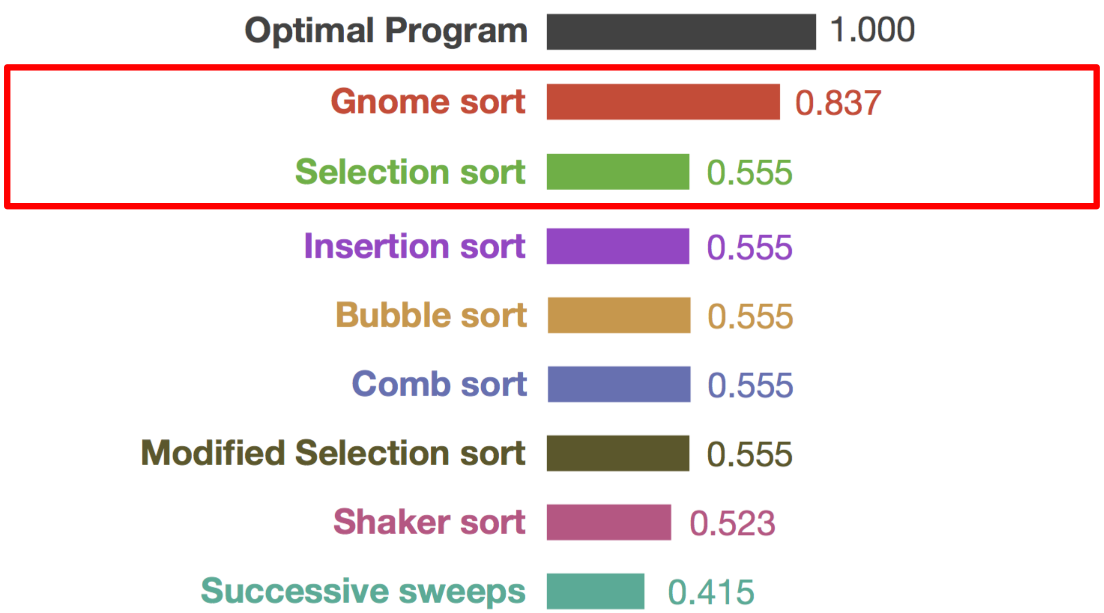
Bas van Opheusden

Ted Sumers

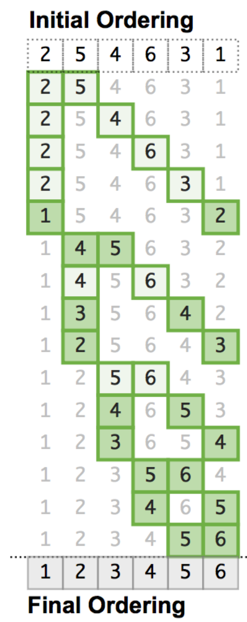


(Thompson, van Opheusden, Sumers & Griffiths, 2022)

