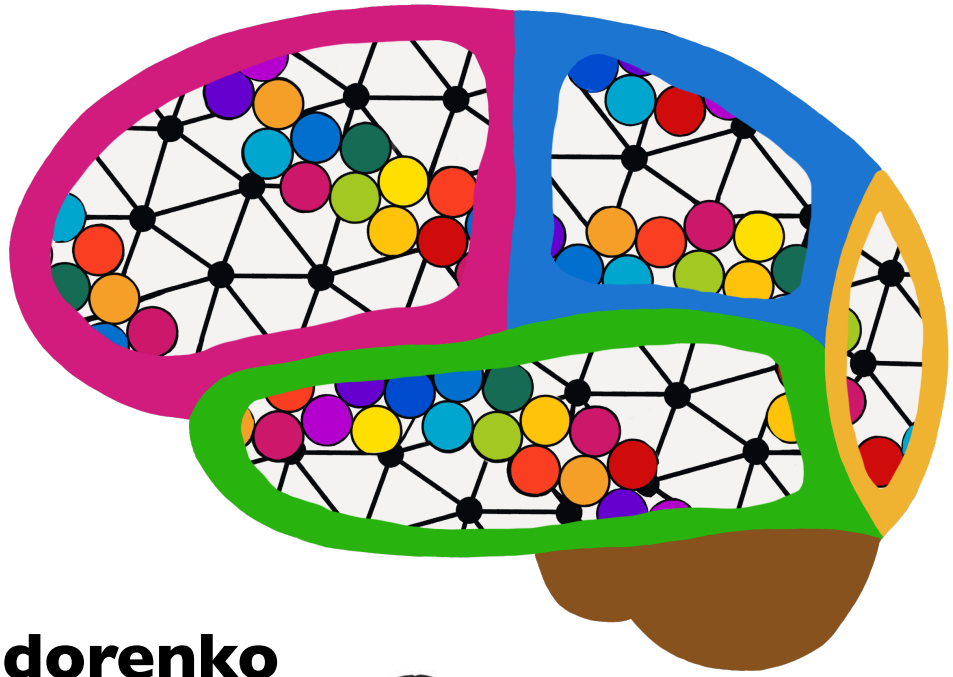
Language and thought in the human brain



LLMs, Cognitive Science, Linguistics, and Neuroscience

Berkeley, CA

Ev Fedorenko

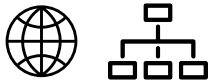
February 6, 2025

Art by Laura Bundesen

How to find us:
evlab.mit.edu
@ev_fedorenko
@evfedorenko.bsky.social

Concepts — the building blocks of thought

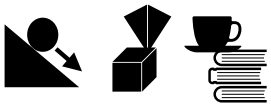
World knowledge + commonsense reasoning



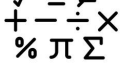
Spatial navigation

Numerical cognition

Physical reasoning

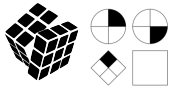


Mathematical reasoning



Music

Abstract problem solving



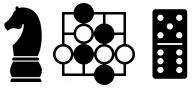
Social reasoning / Theory of mind

Art

Logic

Executive functions

Intellectual / strategy games

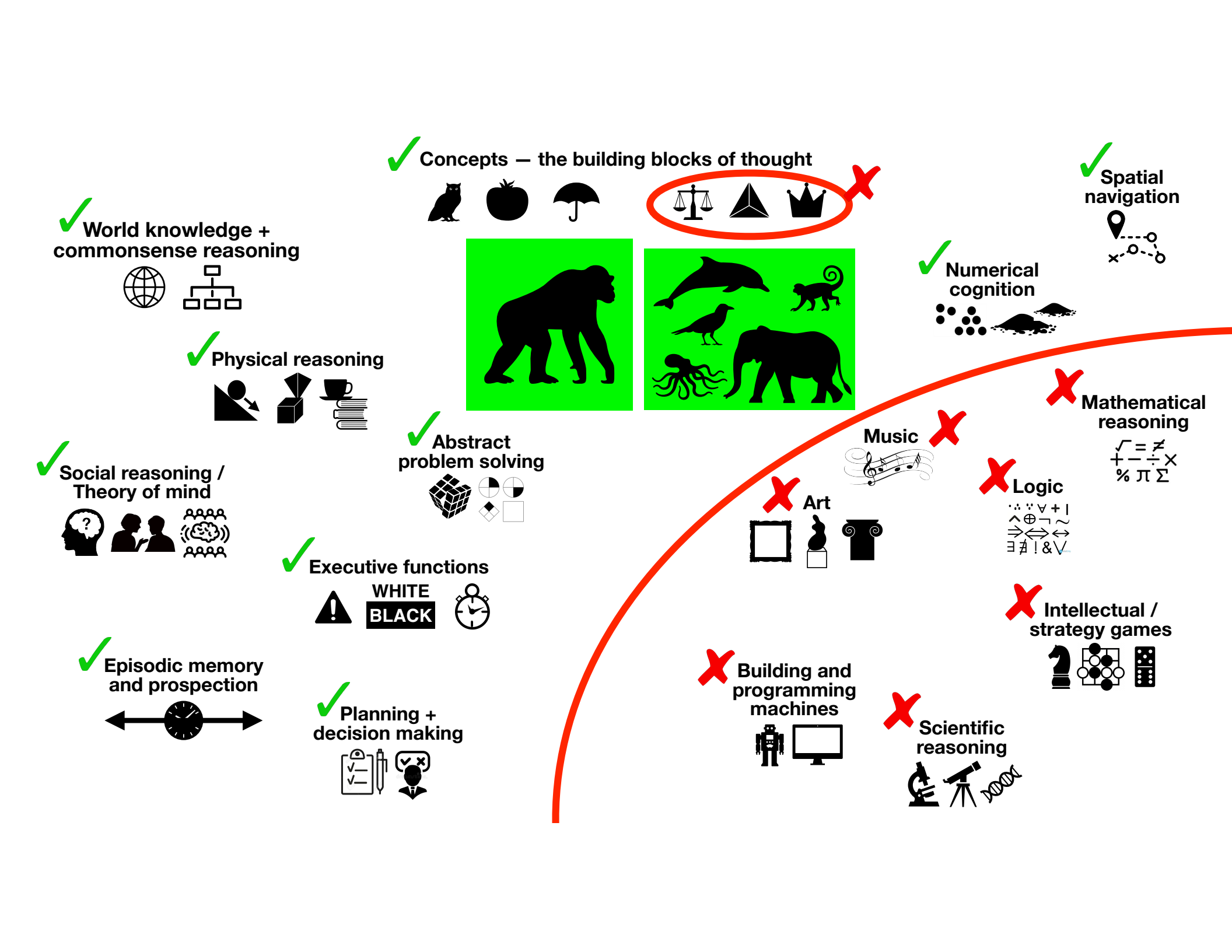


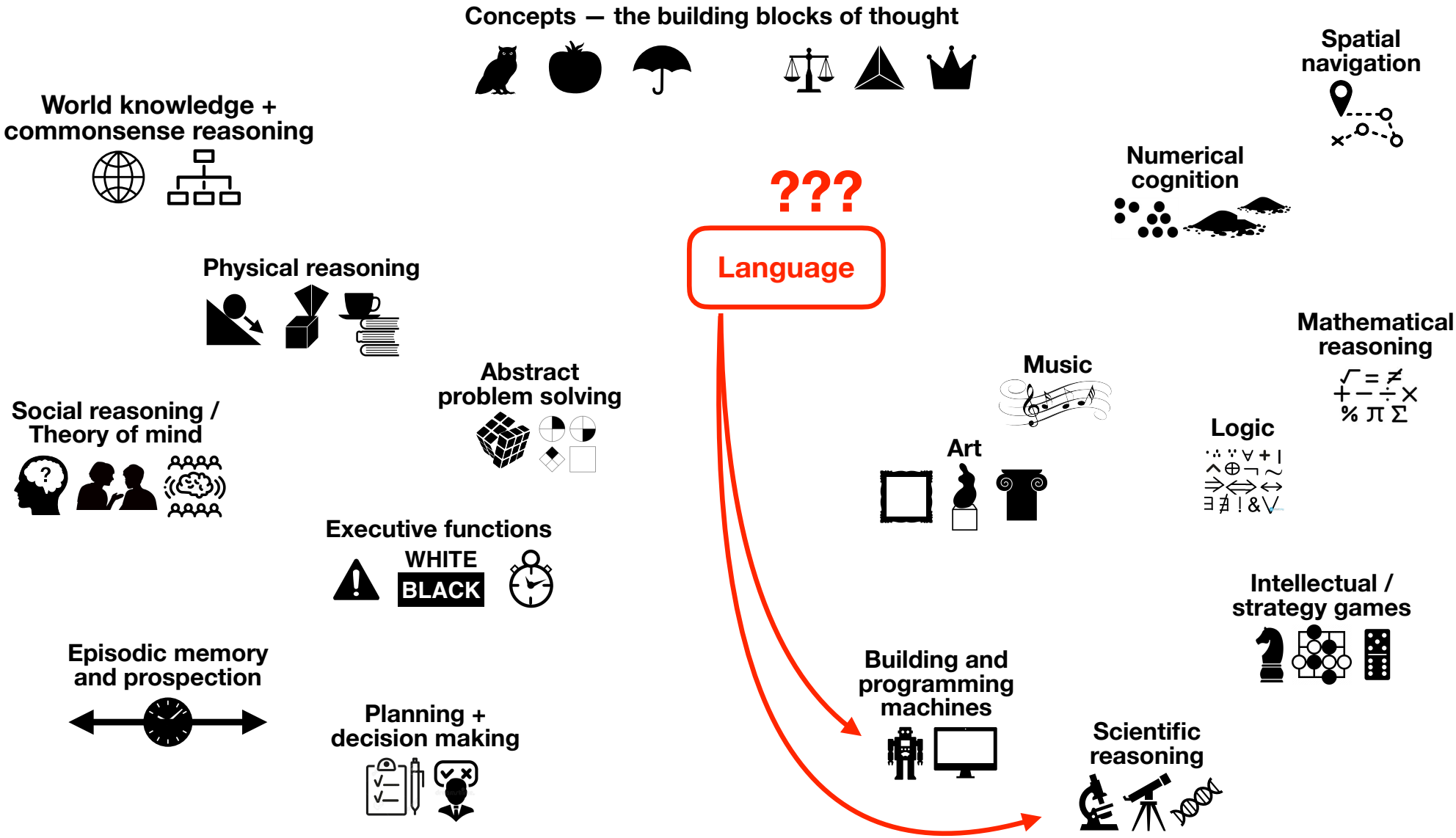
Episodic memory and prospection

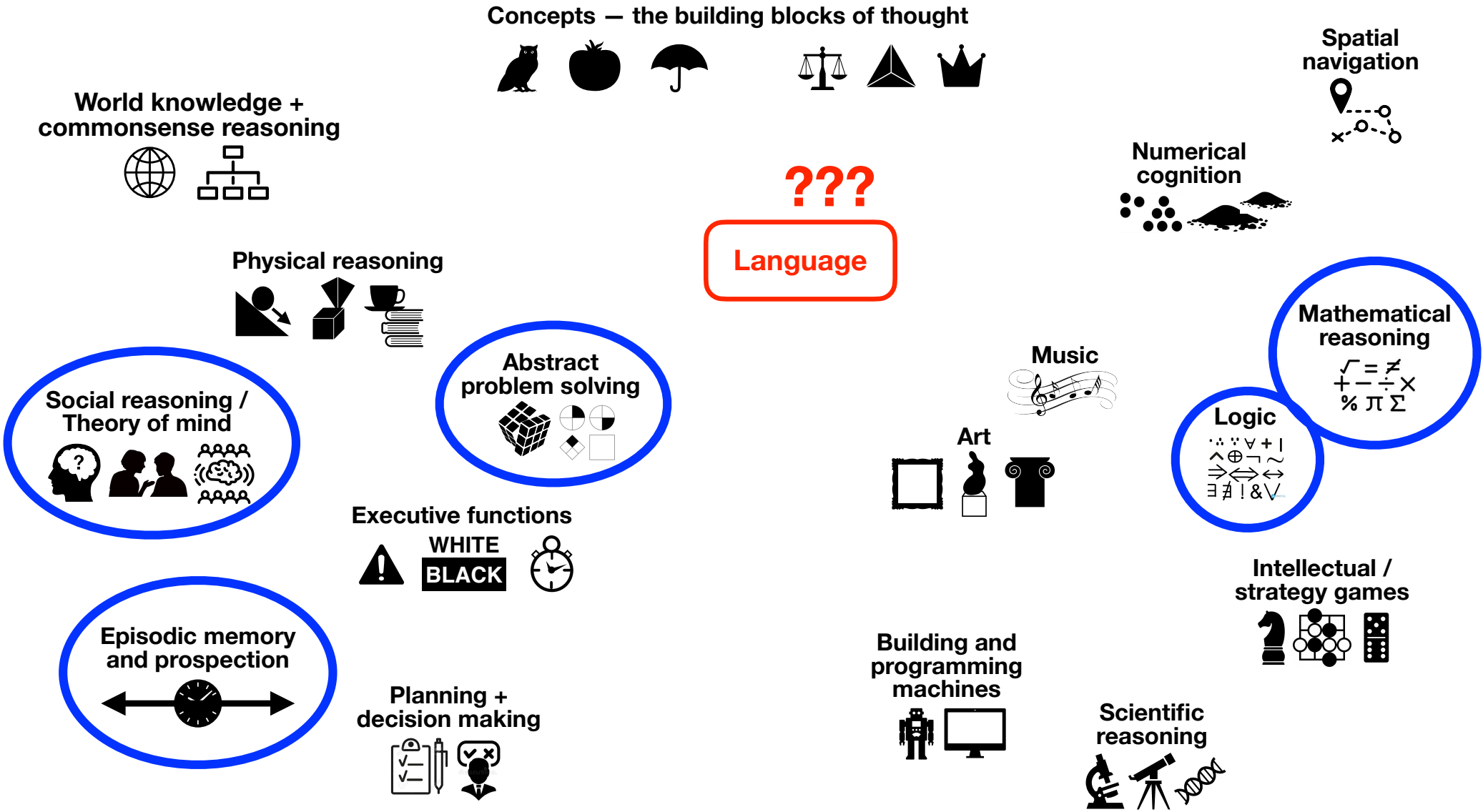
Building and programming machines

Planning + decision making

Scientific reasoning


Concepts — the building blocks of thought
World knowledge + commonsense reasoning
Physical reasoning
Social reasoning / Theory of mind
Abstract problem solving
Executive functions
WHITE
BLACK
Episodic memory and prospection
Planning + decision making
Spatial navigation
Numerical cognition
Mathematical reasoning
Music
Logic
Art
Intellectual / strategy games
Building and programming machines
Scientific reasoning

