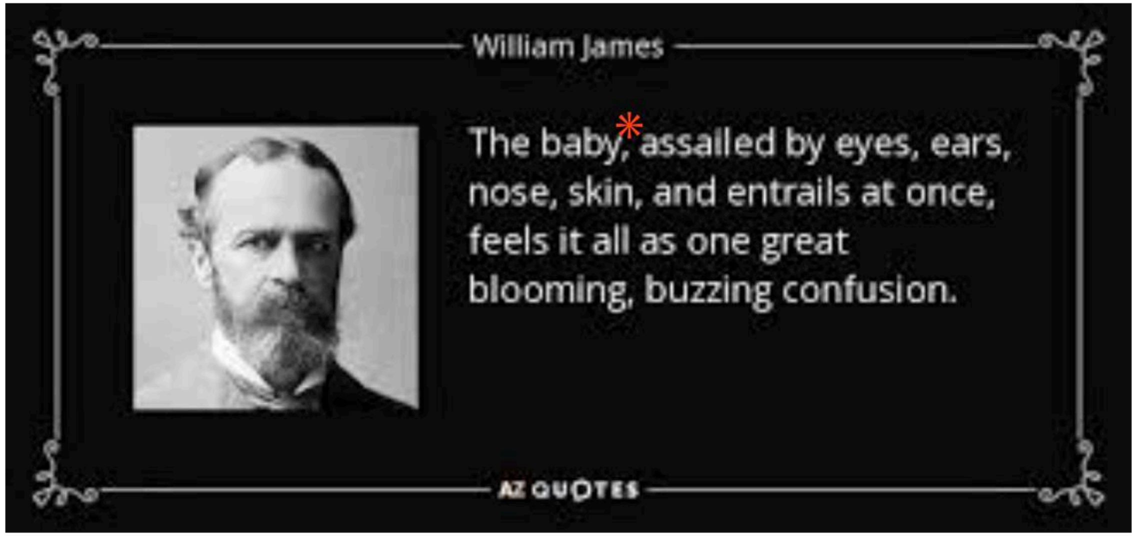
CS375 / Psych 249:
Large-Scale Neural Network Models for Neuroscience

Lecture 7: Modelling Infant Development

2026.01.28

Dr. Clíona O'Doherty
Department of Psychology
Scaffolding of Cognition Team | Stanford NeuroAI Lab
Stanford University

Understanding complex, noisy data streams is a critical part of cognition.



Without sophisticated parsing and entity extraction, the world would be "as one great blooming, buzzing confusion" (for babies or otherwise).

*actually not clearly true for babies . . .

How does intelligent behavior emerge?

Infants must learn to…
parse their sensory input into meaningful knowledge

How does intelligent behavior emerge?

… use their bodies to interact with the objects they see

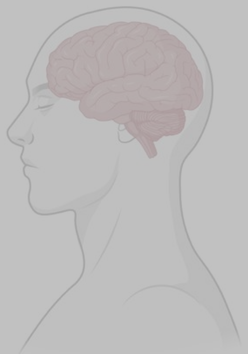
How does intelligent behavior emerge?

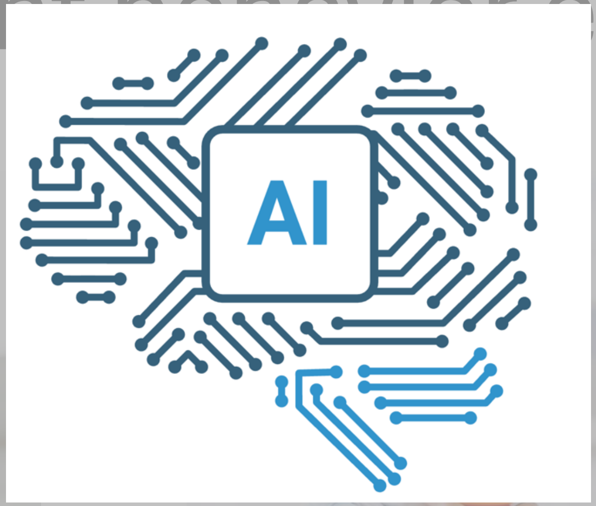
… control their own motion through a space to interact with the environment

How does intelligent behavior emerge?

… language and how to map words to referents

How does intelligen[…] emerge?





How can we model this?

Learn language and how to map words to referents

Overview

- Why should we model development?
- How to study infants? What do we know about early life?
- Recent advances in Developmental NeuroAI.

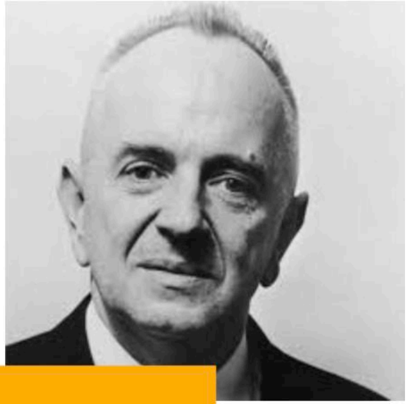
Is it fair to say that AI is really like a baby?

Overview

- Why should we model development?
- How to study infants? What do we know about early life?
- Recent advances in Developmental NeuroAI.

Is it fair to say that AI is really like a baby?

"Nothing in biology makes sense except in light of evolution"

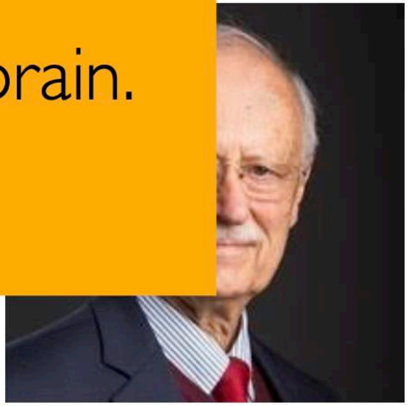


Dobzhansky

"Nothing

Restated:

Behavior is highly constraining of the brain.



Gordon Shepherd

computational

Nothing in neuroscience makes sense except in light of ***optimization***.

Stanford Neuro AI Lab

Behavior is highly constraining of the brain.

Behavior is highly constraining of the brain.