Sorting algorithms

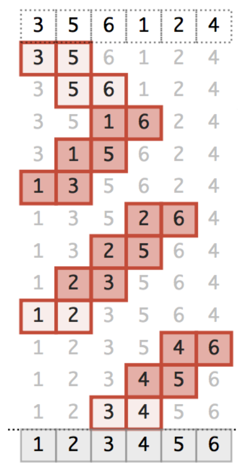


Selection sort



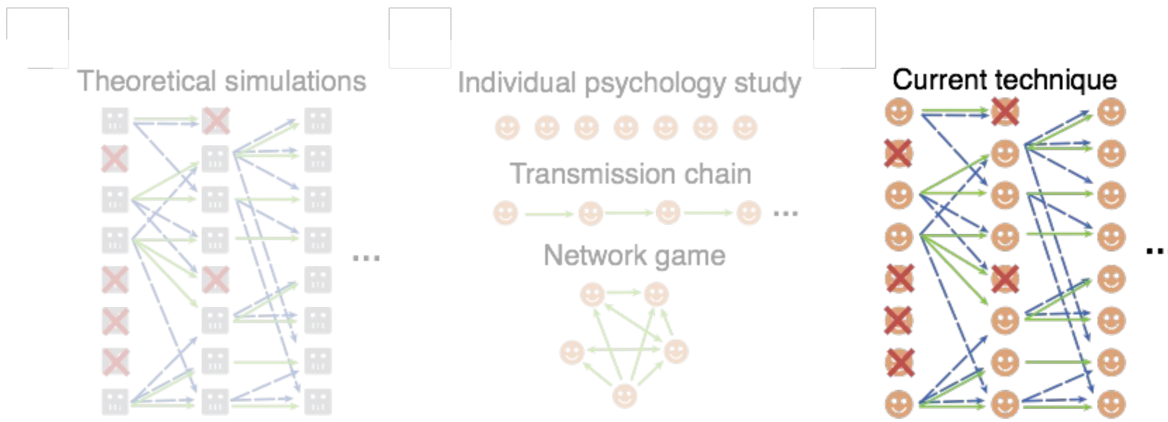
Participant 1546 (G10, RM)
"I started with the first picture, then clicked on picture two. Then back to one, then picture three. Then back to one, to picture 4 and so on. After i was done with the first picture, I started with the second and went second picture to three, second to four, and so on. I did this till I was at the end."

Gnome sort



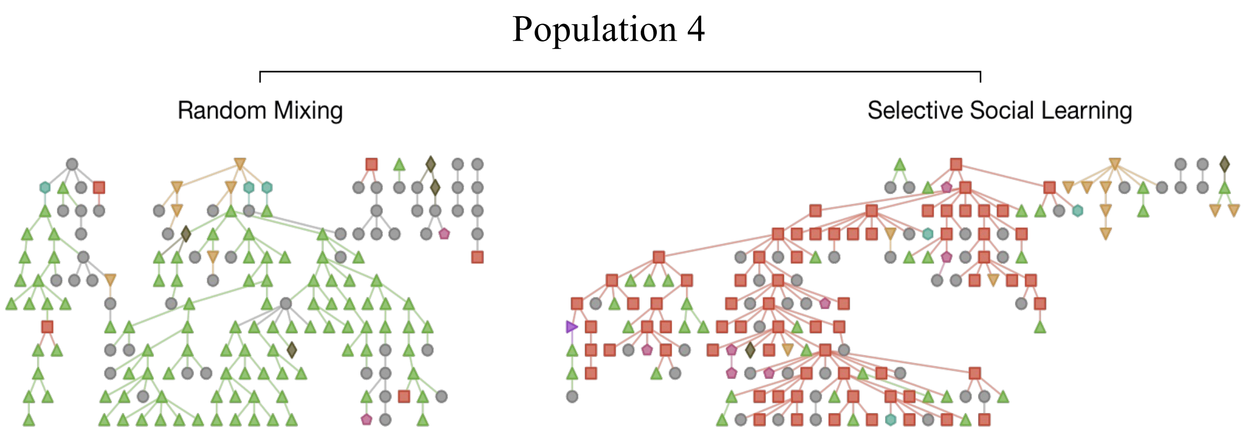
Participant 1281 (G8, SSL)
"Begin with comparing the second picture to the picture on the left. Once it stops moving, move on to the third picture, and compare it with the pictures to the left until it stops moving. Then, move on to the fourth picture and compare it with the pictures to the left. Then, move on to the fifth picture and compare it with the pictures to the left. Finally, choose the sixth picture and compare it with the pictures to the left. Once the sixth picture stops moving, the pictures should be in numerical order."