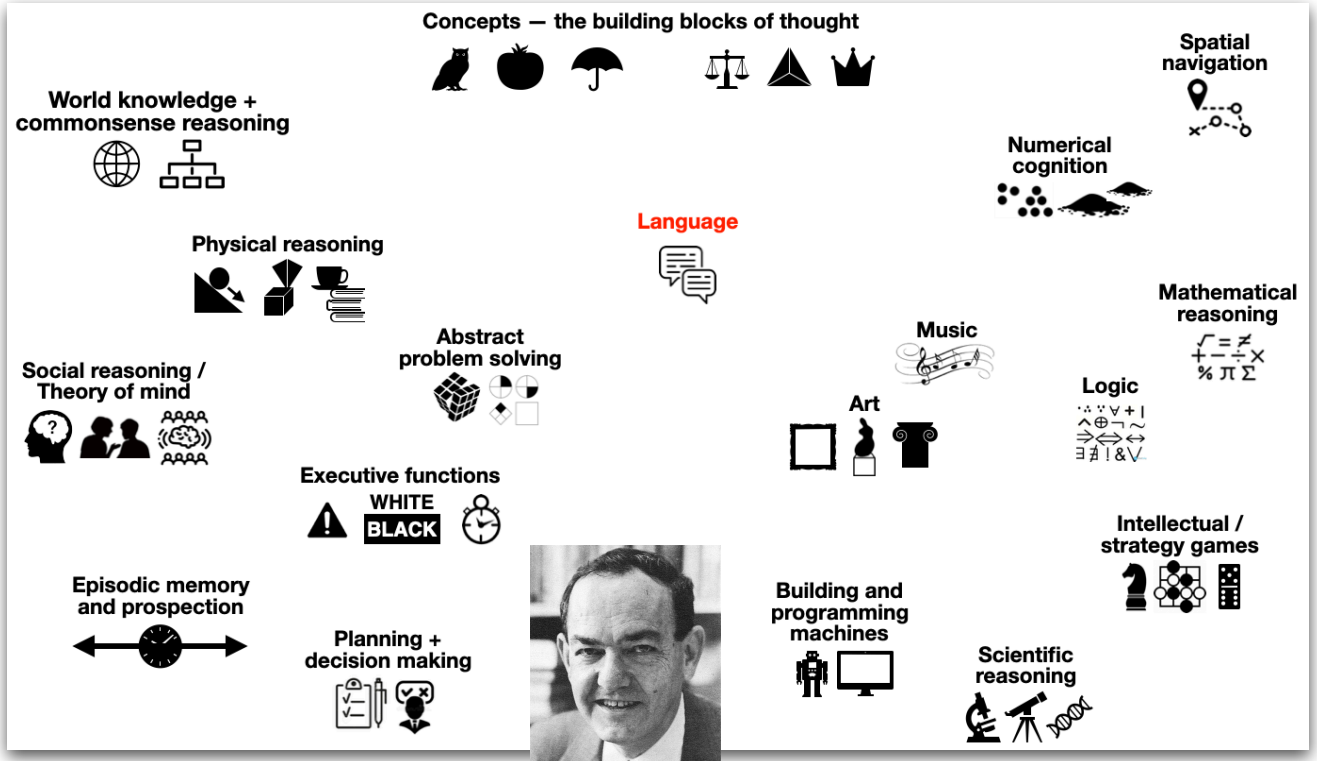

Concepts — the building blocks of thought
World knowledge + commonsense reasoning
Spatial navigation
Numerical cognition
???
Language
Physical reasoning
Mathematical reasoning
Abstract problem solving
Music
Social reasoning / Theory of mind
Art
Logic
Executive functions
WHITE
BLACK
Intellectual / strategy games
Episodic memory and prospection
Building and programming machines
Planning + decision making
Scientific reasoning


Concepts — the building blocks of thought
World knowledge + commonsense reasoning
Spatial navigation
Numerical cognition
???
Language
Physical reasoning
Mathematical reasoning
Music
Abstract problem solving
Social reasoning / Theory of mind
Art
Logic
Executive functions
WHITE
BLACK
Intellectual / strategy games
Episodic memory and prospection
Building and programming machines
Planning + decision making
Scientific reasoning

"The systems of thought … use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts."

"A substantial part of what we call thinking is simply linguistic manipulation, so if there is a severe deficit of language, there will be a severe deficit of thought."

Noam Chomsky



"nearly decomposable systems"

Herbert Simon

Today:

1. The human **language system**: Introduction and key properties

2. The relationship between **language and thought** in humans.
   The **structure** of human thought.

3. **Neural network LMs**—a new model organism for language research

Today:

1. The human **language system**: Introduction and key properties

2. The relationship between **language and thought** in humans.

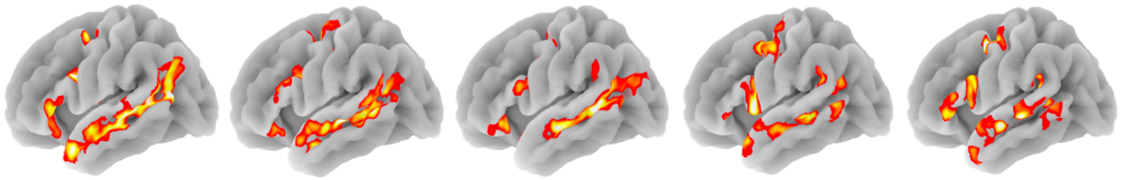
3. **Neural network LMs**—a new model organism for language research

The language system

Language > **Perceptually matched control**

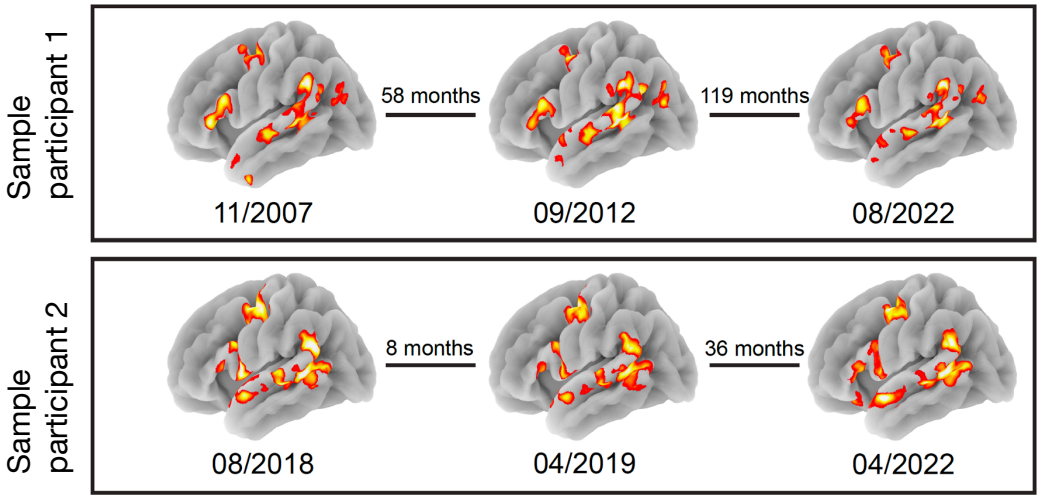
Fedorenko et al. (2010, *J Neurophys*)

Sample individual language maps:



The language system

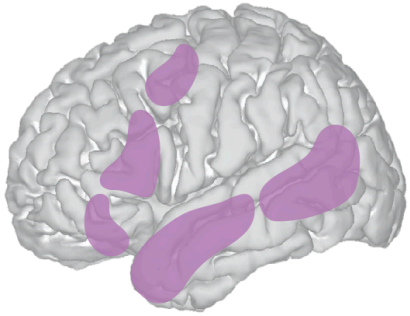
Activations are highly **stable** within individuals **over time**:



Mahowald & Fedorenko (2016, *NeuroImage*);
Lipkin et al. (2022, *Nat Sci Data*)

The language system

- robust response during **comprehension**

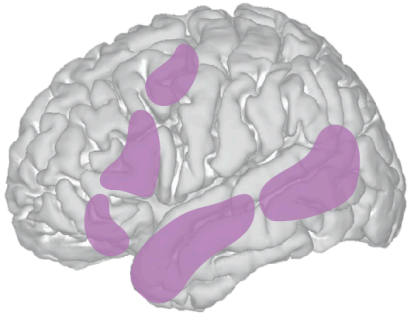


“encoder-decoder”

- robust response during **production**

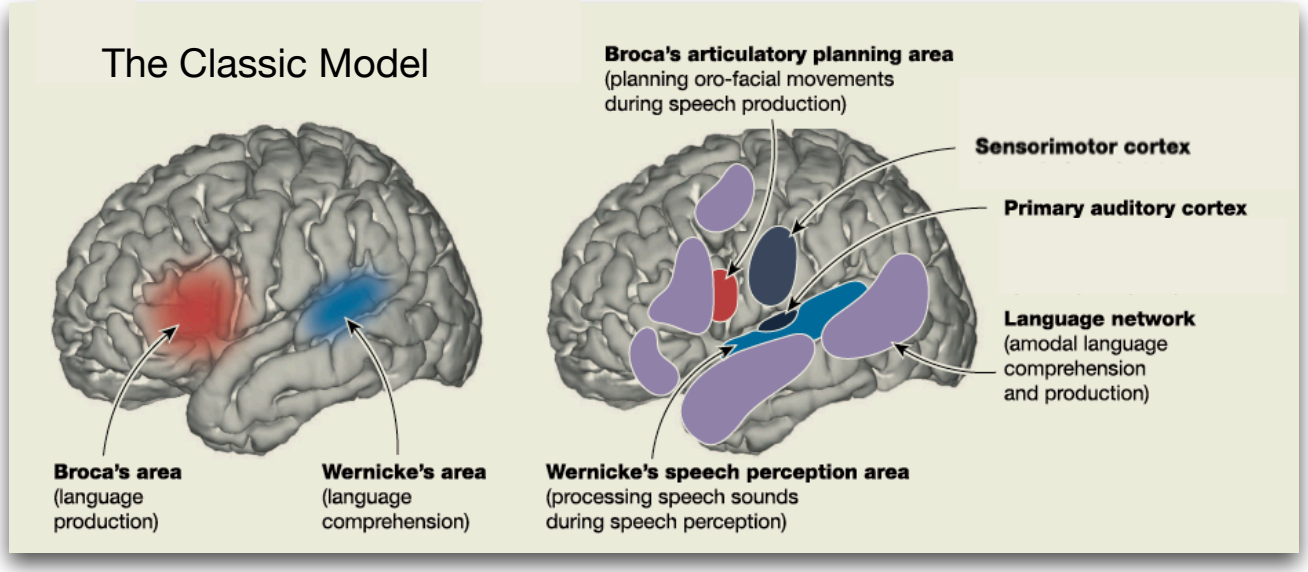
The language system

- robust response during **comprehension**



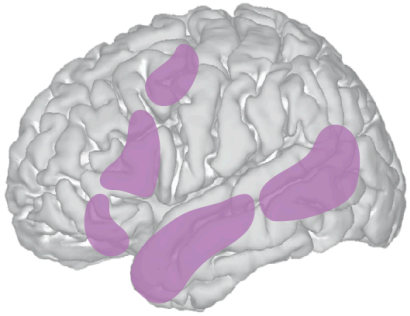
- robust response during **production**

“encoder-decoder”



The language system

- robust response during **comprehension**



“encoder-decoder”

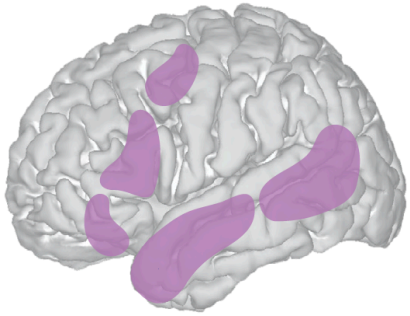
- robust response during **production**

- present and adult-like in topography in **children** (by 3-4y)

Hiersche et al. (2023); Ozernov-Palchik, O’Brien et al. (2024); Olson et al. (in prep.)