CellPress

Neuron
Perspective

Neuroscience Needs Behavior: Correcting a Reductionist Bias

John W. Krakauer,[1,*] Asif A. Ghazanfar,[2] Alex Gomez-Marin,[3] Malcolm A. MacIver,[4] and David Poeppel[5,6]

There are ever more compelling tools available for neuroscience research, ranging from selective genetic targeting to optogenetic circuit control to mapping whole connectomes. These approaches are coupled with a deep-seated, often tacit, belief in the reductionist program for understanding the link between the brain and behavior. The aim of this program is causal explanation through neural manipulations that allow testing of necessity and sufficiency claims. We argue, however, that another equally important approach seeks an alternative form of understanding through careful theoretical and experimental decomposition of behavior. Specifically, the detailed analysis of tasks and of the behavior they elicit is best suited for discovering component processes and their underlying algorithms. In most cases, we argue that study of the neural implementation of behavior is best investigated *after* such behavioral work. Thus, we advocate a more pluralistic notion of neuroscience when it comes to the brain-behavior relationship: behavioral work provides understanding, whereas neural interventions test causality.

Behavior is highly constraining of the brain.




But how does that behavior arise?

Learning to adapt and behave in the first year of life

What are the mechanisms of neural development and cognitive function during infancy?

Nature *versus* Nurture

Nativism

Our development is preprogrammed by genetics

Plato - our sense data do not provide sufficient information to specify the abstract ideas and knowledge that humans possess

Empiricism

Our environments and experiences shape our development

Aristotle - our sense data are sufficient to specify abstract concepts and ideas and, therefore, that human knowledge is acquired through everyday experience.

It's not that simple …

Empirical findings show that this dichotomy is implausible (Lewkowicz, 2011)

Domain-specific

Infants possess fragile, specific innate knowledge

This can be sophisticated knowledge but might break under seemingly trivial circumstances

Learning plays the critical role of optimizing these basic building blocks

Domain-general

Infants possess learning mechanisms that allow them to quickly acquire sophisticated knowledge

Experience is important but not necessary since infants can make inferences

These learning and reasoning mechanisms are innate

It's not that simple …

Empirical findings show that this dichotomy is implausible (Lewkowicz, 2011)

Domain-specific

Infants possess fragile, specific innate knowledge

This can be sophist… but might break un… circumstances

Learning plays the critical role of optimizing these basic building blocks

Domain-general

Infants possess learning mechanisms … to quickly acquire …nowledge

…mportant but not …e infants can make

These learning and reasoning mechanisms are innate

We need to figure out the relative importance of each

How to study preverbal infants?

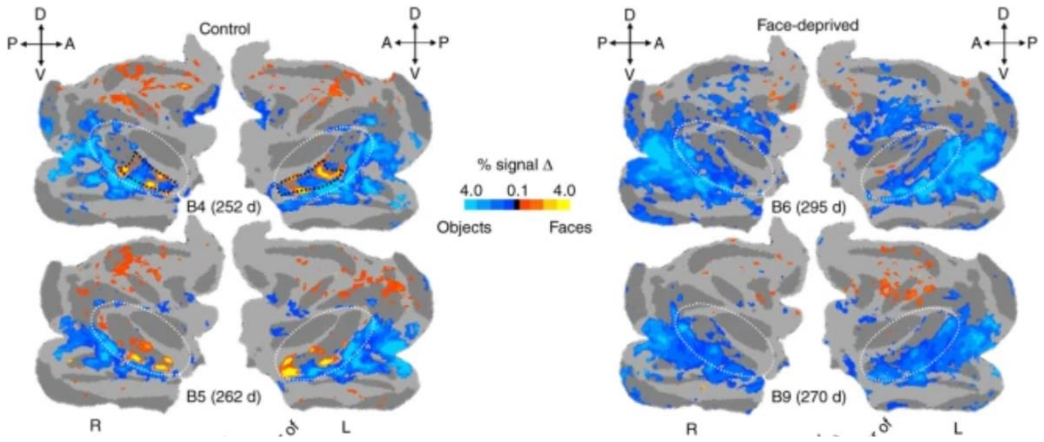
We can’t ask them to do complex experimental tasks

We can’t interfere with their learning and development

- Although we can use naturally arising differences such as preterm birth or twin studies

How to study preverbal infants?

Figure 1: Faces>objects and hands>objects activations in control and face-deprived monkeys.



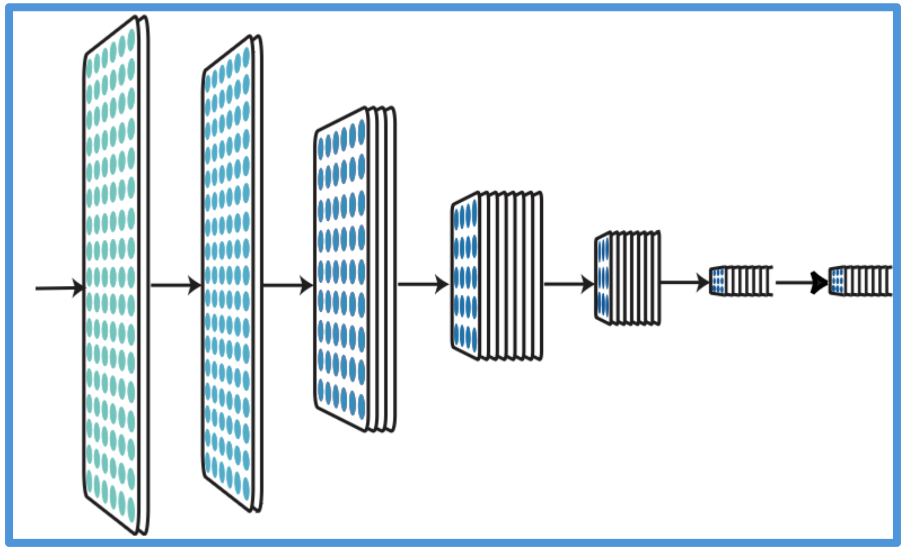
No controlled rearing like with animal models Arcaro et al., (2017)

We can't ask them to do complex experimental tasks

We can't interfere with their learning and development

- Although we can use naturally arising differences such as preterm birth or twin studies

Could computational models be the answer?



Brain

Model

What do developmentalists want from models?

1. Fit to data: Models must be good. However, they don't have to be perfect to be useful

2. Open: Being able to access both the training sample and trained weights is critical for experimental research
Frank, 2023

Nice to have: Interpretable

Once we have this, we can ask why questions
Kanwisher, et al., 2023

What do developmentalists want from models?