Evolutionary simulations with people



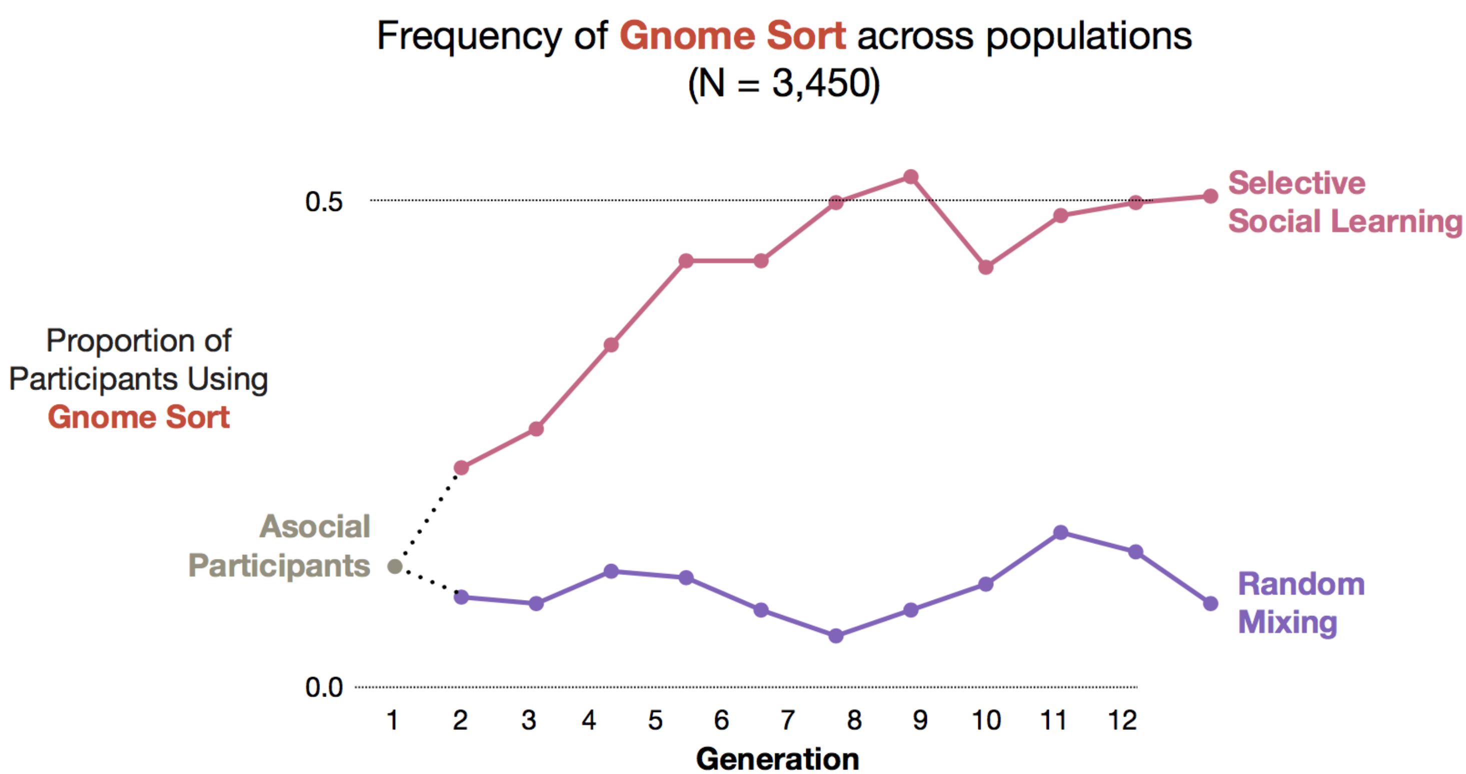
Manipulate *selection*: whether scores are visible
(10 populations of 12 generations of 15 people)