The language system

- robust response during **comprehension**



“encoder-decoder”

- robust response during **production**

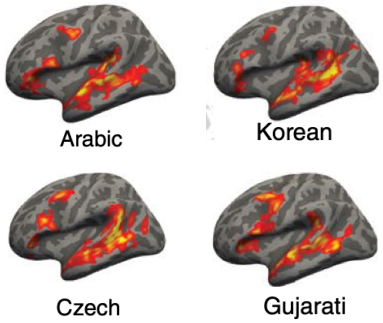
- present and adult-like in topography in **children** (by 3-4y)

Hiersche et al. (2023); Ozernov-Palchik, O’Brien et al. (2024); Olson et al. (in prep.)

- similar **across languages** across and within speakers

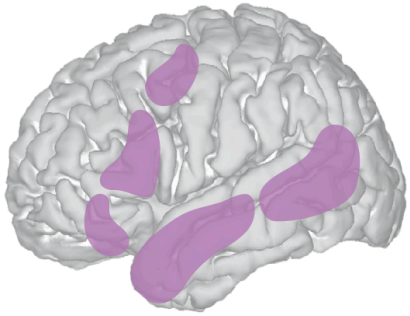
Malik-Moraleda, Ayyash et al. (2022); Malk-Moraleda, Jouravlev et al. (2024)

Saima Malik-Moraleda



The language system

- robust response during **comprehension**



“encoder-decoder”

- robust response during **production**

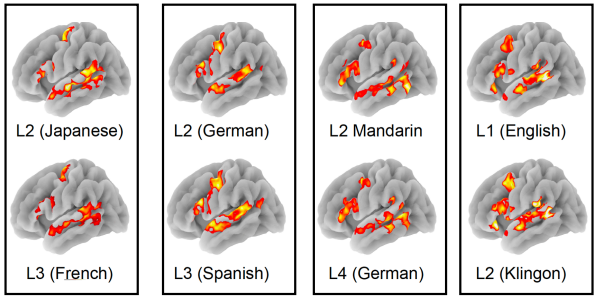
- present and adult-like in topography in **children** (by 3-4y)

Hiersche et al. (2023); Ozernov-Palchik, O’Brien et al. (2024); Olson et al. (in prep.)

- similar **across languages** across and within speakers

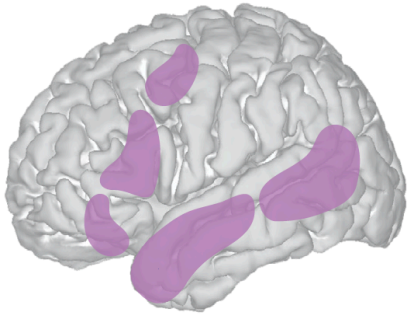
Malik-Moraleda, Ayyash et al. (2022); Malk-Moraleda, Jouravlev et al. (2024)

Saima Malik-Moraleda



The language system

- robust response during **comprehension**



“encoder-decoder”

- robust response during **production**

- present and adult-like in topography in **children** (by 3-4y)

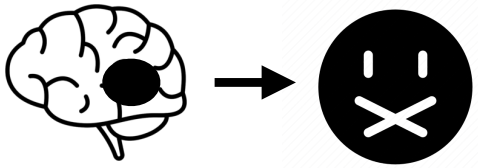
Hiersche et al. (2023); Ozernov-Palchik, O’Brien et al. (2024); Olson et al. (in prep.)

- similar **across languages** across and within speakers

Malik-Moraleda, Ayyash et al. (2022); Malk-Moraleda, Jouravlev et al. (2024)

- **causally** important for language function

a large body of work on aphasia



The language system

What **linguistic computations** does the language system support?

A **red-haired woman** is playing with her **dog** ...

The language system

What **linguistic computations** does the language system support?

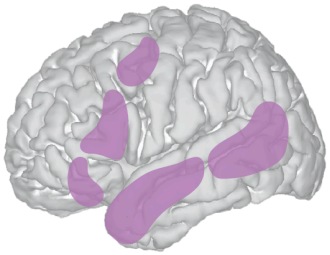
A **red-haired woman** is playing with her **dog** ...

The language system

What **linguistic computations** does the language system support?

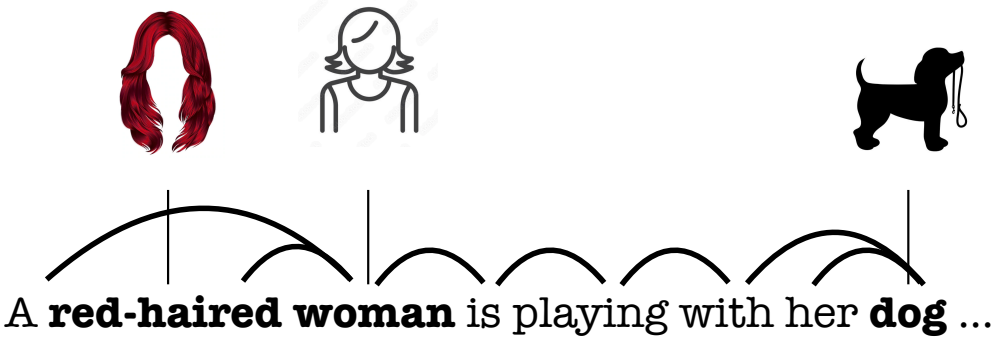
The language system supports computations that are related to:

- word retrieval
- syntactic structure building
- semantic composition



Fedorenko *et al. (2010 J Neurophys);*
Fedorenko *et al. (2016 PNAS);*
Shain, Kean *et al. (2024 JOCN);*
Kauf *et al. (2024 bioRiv)*

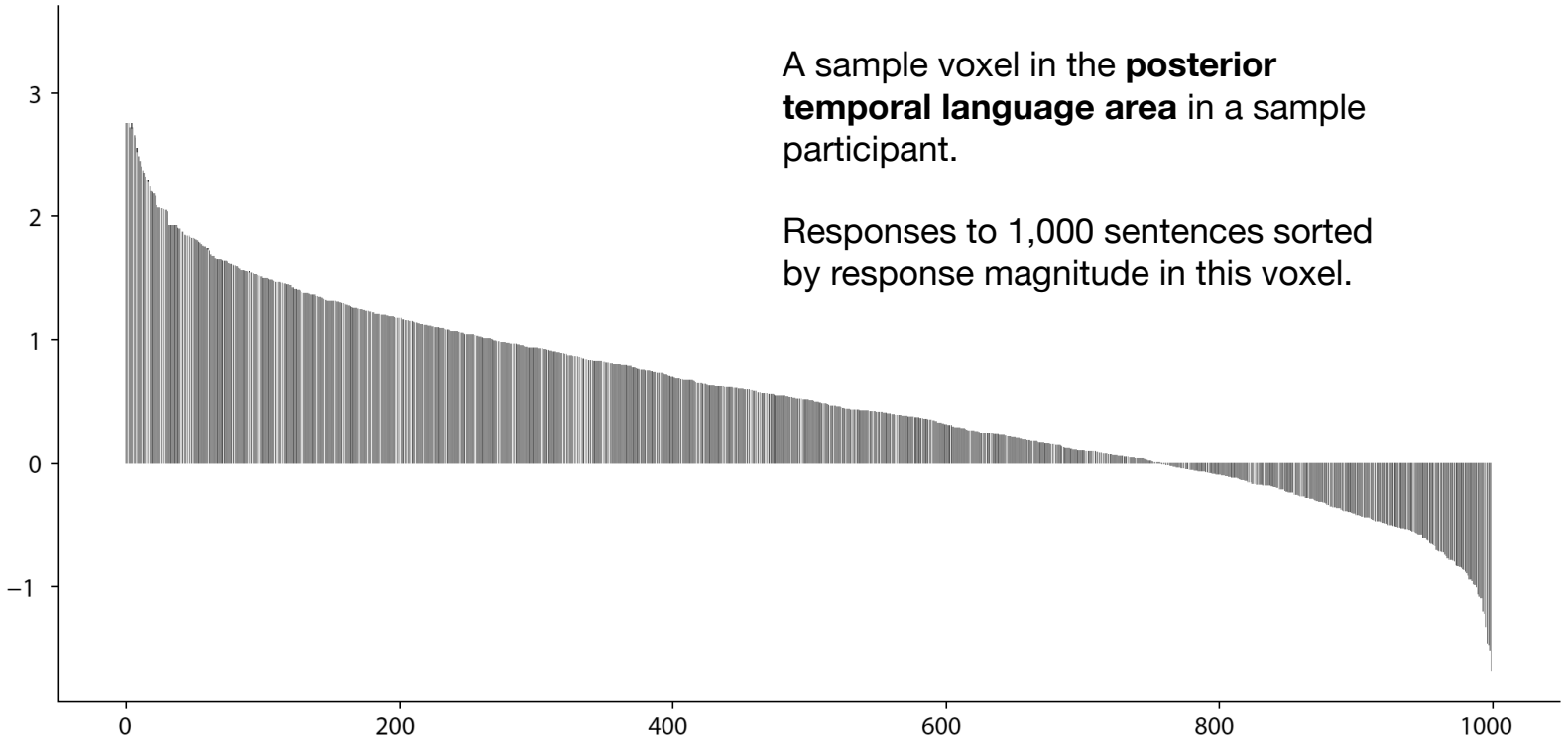
Shain, Blank *et al. (2020 Np'logia);*
Shain *et al. (2023 JNeuro)*



The language system

What **linguistic computations** does the language system support?

Many language voxels do not show semantic tuning.



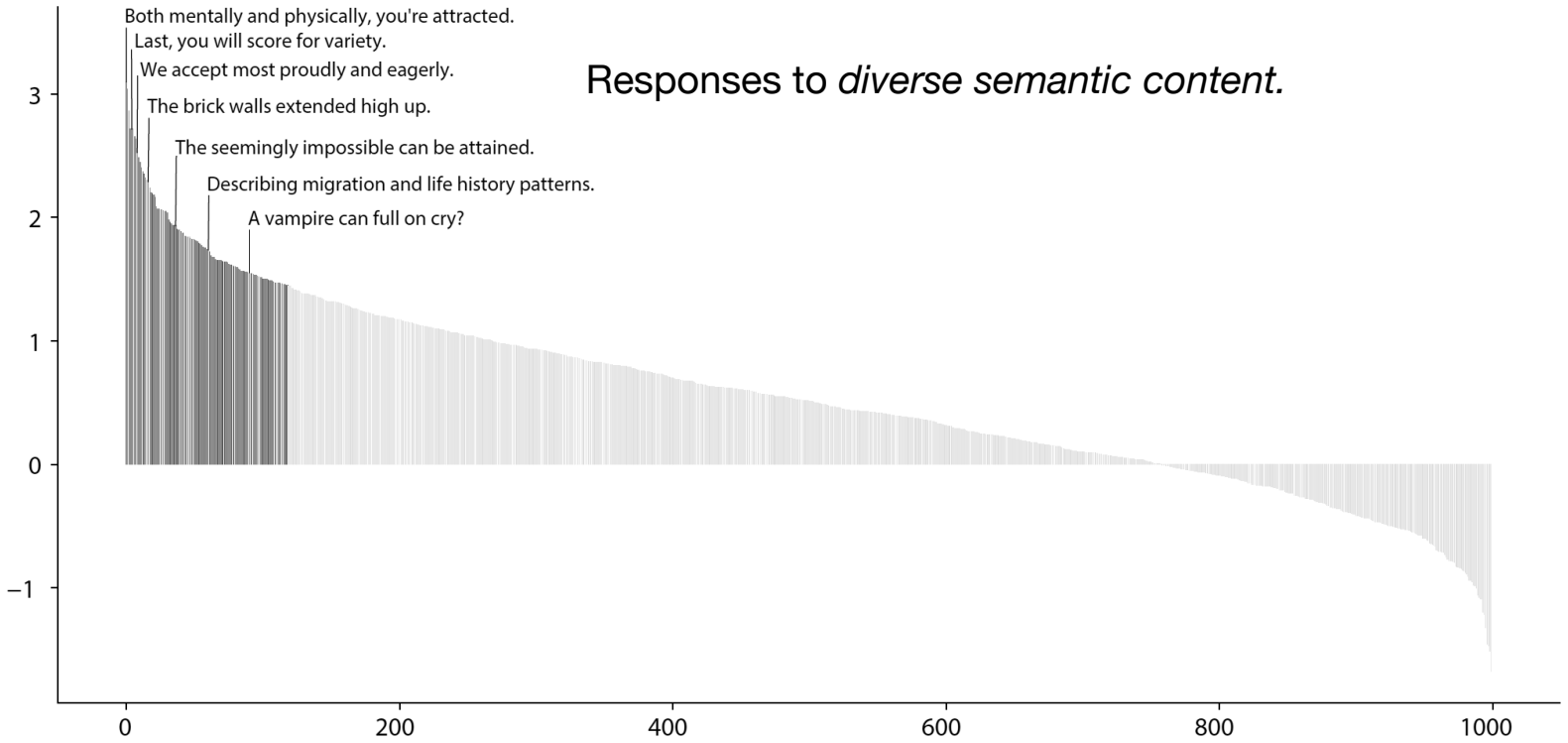
A sample voxel in the **posterior temporal language area** in a sample participant.