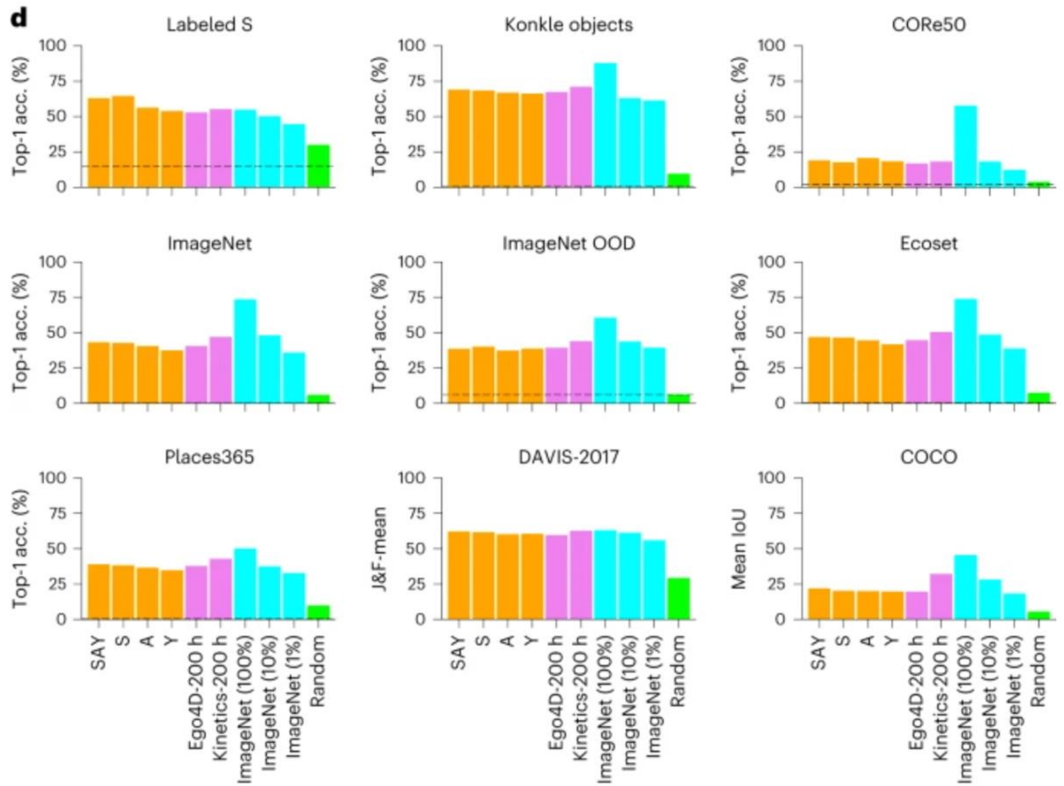
1. Fit to data: Models must be good. However, they don't have to be perfect to be useful

Article | Published: 07 March 2024

Learning high-level visual representations from a child's perspective without strong inductive biases

A. Emin Orhan & Brenden M. Lake

Nature Machine Intelligence **6**, 271–283 (2024) | Cite this article



What do developmentalists want from models?