Results


Population 4
Random Mixing
Selective Social Learning


Gnome sort
0.837
Selection sort
0.555

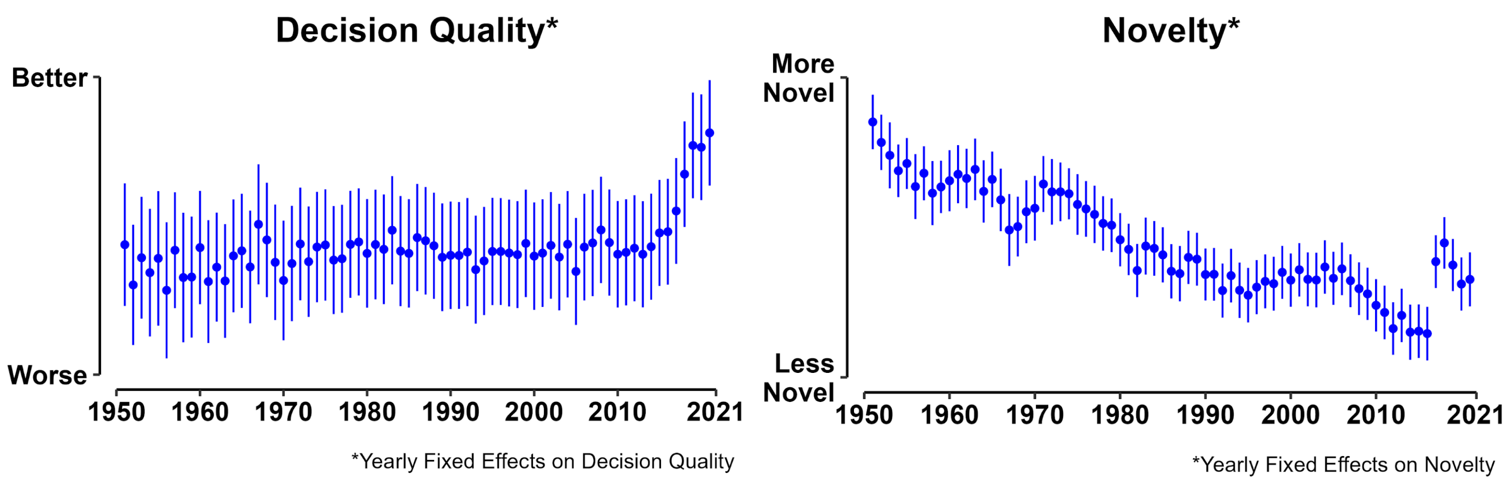
Selection has a substantial effect



(Thompson, van Opheusden, Sumers & Griffiths, 2022)


ALPHAGO
01:59:50
LEE SEDOL
AlphaGo
Lee Sedol

Learning from AI



(Shin, Kim, van Opheusden, & Griffiths, 2023)

Human computational problems

1. Humans have limited time

2. Humans have limited computation

3. Humans have limited communication

What math do we need for understanding human minds?

Human computational problems

1. Humans have limited time

Bayesian inference, metalearning

2. Humans have limited computation

resource rationality

3. Humans have limited communication

distributed computation, evolution

What math do we need for understanding human minds?

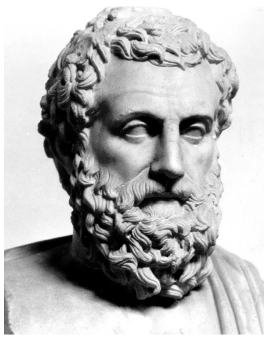
Implications for AI

- We shouldn't expect the intelligence produced by machines to look like that of people...
- ...unless they are solving the same kind of problems
  - e.g., autonomous, low-compute, low-bandwidth settings
- But that doesn't mean we can't learn from humans
  - rapid learning, efficient compute, and using language and teaching are all nice to have
  - the difference is that humans *need* to have them

Conclusions

- Human computational problems reflect three human limitations, potentially distinct from AI
  - limited time, computation, and communication
- Studying these limitations uses three formalisms distinctly relevant to cognitive science
  - Bayes, resource rationality, and distributed computation
- Changing the reference class changes how we characterize what is unique about humans

What makes humans unique?



"animals that think"

Computational Cognitive Science Lab
http://cocosci.princeton.edu/

Credits

Erin Grant
Tom McCoy
Fred Callaway
Paul Krueger
Falk Lieder
Sayan Gul
Evan Russek
Dan Acosta-Kane
Marcelo Mattar
Bill Thompson
Bas van Opheusden
Ted Sumers