Responses to 1,000 sentences sorted by response magnitude in this voxel.

The language system

What **linguistic computations** does the language system support?

Many language voxels do not show semantic tuning.



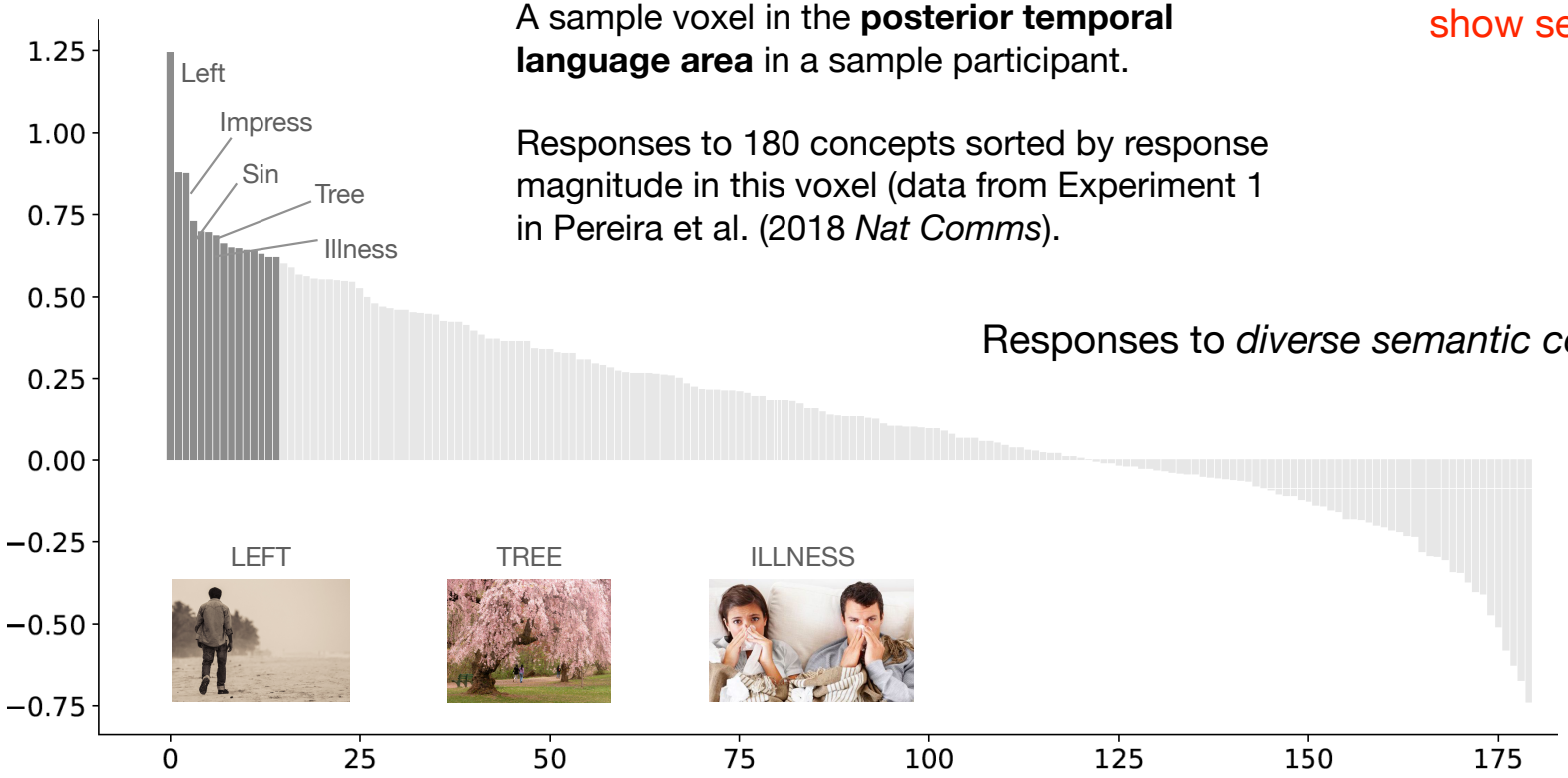
The language system

What **linguistic computations** does the language system support?

A sample voxel in the **posterior temporal language area** in a sample participant.

Responses to 180 concepts sorted by response magnitude in this voxel (data from Experiment 1 in Pereira et al. (2018 *Nat Comms*).

Many language voxels do not show semantic tuning.



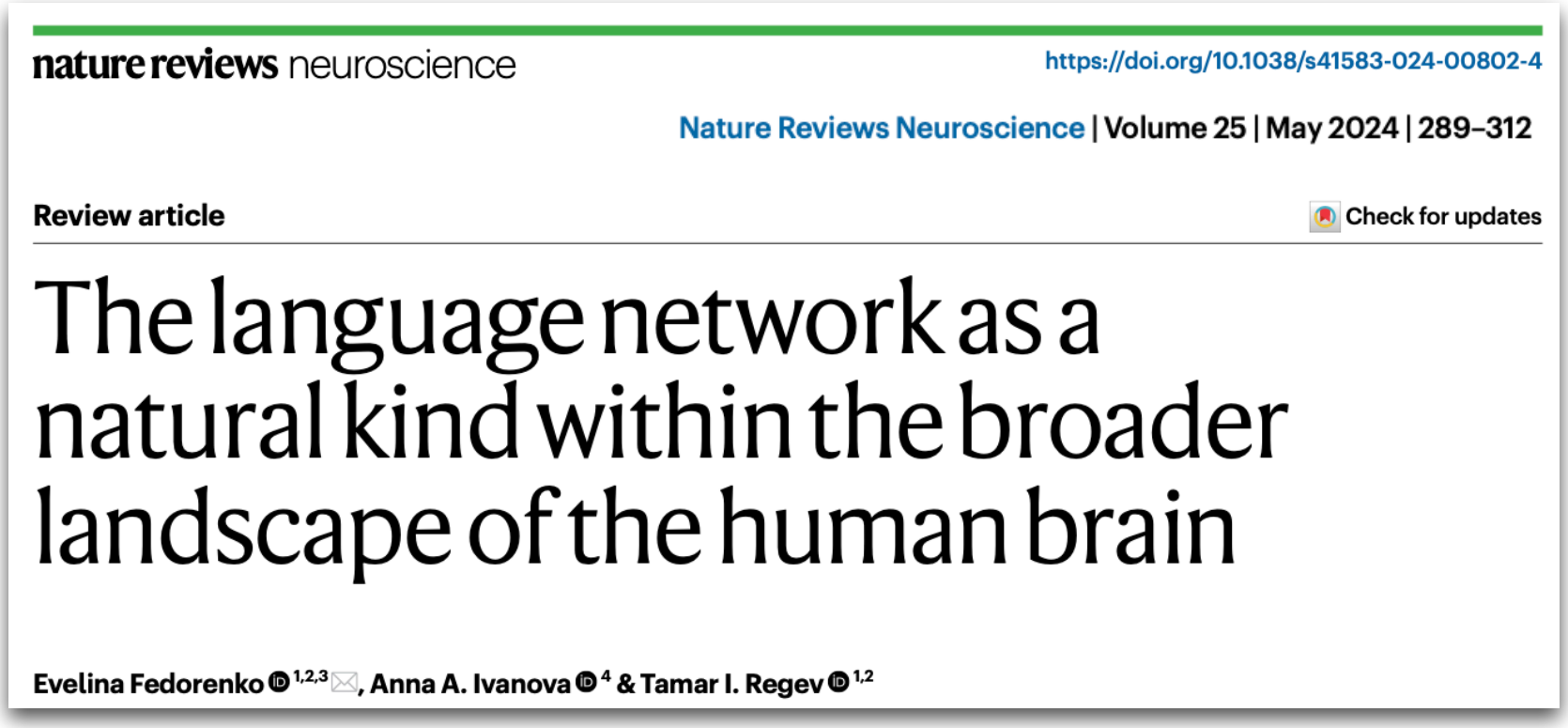
Responses to *diverse semantic content.*


Maria Ryskina


Greta Tuckute

The language system

To learn more:



nature reviews neuroscience

https://doi.org/10.1038/s41583-024-00802-4

Nature Reviews Neuroscience | Volume 25 | May 2024 | 289–312

Review article

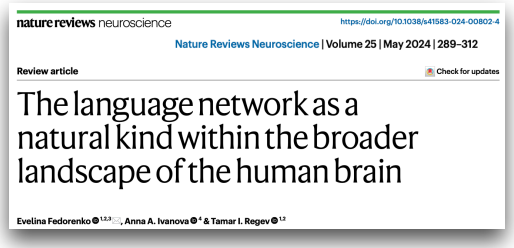
Check for updates

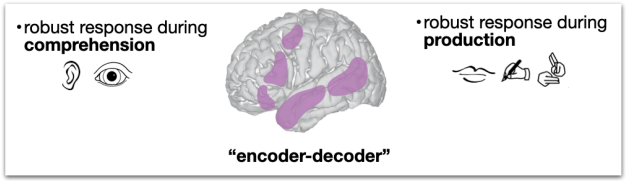
The language network as a natural kind within the broader landscape of the human brain

Evelina Fedorenko [1,2,3] ✉, Anna A. Ivanova [4] & Tamar I. Regev [1,2]

Today:

1 The human **language system**: Introduction and key properties





2 The relationship between **language and thought** in humans.

3 **Neural network LMs**—a new model organism for language research

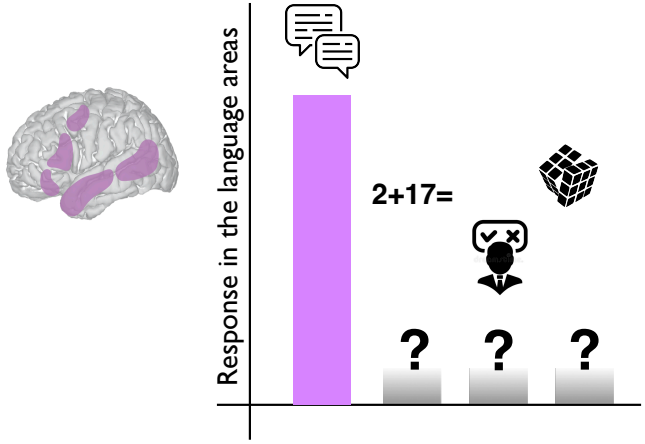
"The systems of thought … use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts."

"A substantial part of what we call thinking is simply linguistic manipulation, so if there is a severe deficit of language, there will be a severe deficit of thought."

Noam Chomsky

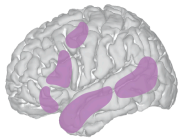
How do we test this hypothesis?

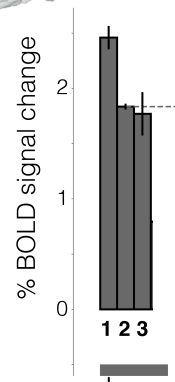
Is the language system engaged when we think?



Language vs. thought (and other non-linguistic functions)

Language areas are highly **selective** relative to diverse **non-linguistic inputs and tasks**.





34 experiments, 79 conditions;
780 participants across 1,051 scanning sessions

Language:
1 - Auditory sentence comprehension (listening)
2 - Visual sentence comprehension (reading)
3 - Spoken sentence production (speaking)

Speech perception
(syllables+nonwords)

Speech perception
(Foreign lang)

Letter perception
(Consonant strings+nonwords)

Articulation

Music perception

Numerical cognition

Executive functions

Fedorenko *et al. (2011 PNAS);*
Fedorenko *et al. (2012 Curr Blo);*
Pritchett *et al. (2018 J Neurophys);*
Jouravlev *et al. (2019 Np'logia);*
Shain, Paunov, Chen *et al. (2023 Cer Cort);*
Ivanova *et al. (2020 eLife);*
Benn, Ivanova *et al. (2023 Cer Cort);*
Chen *et al. (2023 Cer Cort), inter alia*