1. Fit to data: Models must be good. However, they don't have to be perfect to be useful

2. Open: Being able to access both the training sample and trained weights is critical for experimental research

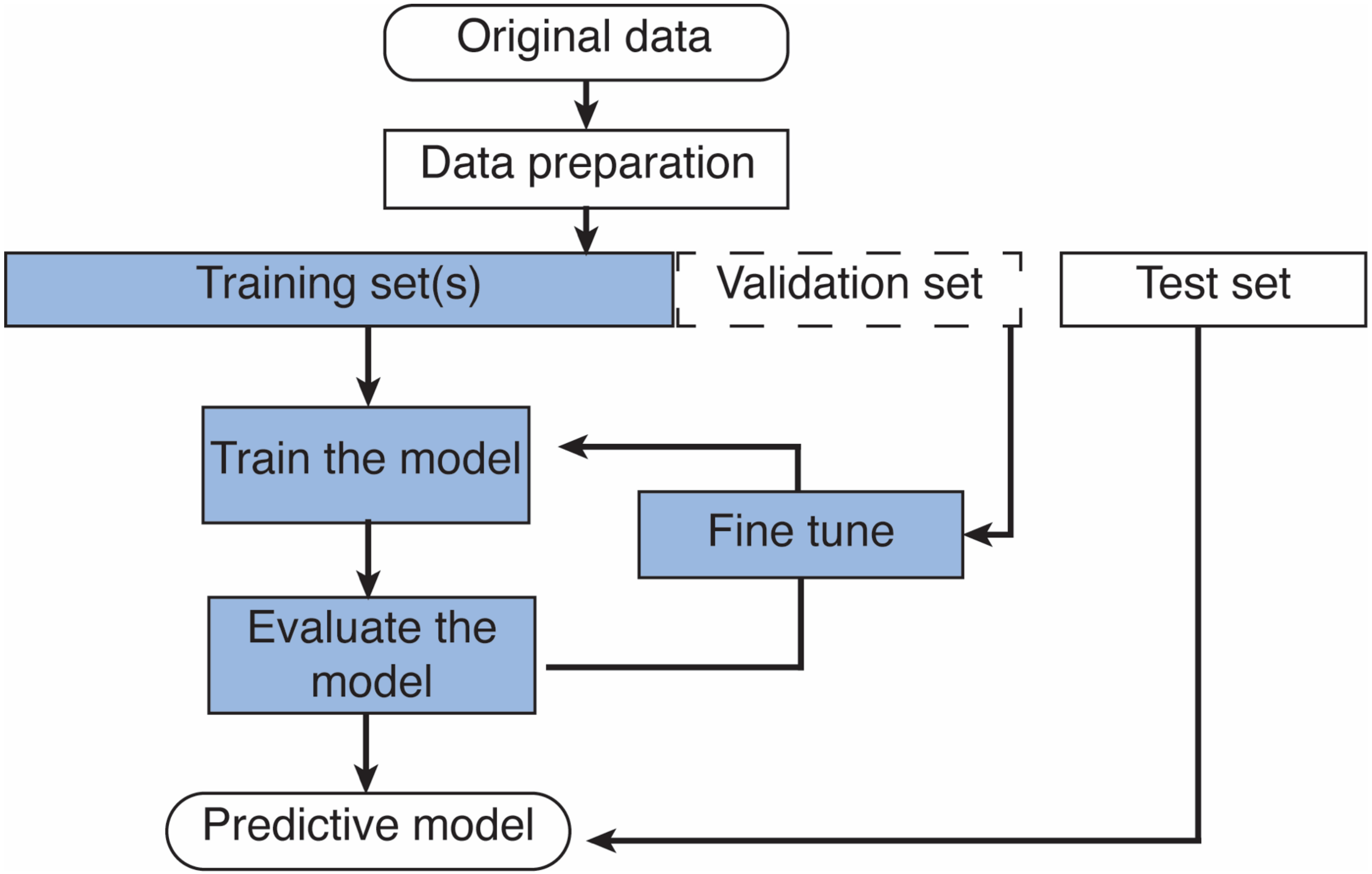
Frank, 2023

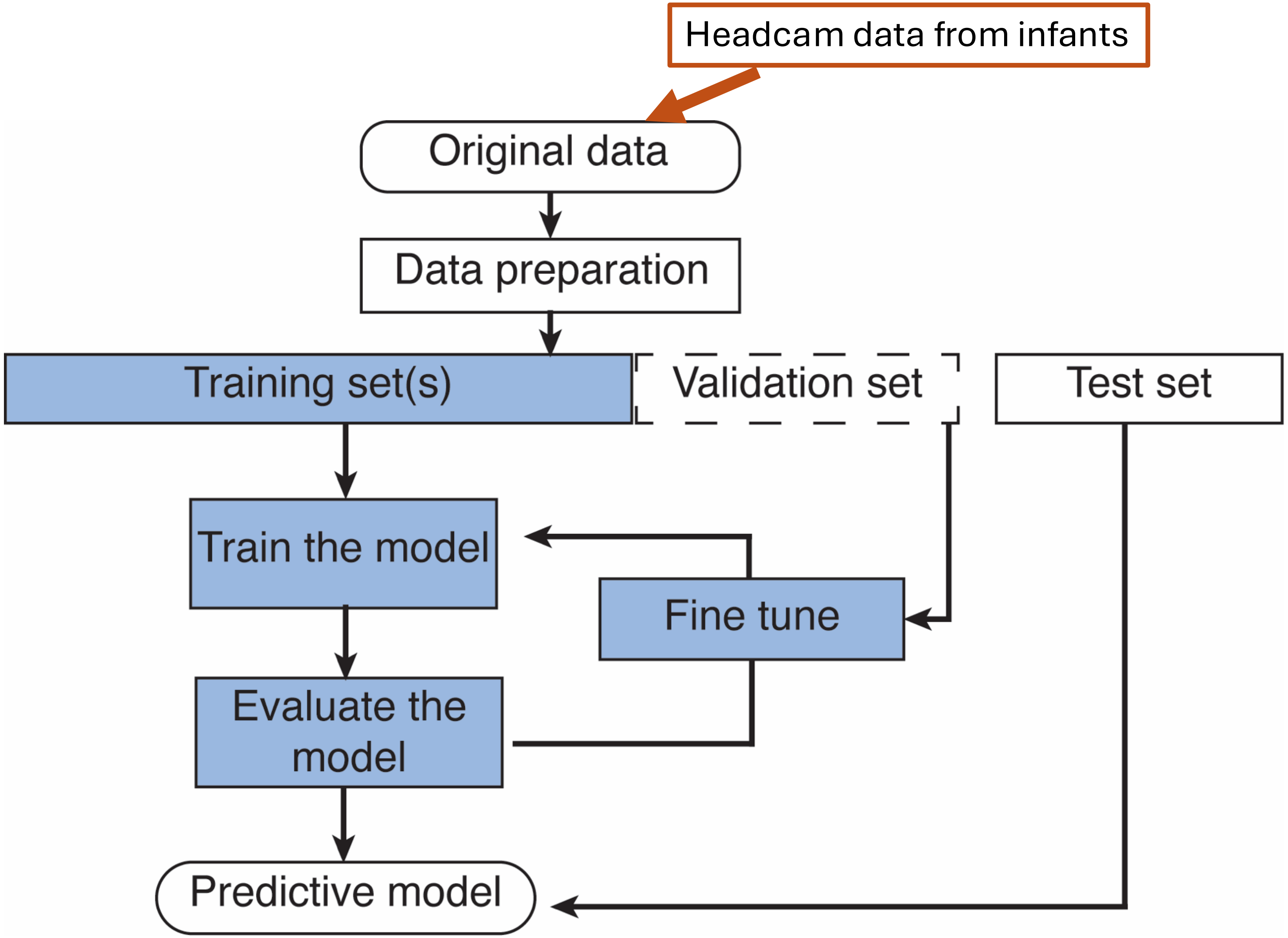
Nice to have: Interpretable

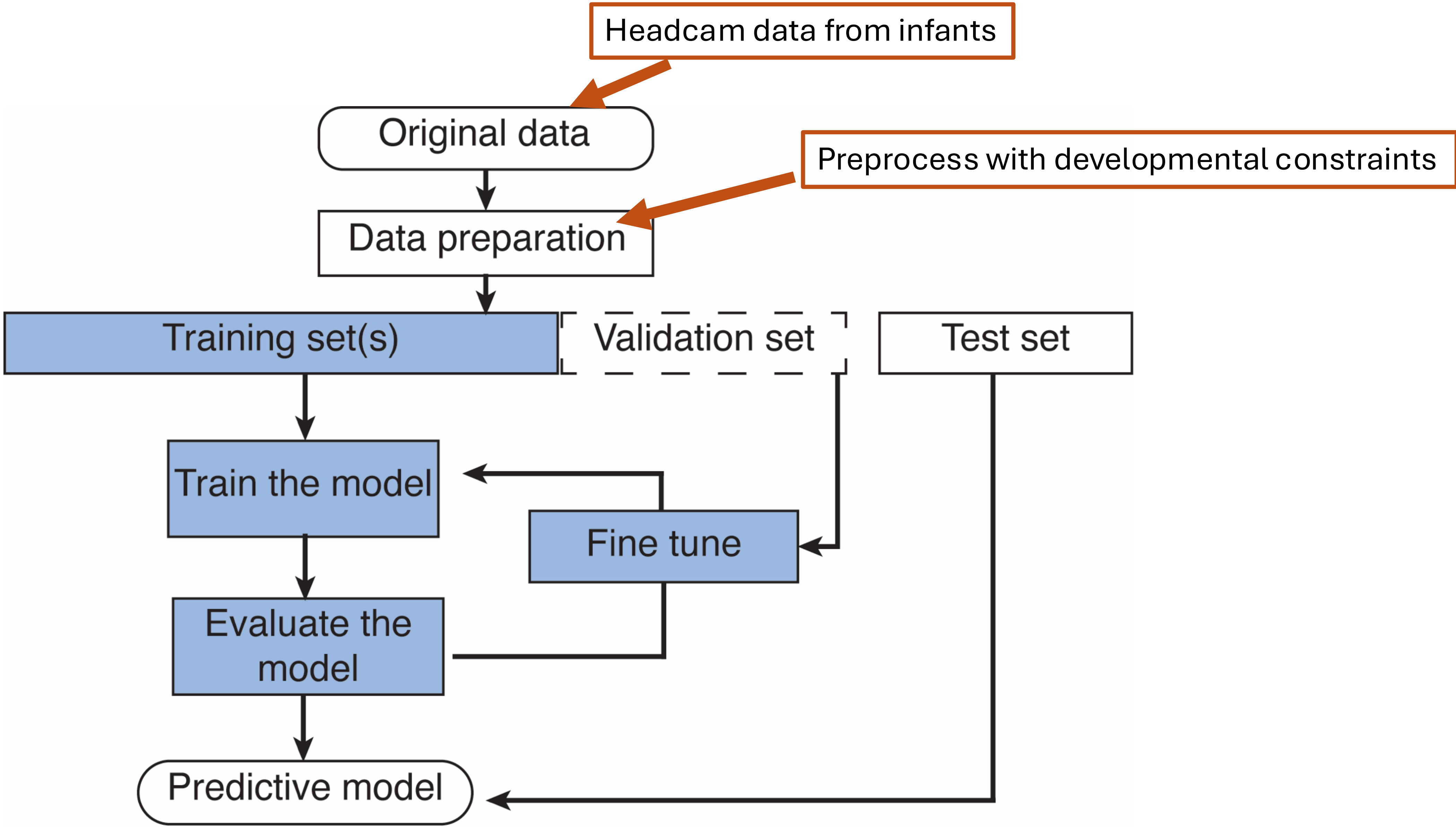
Once we have this, we can ask why questions

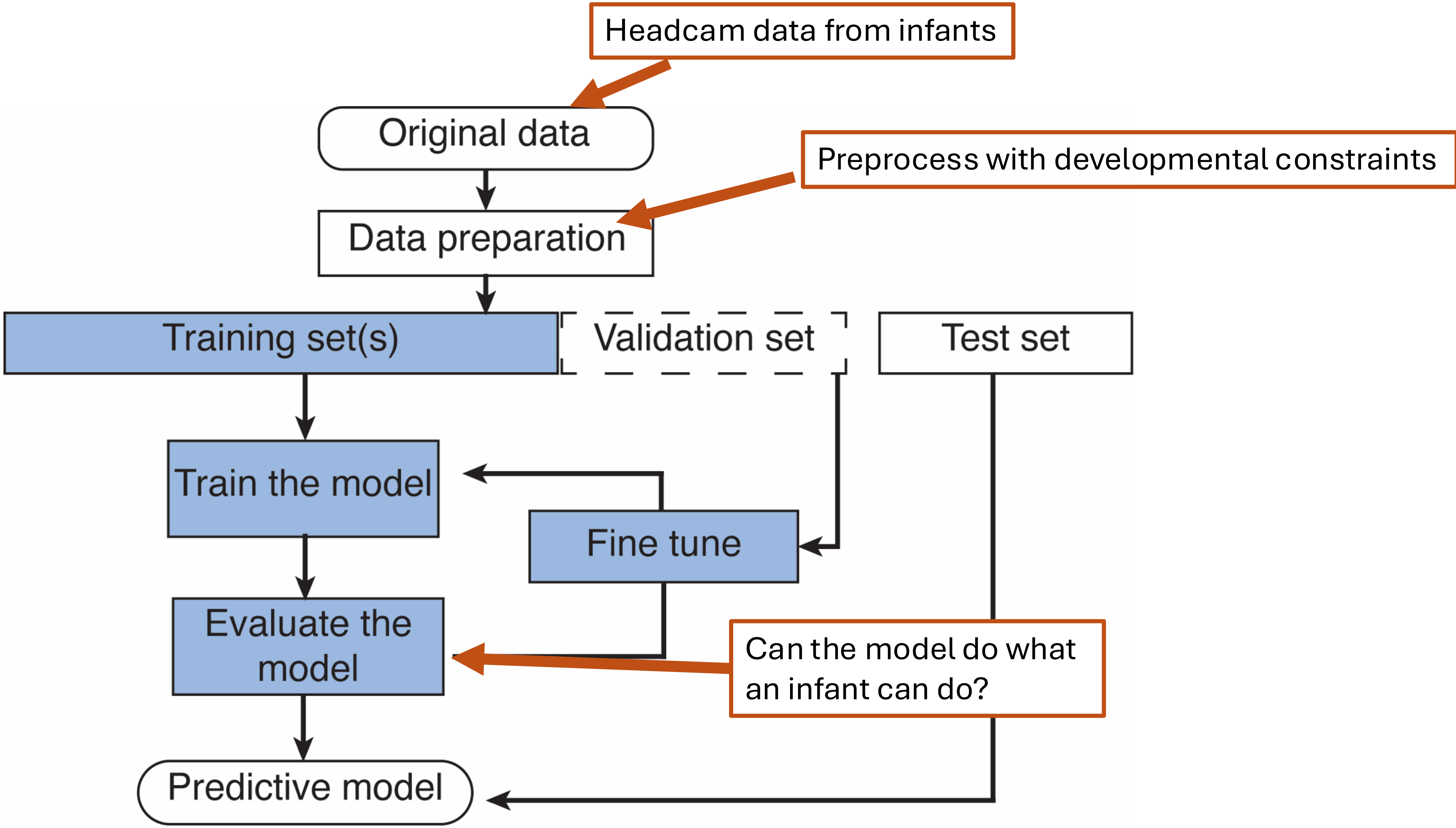
Kanwisher, et al., 2023

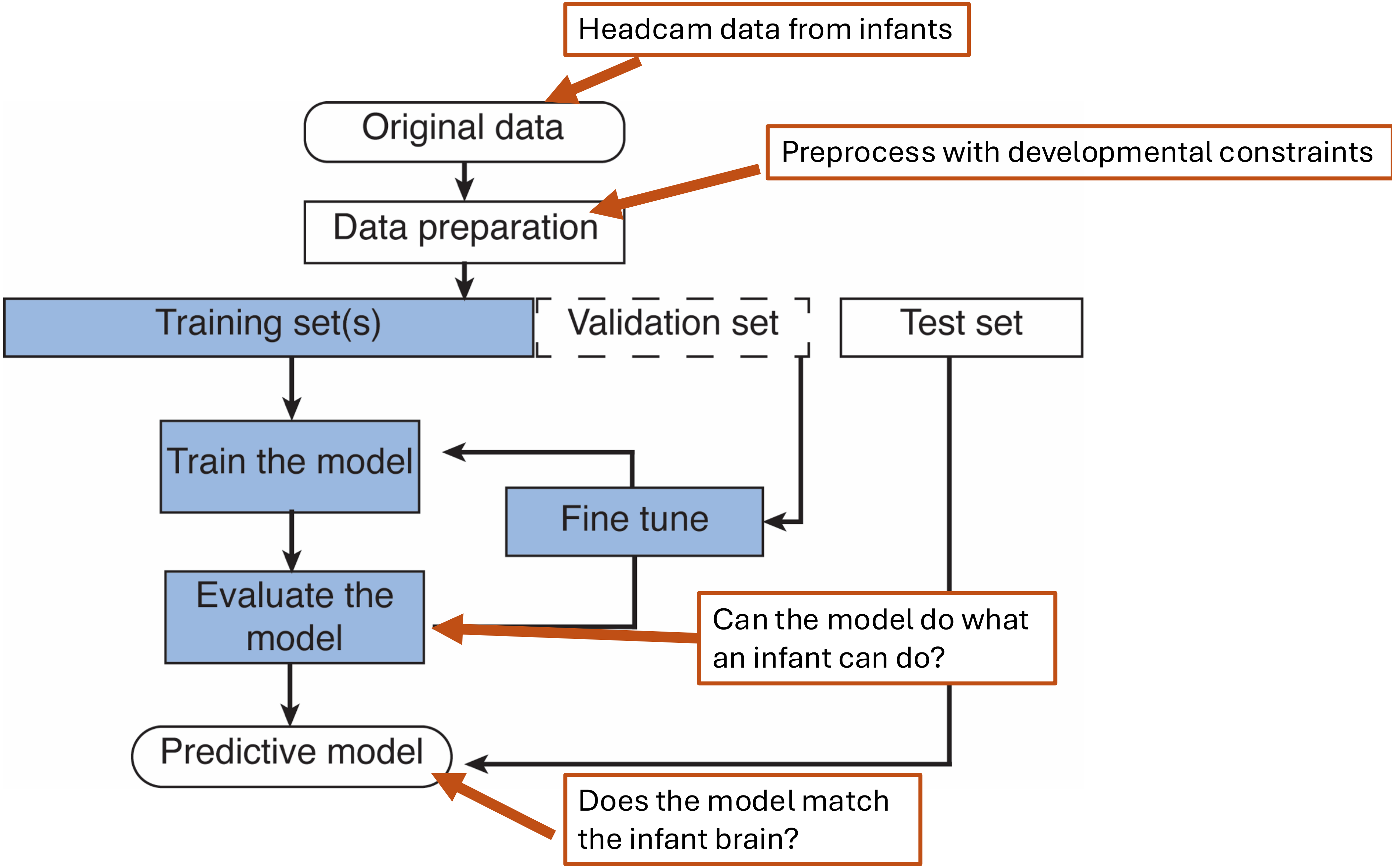
What does development have to offer modelers?