Object semantics

Scene semantics

Event semantics

Face perception

Hand+body perception

Gesture perception

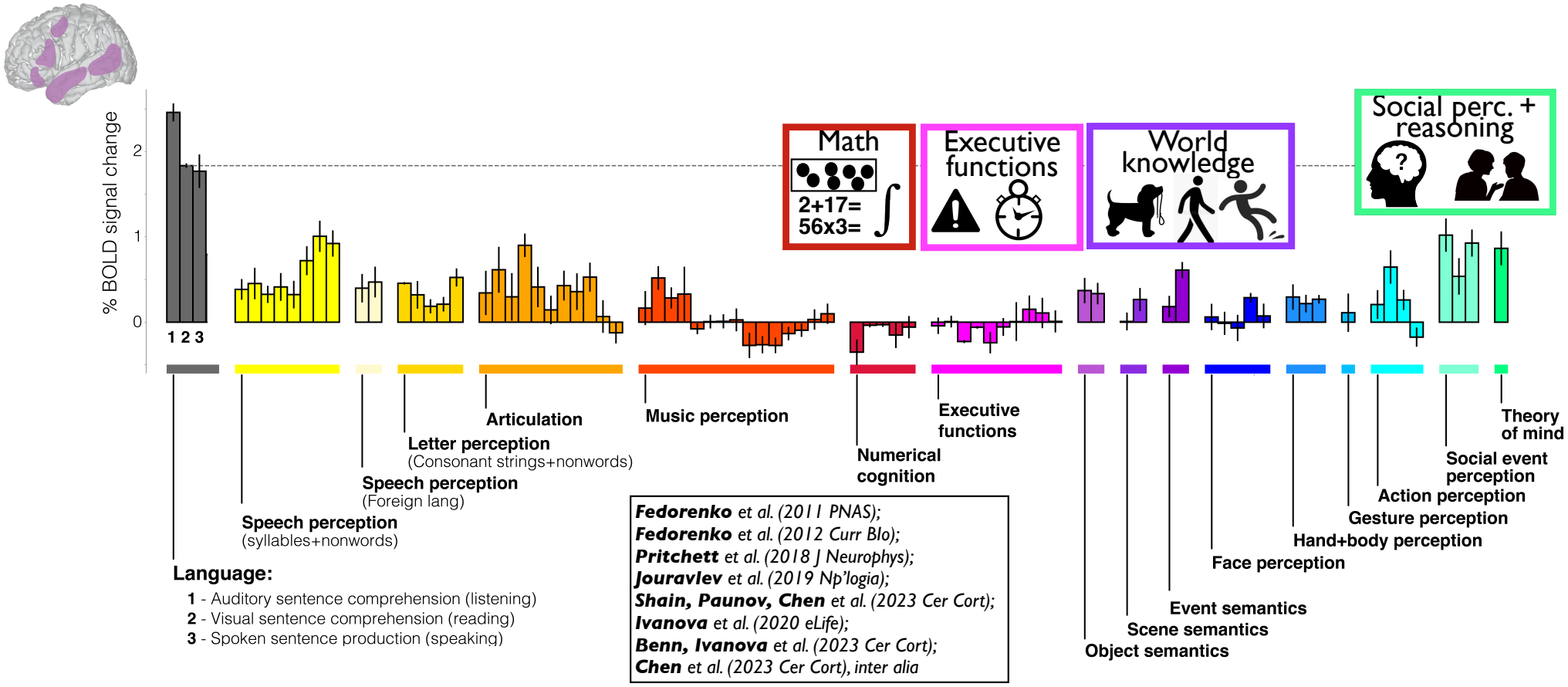
Action perception

Social event perception

Theory of mind

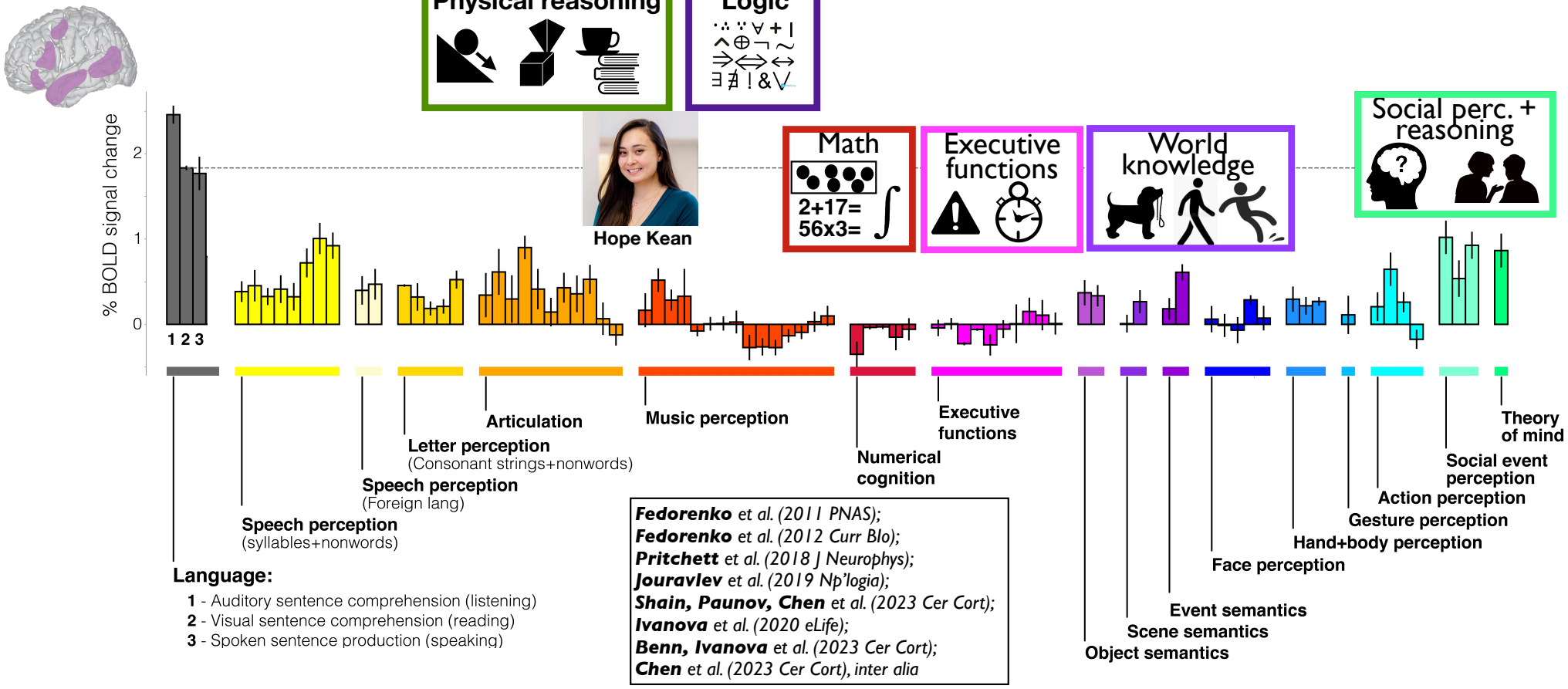
Language vs. thought (and other non-linguistic functions)

Language areas **show little/no response** when we engage in diverse thought-related activities.



Language vs. thought (and other non-linguistic functions)

Language areas **show little/no response** when we engage in diverse thought-related activities.



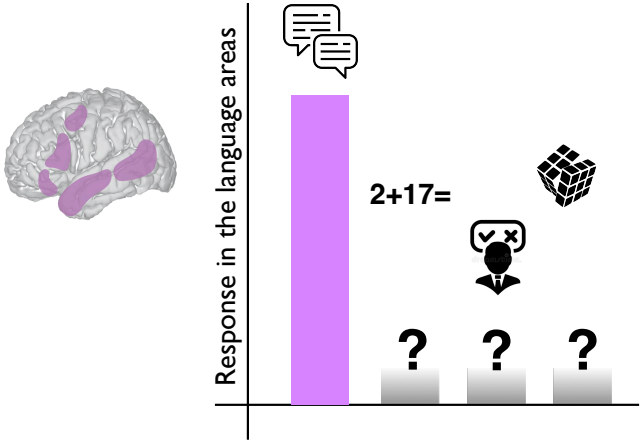
"The systems of thought … use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts."

"A substantial part of what we call thinking is simply linguistic manipulation, so if there is a severe deficit of language, there will be a severe deficit of thought."

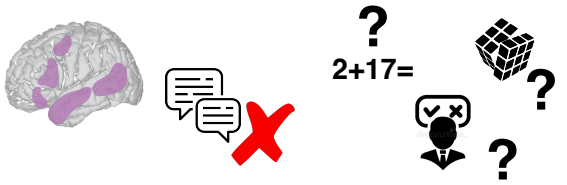
Noam Chomsky

How do we test this hypothesis?

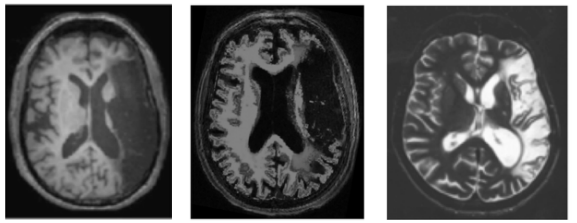
Is the language system engaged when we think?



Can we think without language?



Sample lesions of patients with global aphasia:



Language vs. thought

✓ **World knowledge + commonsense reasoning**

✓ **Physical reasoning**

✓ **Social reasoning / Theory of mind**

✓ **Abstract problem solving**

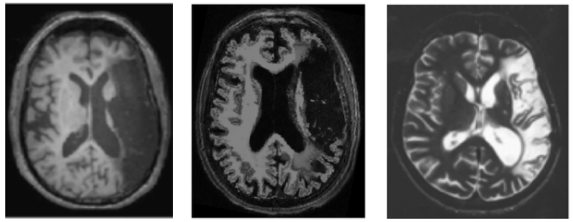
✓ **Executive functions**

WHITE
BLACK

✓ **Episodic memory and prospection**

✓ **Planning + decision making**

Sample lesions of patients with global aphasia:



✓ **Spatial navigation**

✓ **Numerical cognition**

✓ **Music**

✓ **Mathematical reasoning**

✓ **Logic**

✓ **Art**

✓ **Scientific reasoning**

✓ **Intellectual / strategy games**

Rosemary Varley
(UCL)

Anya Ivanova
(Georgia Tech)

Hope Kean
(MIT)

Fedorenko & Varley *(2016 ANYAS);*
Woolgar *et al. (2018 Nat Hum Beh);*
Ivanova *et al. (2021 NOL);*
Chen *et al. (2022 Cer Cort);*
Benn, Ivanova *et al. (2023 Cer Cort), inter alia*

"The systems of thought … use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts."

"A substantial part of what we call thinking is simply linguistic manipulation, so if there is a severe deficit of language, there will be a severe deficit of thought."

Noam Chomsky

Is the language system engaged when we think?

No

Can we think without language?

Yes

Perspective

Nature | Vol 630 | 20 June 2024 | 575

Language is primarily a tool for communication rather than thought

https://doi.org/10.1038/s41586-024-07522-w

Received: 15 February 2023

Accepted: 3 May 2024

Published online: 19 June 2024

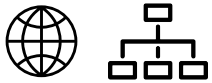
Check for updates

Evelina Fedorenko[1,2]✉, Steven T. Piantadosi[3] & Edward A. F. Gibson[1]

Language is a defining characteristic of our species, but the function, or functions, that it serves has been debated for centuries. Here we bring recent evidence from neuroscience and allied disciplines to argue that in modern humans, language is a tool for communication, contrary to a prominent view that we use language for thinking. We begin by introducing the brain network that supports linguistic ability in humans. We then review evidence for a double dissociation between language and thought, and discuss several properties of language that suggest that it is optimized for communication. We conclude that although the emergence of language has unquestionably transformed human culture, language does not appear to be a prerequisite for complex thought, including symbolic thought. Instead, language is a powerful tool for the transmission of cultural knowledge; it plausibly co-evolved with our thinking and reasoning capacities, and only reflects, rather than gives rise to, the signature sophistication of human cognition.

The structure of thought

World knowledge + commonsense reasoning

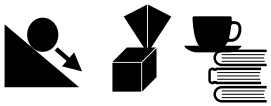


Concepts — the building blocks of thought

Spatial navigation

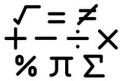
Numerical cognition

Physical reasoning

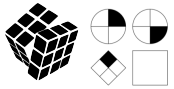


Language

Mathematical reasoning



Abstract problem solving

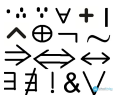


Music

Social reasoning / Theory of mind

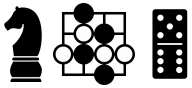
Art

Logic



Executive functions

Intellectual / strategy games



Episodic memory and prospection

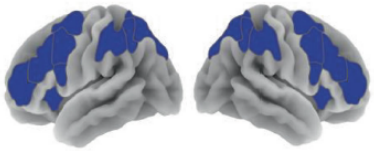
Planning + decision making

Building and programming machines

Scientific reasoning

The structure of thought

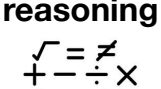
Multiple demand network



e.g., Duncan (2010); Assem et al. (2020)

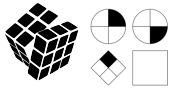
Numerical cognition

Mathematical reasoning



Logic

Abstract problem solving

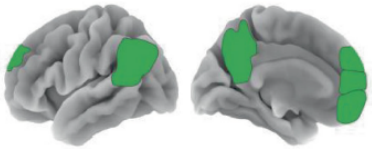


Executive functions

Building and programming machines

The structure of thought

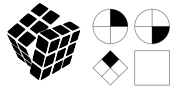
Theory of mind network



e.g., Saxe & Kanwisher (2003)

Social reasoning / Theory of mind

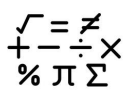
Abstract problem solving



Executive functions

Numerical cognition

Mathematical reasoning



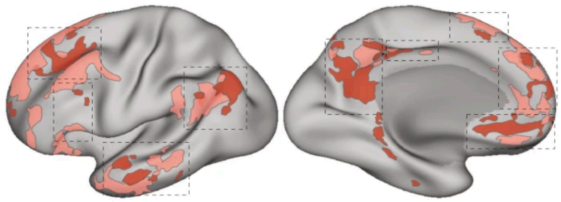
Logic

Building and programming machines

The structure of thought

Default network

Broadly similar areas as the Theory of Mind network, but robustly dissociable within individuals.



Braga & Buckner (2017)

Default

Theory of Mind

Numerical cognition

Mathematical reasoning

Logic

Abstract problem solving

Social reasoning / Theory of mind

Executive functions

WHITE

BLACK

Episodic memory and prospection

Spatial navigation

Building and programming machines

The structure of thought

Fischer et al. (2016)

Physical reasoning

Social reasoning / Theory of mind

Abstract problem solving

Executive functions

WHITE BLACK

Episodic memory and prospection

Spatial navigation

Pramod et al. (in prep.)

Causal reasoning

Numerical cognition

Mathematical reasoning

Logic

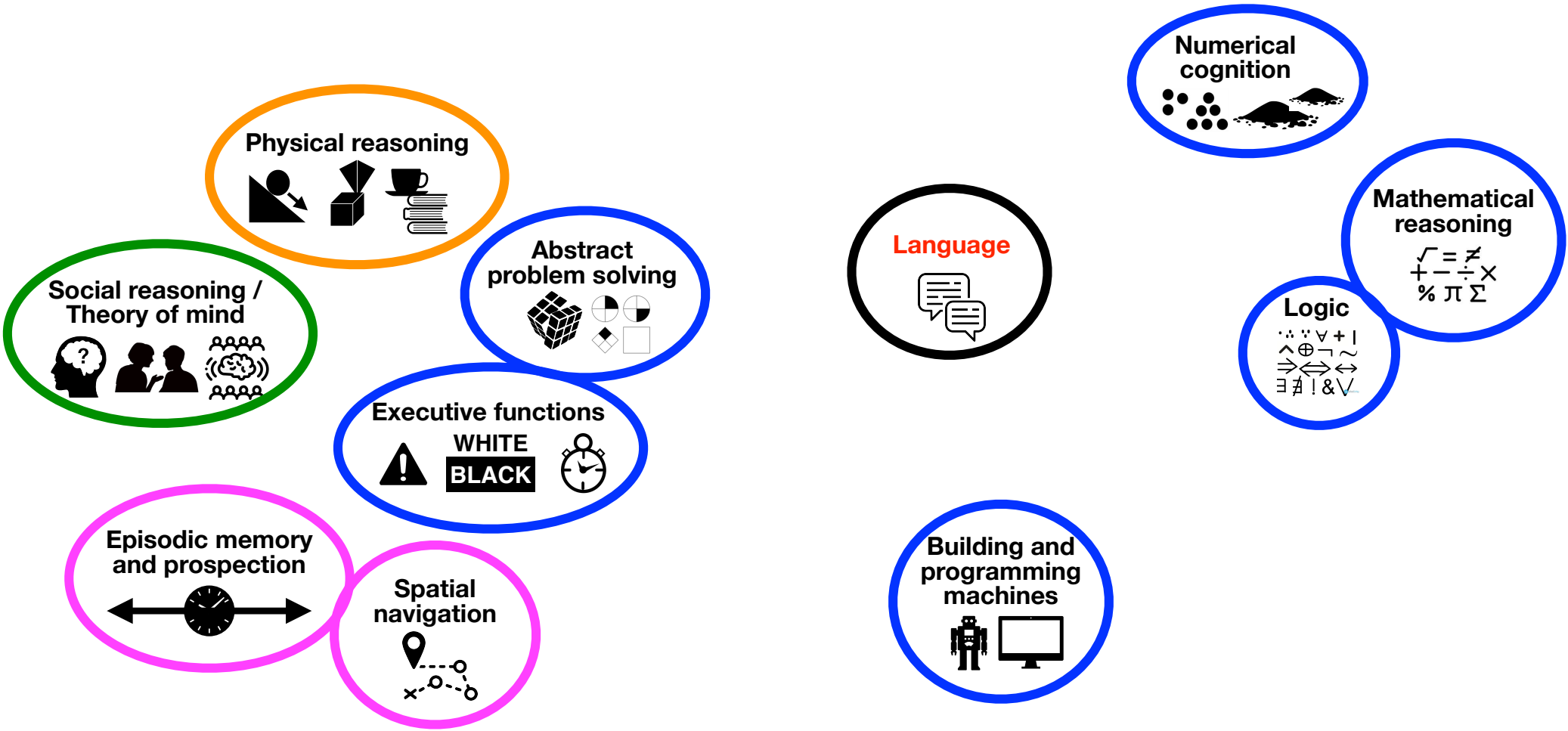
Kean et al. (in prep.)

Deductive reasoning

Building and programming machines

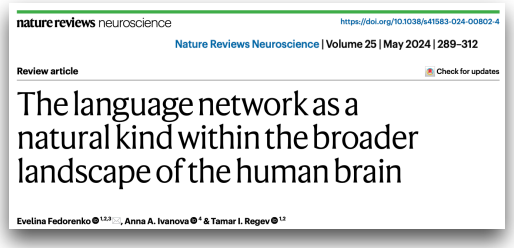
Hope Kean
(MIT)

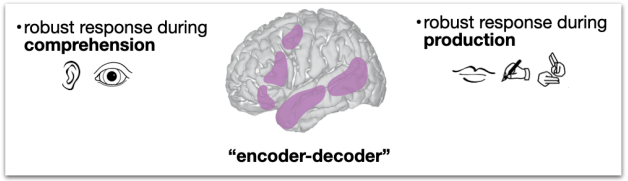
The structure of thought



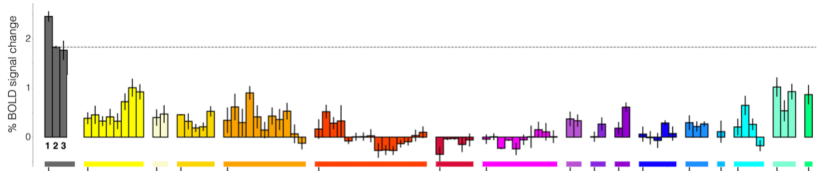
Today:

1 The human **language system**: Introduction and key properties





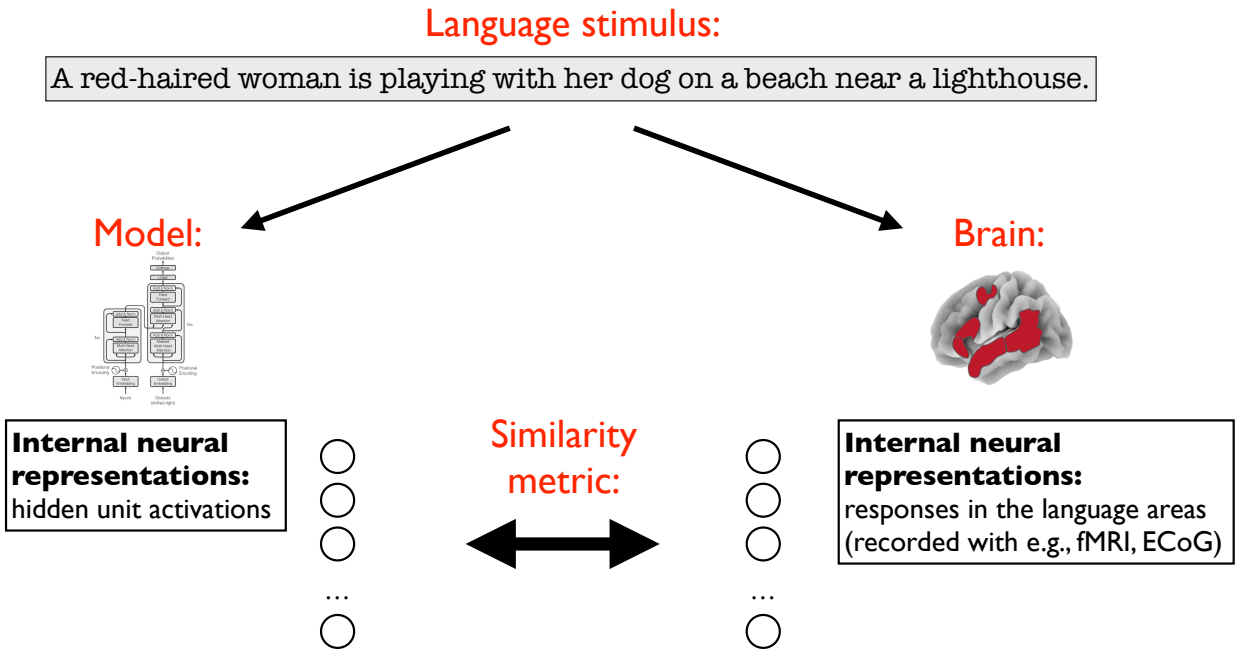
2 The relationship between **language and thought** in humans.



3 **Neural network LMs**—a new model organism for language research

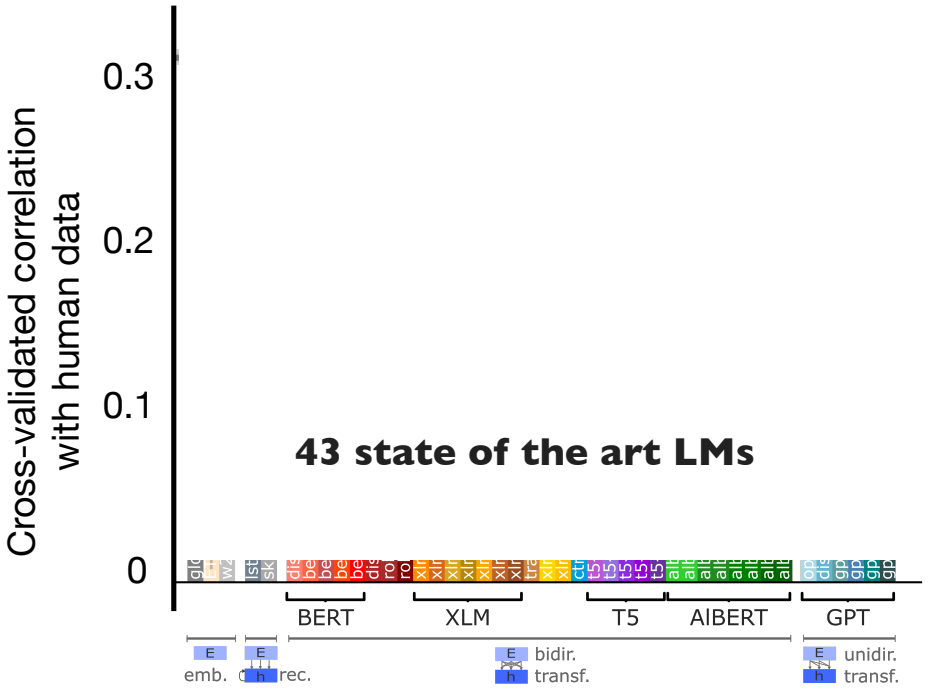
Neural network LMs as models of human language processing

Are LLMs similar to the human language system in their representations?



Neural network LMs as models of human language processing

Are LLMs similar to the human language system in their representations?

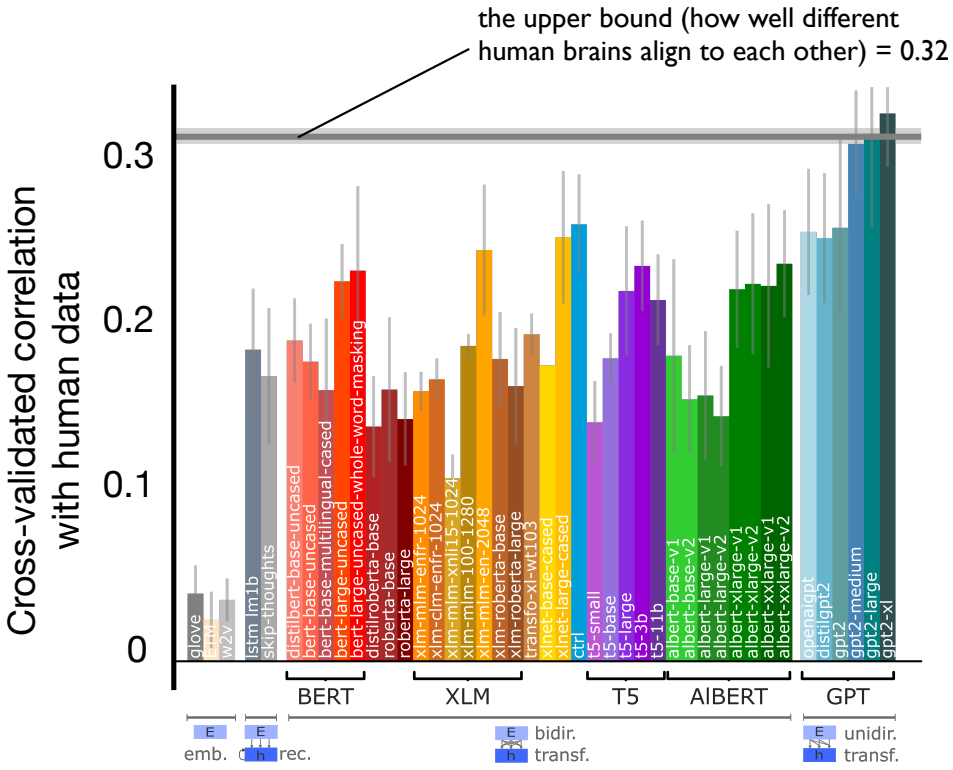


Schrimpf *et al. (2021 PNAS)*

Neural network LMs as models of human language processing

Are LLMs similar to the human language system in their representations?