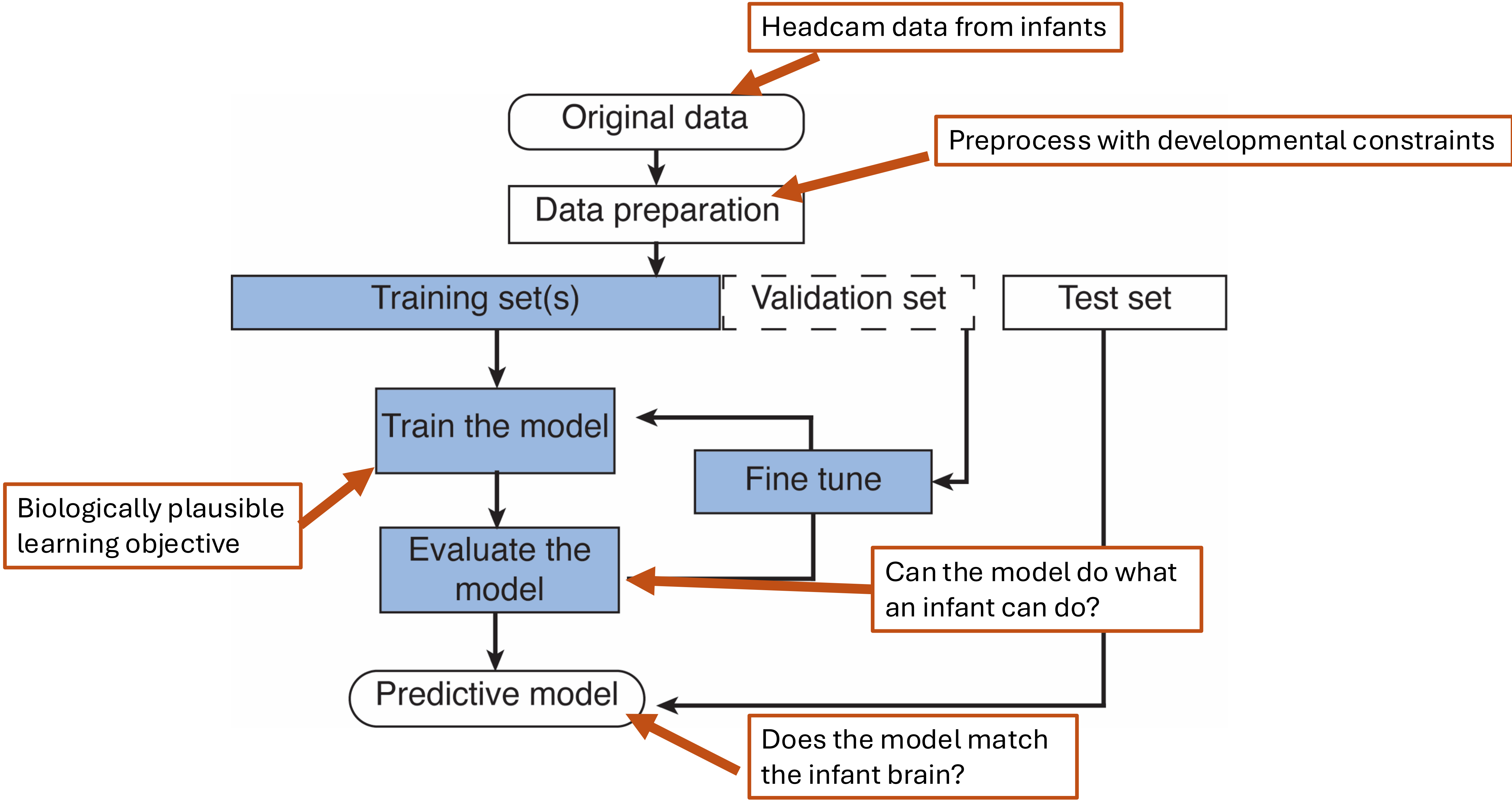


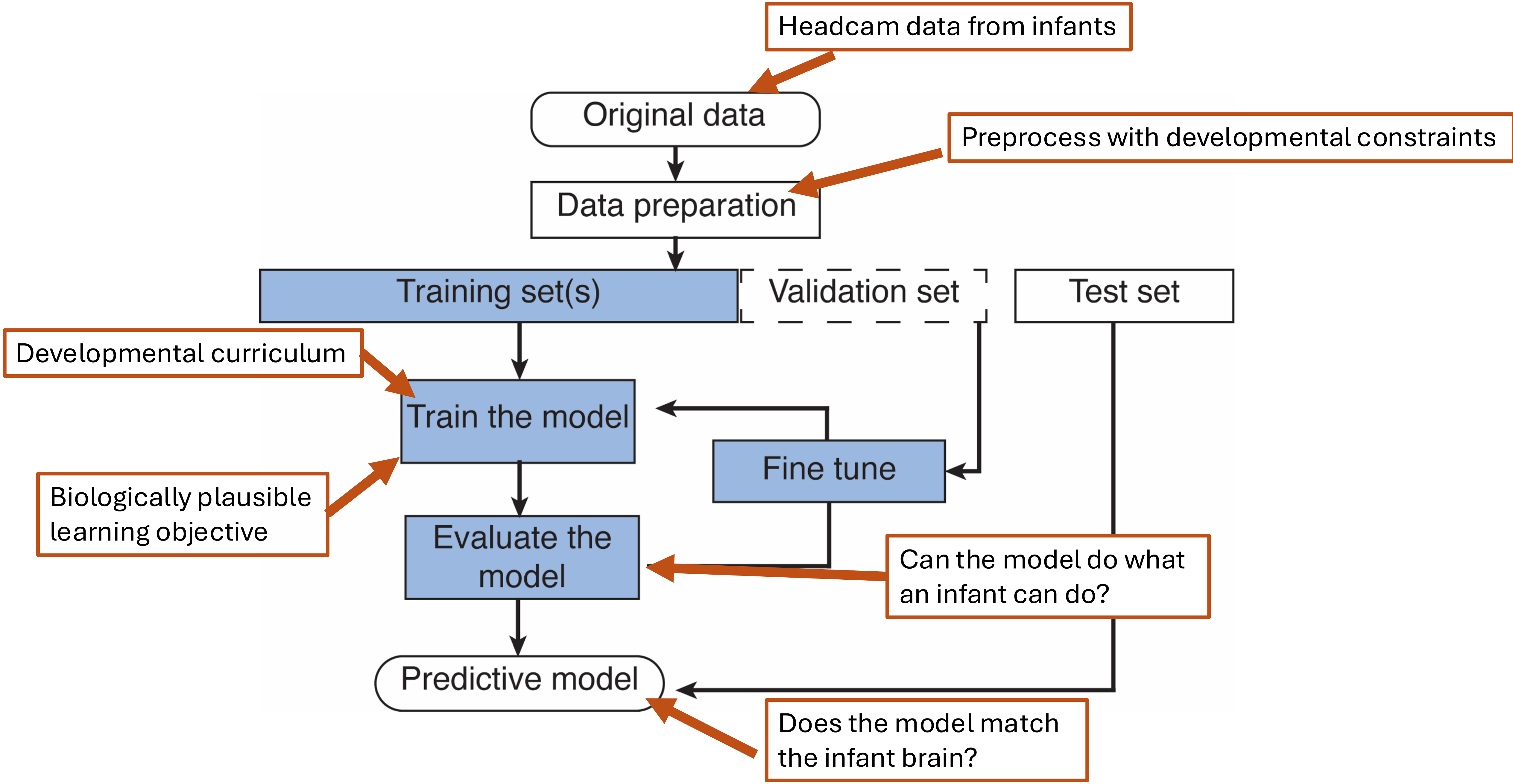

Headcam data from infants
Original data
Data preparation
Training set(s)
Validation set
Test set
Train the model
Fine tune
Evaluate the model
Predictive model


Headcam data from infants
Original data
Preprocess with developmental constraints
Data preparation
Training set(s)
Validation set
Test set
Train the model
Fine tune
Evaluate the model
Predictive model


Headcam data from infants
Original data
Preprocess with developmental constraints
Data preparation
Training set(s)
Validation set
Test set
Train the model
Fine tune
Evaluate the model
Can the model do what an infant can do?
Predictive model