Yes



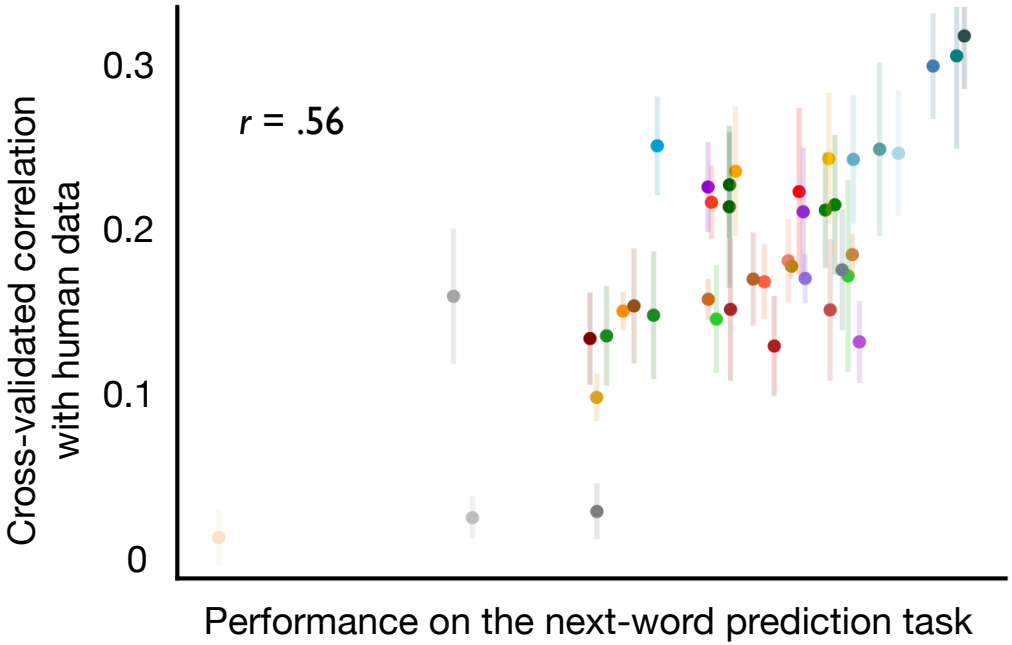
These findings have been replicated by many independent research groups, across many neural datasets.

***Schrimpf** et al. (2021 PNAS)*

Neural network LMs as models of human language processing

Are LLMs similar to the human language system in their representations?

Yes



Optimizing for **predictive representations** may be a critical objective of both biological and artificial language models.

Schrimpf *et al. (2021 PNAS)*

Neural network LMs as models of human language processing

Are LLMs similar to the human language system in their representations?

Yes

Are the representations similar enough to “control” activity in the language system?

Greta Tuckute

Tuckute *et al. (2024 NatHumBeh)*

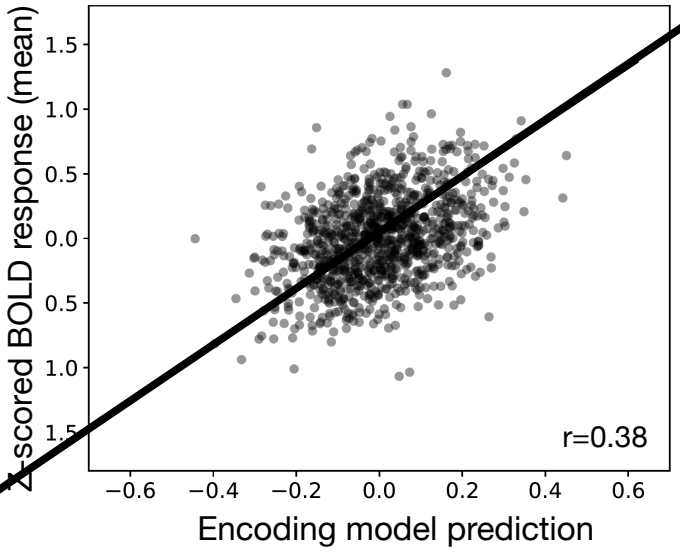
Neural network LMs as models of human language processing

Non-invasive closed-loop control of the language circuits

Greta Tuckute

***Tuckute** et al. (2024 NatHumBeh)*

Training the encoding model on 1,000 diverse sentences:



??

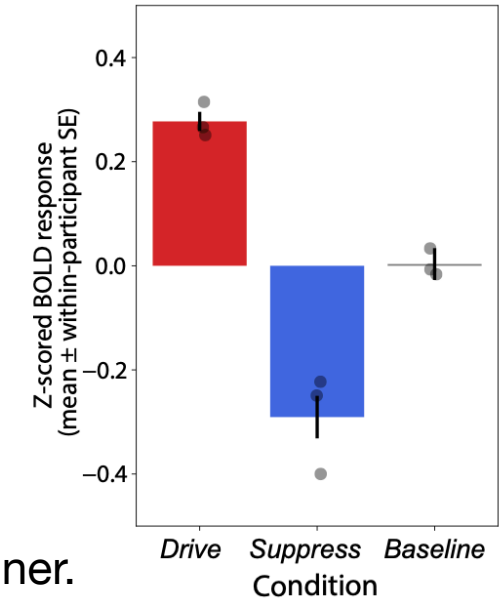
??

"Suppress" sentences

We were sitting on the couch.
That is such a beautiful picture!
They stood there for a moment.
They went up the stairs together.
Inside was a tiny silver sculpture.
They walked out onto the balcony.
Cas gazed up at the sky.
What else is there to do?

"Drive" sentences

Changing PhD group: Yes or Not?
Notice how you reacted to WTF.
Add, some feminists are call male.
Jiffy Lube of -- of therapies, yes.
People on Insta Be Like, "Gross!"
Buy sell signals remains a particular.
Turin loves me not, nor will.
URL right, or report reviewing Vimeo.



Successful modulation of brain responses to language in a closed-loop manner.

Neural network LMs as models of human language processing

Are LLMs similar to the human language system in their representations?

Yes

Are the representations similar to the language system across languages?

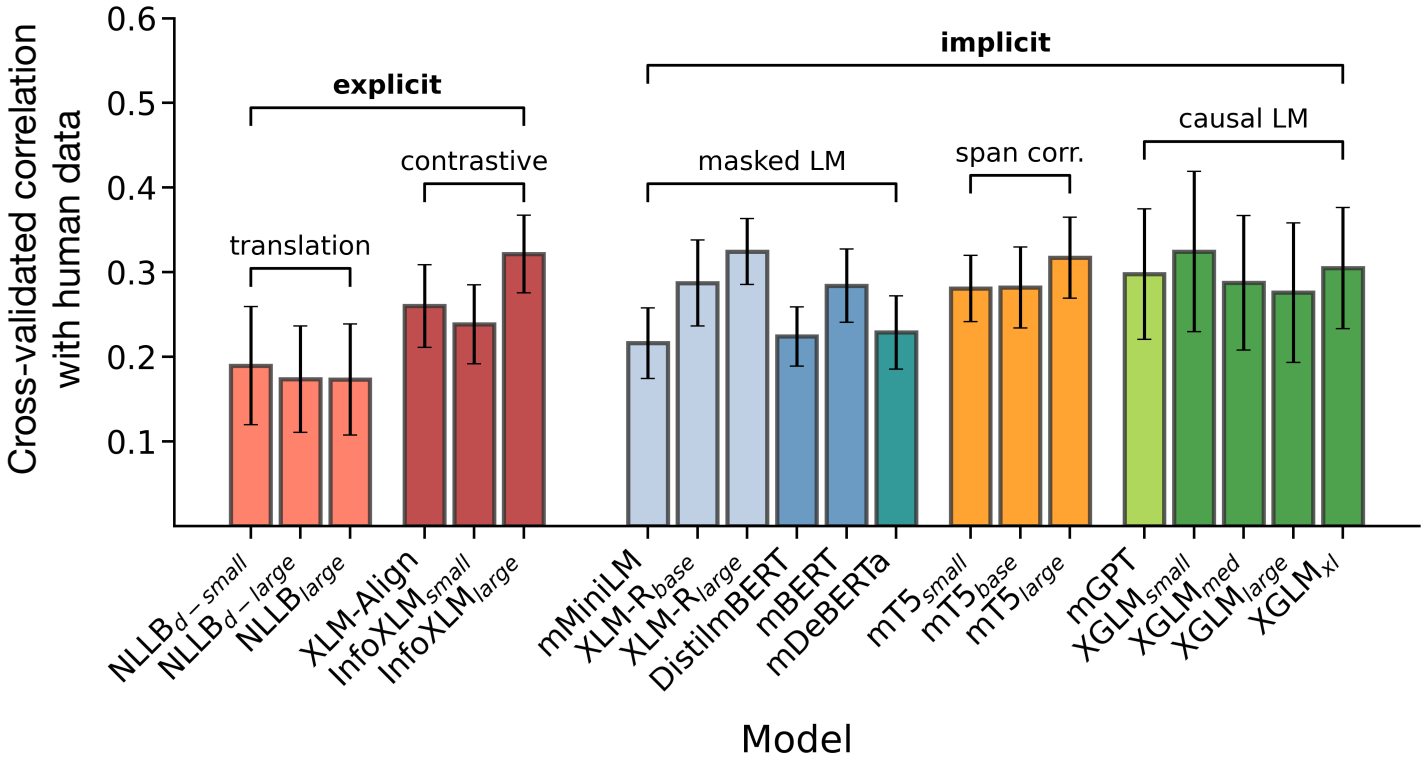
Do multilingual LMs capture some features shared across languages
to enable generalization to neual data from new languages?

Andrea de Varda

de Varda *et al. (2025 bioRxiv)*

Neural network LMs as models of human language processing

Generalization to new languages with multilingual LMs



Human neural (fMRI) data from 12 diverse languages:

- Afrikaans
- Dutch
- Farsi
- French
- Lithuanian
- Marathi
- Norwegian
- Romanian
- Spanish
- Tamil
- Turkish
- Vietnamese

***de Varda** et al. (2025 bioRxiv)*

Neural network LMs as models of human language processing

Are LLMs similar to the human language system in their representations?

Yes

Annual Review of Neuroscience

Language in Brains, Minds, and Machines

Greta Tuckute, Nancy Kanwisher, and Evelina Fedorenko

Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA;
email: evelina9@mit.edu

Neural network LMs as models of human language processing

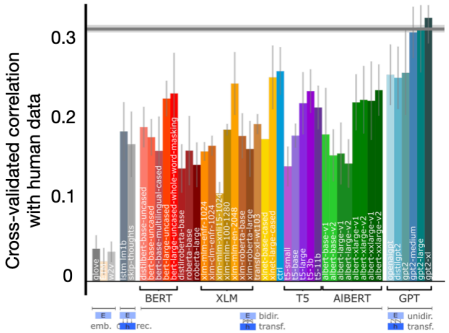
Are LLMs similar to the human language system in their representations?

Yes

Some things I find exciting:

- **What properties** make an LM similar (in its behavior and/or internal representations) to the human language system?

  (!) importance of controlled experimentation

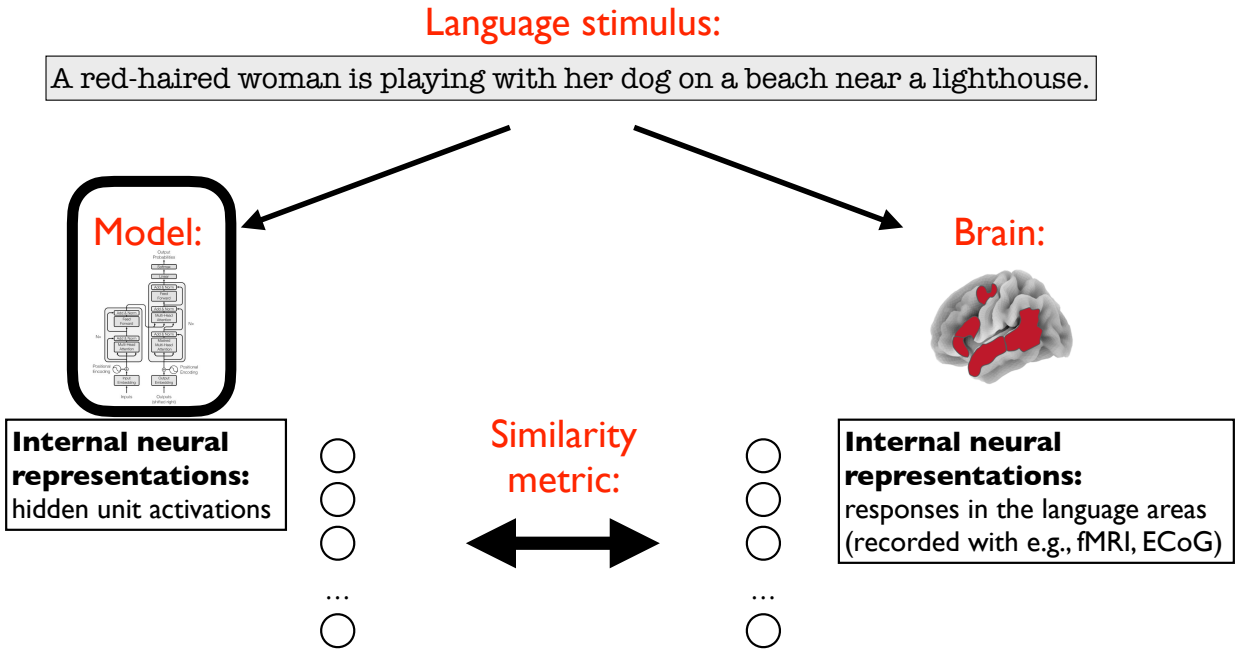


Neural network LMs as models of human language processing

Distilling the necessary and sufficient conditions for an LM to resemble the human language system

minimal model pairs varying in:

- architecture
- training data
- training objectives



Neural network LMs as models of human language processing

Distilling the necessary and sufficient conditions for an LM to resemble the human language system

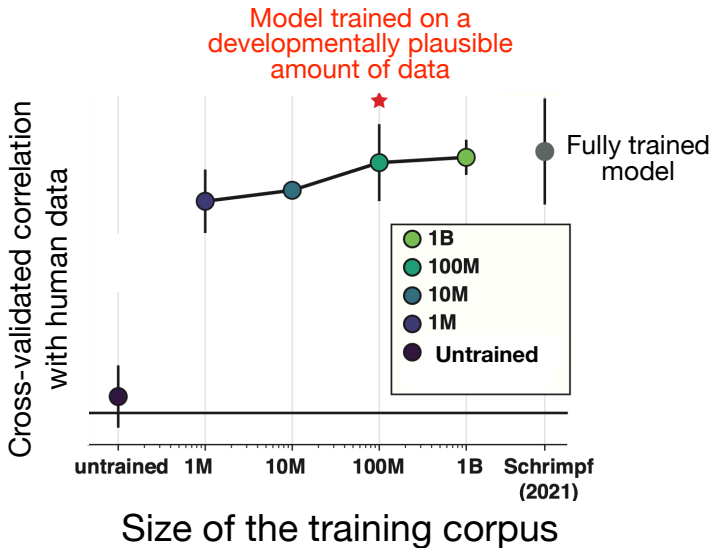
minimal model pairs varying in:

- architecture
- **training data**
- training objectives

Even when trained on a **developmentally plausible amount** of data, a GPT-style model can predict human neural responses.

Eghbal Hosseini

Hosseini *et al. (2024 NerobioLang)*



Neural network LMs as models of human language processing

Distilling the necessary and sufficient conditions for an LM to resemble the human language system

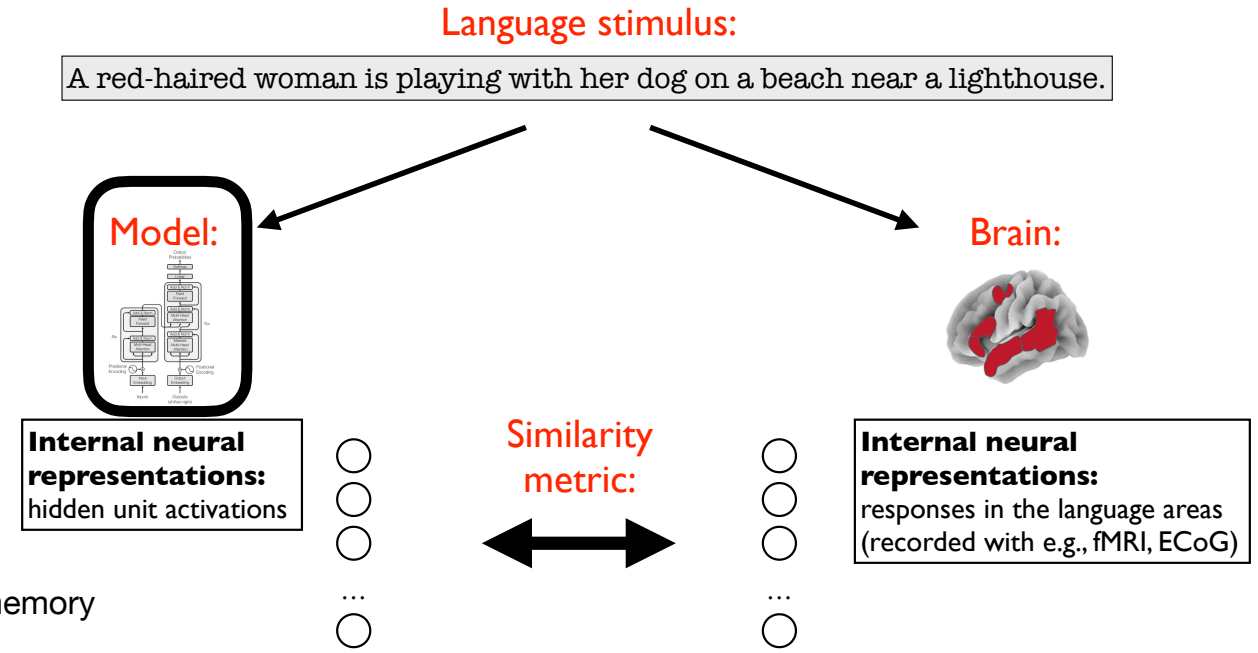
minimal model pairs varying in:

- architecture
- training data
- training objectives

(including building more ***biologically*** and ***cognitively*** plausible models)

e.g.,
- recurrent NNs
- more human-like neurons
- wiring lenght costs

e.g.,
- human-like memory limitations



Neural network LMs as models of human language processing

Are LLMs similar to the human language system in their representations?

Yes

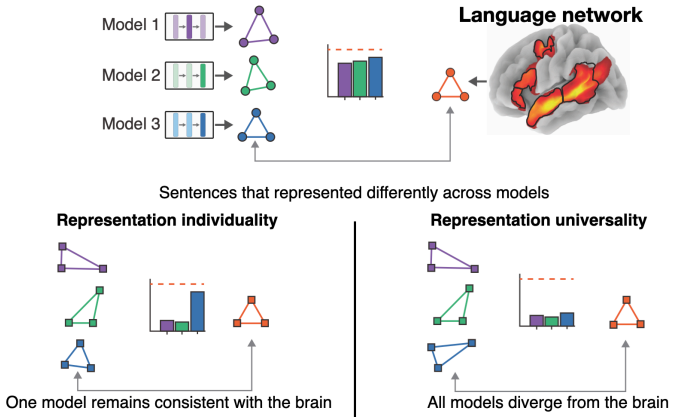
Some things I find exciting:

- **What properties** make an LM similar (in its behavior and/or internal representations) to the human language system?
- What **linguistic features** are shared betwen LM and human representations? What are the **core dimensions** of linguistic representations?

Neural network LMs as models of human language processing

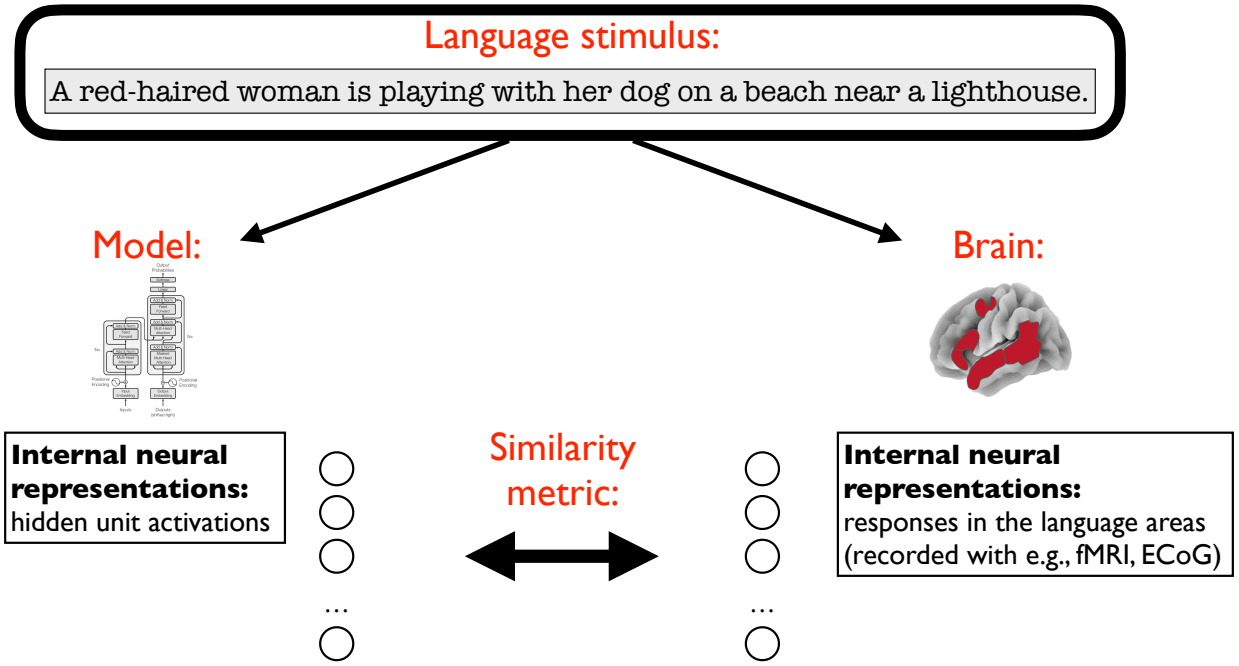
Linguistic representations in LMs vs. humans

Uncovering the **representational axes** of language via 'controversial stimuli' ("stress-testing" the models):



Eghbal Hosseini

***Hosseini** et al. (2023 CCN; 2025 bioRxiv)*



Neural network LMs as models of human language processing

Are LLMs similar to the human language system in their representations?

Yes

Some things I find exciting:

- **What properties** make an LM similar (in its behavior and/or internal representations) to the human language system?
- What **linguistic features** are shared betwen LM and human representations? What are the **core dimensions** of linguistic representations?
- Using LMs as tools for understanding typical and atypical **language development**, and acquired **language disorders**.

Neural network LMs as models of human language processing

Language development

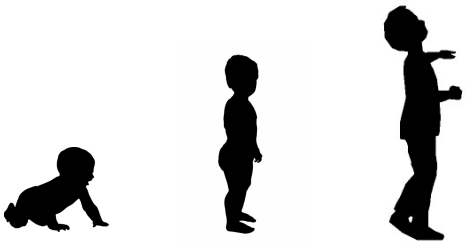
("controlled rearing" approaches)

Building and evaluating **developmentally plausible language models**, including:

speech-based models

multimodal (language+vision) models

Relating **representations from 'baby language models'** to **neural data** from children across the developmental trajectory.



Greta Tuckute

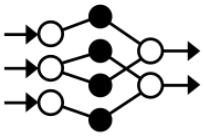
Chengxu Zhuang

***Zhuang** et al. (2023, 2024)*

Halie Olson

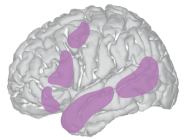
Language disorders

(model ablation and related approaches)



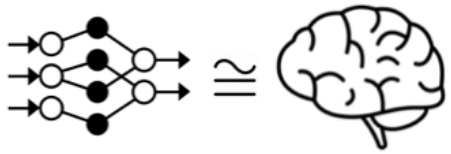
Take-aways

Language and thought are robustly **distinct** in the human brain.

→ **Language** is supported by a **specialized** brain network. 

→ Different aspects of **thought** rely on distinct brain networks, but the ontology of thought requires more work.

Representations from **neural network LMs** are **similar** to those in the human language system.



Thank you!

My amazing labbies!

Not pictured current/incoming members:
Halie Olson, Sara Swords, Alex Fung, Selena She, Agata Wolna, Chiebuka Ohams, Anvitha Kachinthaya, Aaron Wright, Kumar Duraivel

Former labbies:

- Idan Blank
- Alex Paunov
- Zuzanna Balewski
- Olessia Jouravlev
- Terri Scott
- Zach Mineroff
- Bri Pritchett
- Caitlyn Hoeflin
- Melissa Kline
- Nafisa Syed
- Moataz Assem
- Jeanne Gallée
- Dima Ayyash
- Yev Diachek
- Matt Siegelman
- Yotaro Sueoka
- Jessica Chen
- Alvincé Pongos
- Miriam Hauptman
- Rachel Ryskin
- Josef Affourtit
- Hannah Small
- Maya Taliaferro
- Sammy Floyd
- Anna Ivanova
- Aalok Sathe
- Hee So Kim
- Niharika Jhingan
- Carina Kauf
- Chengxu Zhuang
- Cory Shain

Select collaborators:

- **Nancy Kanwisher**
- **Ted Gibson**
- **Steve Piantadosi**
- **Kyle Mahowald**

- Jacob Andreas
- Anne Billot
- Peter Brunner
- Anila D'Mello
- Simon Fisher
- John Gabrieli
- Swathi Kiran
- Roger Levy
- Frank Mollica
- Alfonso Nieto-Castañón
- Sam Norman-Haignere
- Amanda O'Brien
- Ola Ozernov-Palchik
- Mark Richardson
- Rebecca Saxe
- Zeynep Saygin
- Martin Schrimpf
- Josh Tenenbaum
- Rosemary Varley
- Maria Varkanitsa
- Noga Zaslavsky

Funding support:

NIH NIDCD
NIH NICHD
NIH NINDS
McGovern Institute
brain+cognitive sciences
SCSB Simons Center for the Social Brain
K. Lisa Yang ICoN Center
MIT Quest for Intelligence

How to find us:
evlab.mit.edu
@ev_fedorenko
@evfedorenko.bsky.social