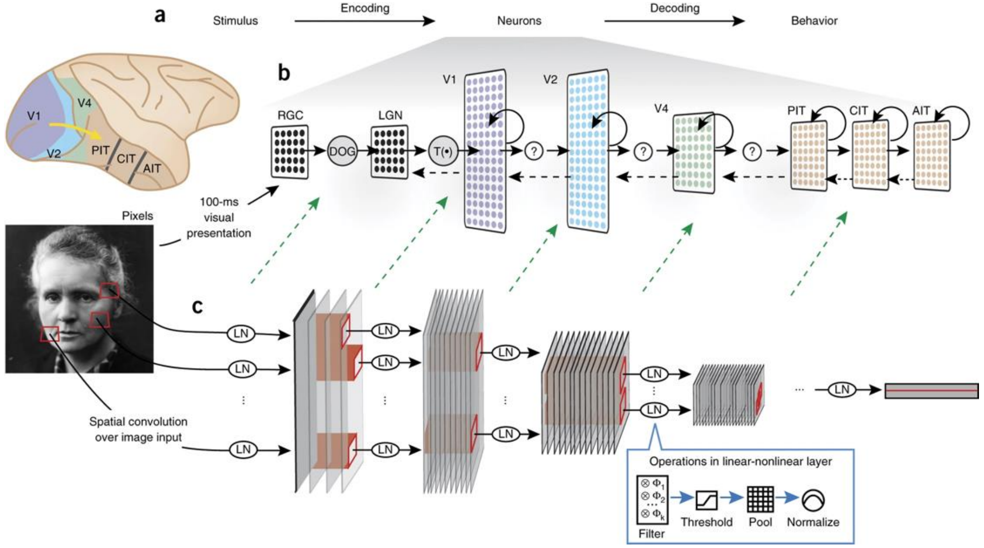

Headcam data from infants
Original data
Preprocess with developmental constraints
Data preparation
Training set(s)
Validation set
Test set
Train the model
Fine tune
Evaluate the model
Can the model do what an infant can do?
Predictive model
Does the model match the infant brain?


Headcam data from infants
Original data
Preprocess with developmental constraints
Data preparation
Training set(s)
Validation set
Test set
Train the model
Fine tune
Biologically plausible learning objective
Evaluate the model
Can the model do what an infant can do?
Predictive model
Does the model match the infant brain?


Headcam data from infants
Original data
Preprocess with developmental constraints
Data preparation
Training set(s)
Validation set
Test set
Developmental curriculum
Train the model
Fine tune
Biologically plausible learning objective
Evaluate the model
Can the model do what an infant can do?
Predictive model
Does the model match the infant brain?

Cognitive Computational Neuroscience

Deep neural network modelling of object recognition and high-level vision.



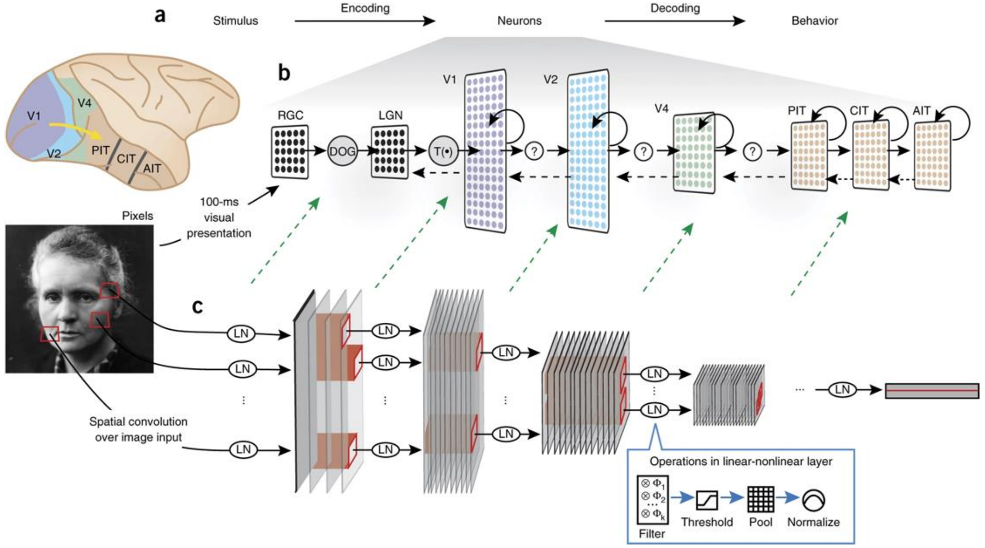
Yamins & DiCarlo, 2016

Learning to recognise in the first year of life

What are the mechanisms of neural development and visual function during infancy?

Cognitive Computational Neuroscience

Deep neural network modelling of object recognition and high-level vision.



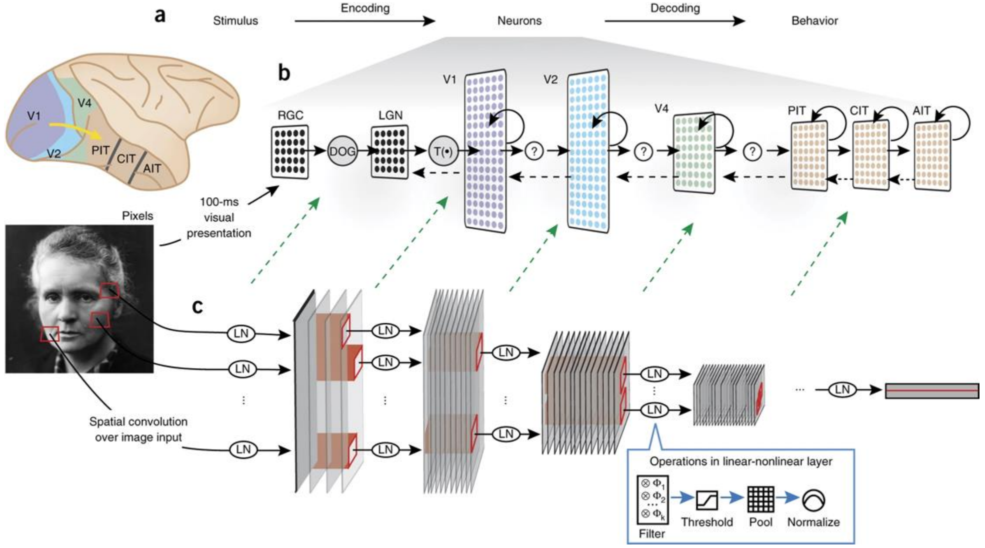
Yamins & DiCarlo, 2016

Learning to recognise in the first year of life

What are the mechanisms of neural development and visual function during infancy?

Developmental Cognitive Computational Neuroscience

Deep neural network modelling of object recognition and high-level vision.



Yamins & DiCarlo, 2016

Overview

- Why should we model development?
- How to study infants? What can we do in early life?
- Recent advances in Developmental NeuroAI.

Is it fair to say that AI is really like a baby?

How to study preverbal infants?

Where we look reveals something about what we know

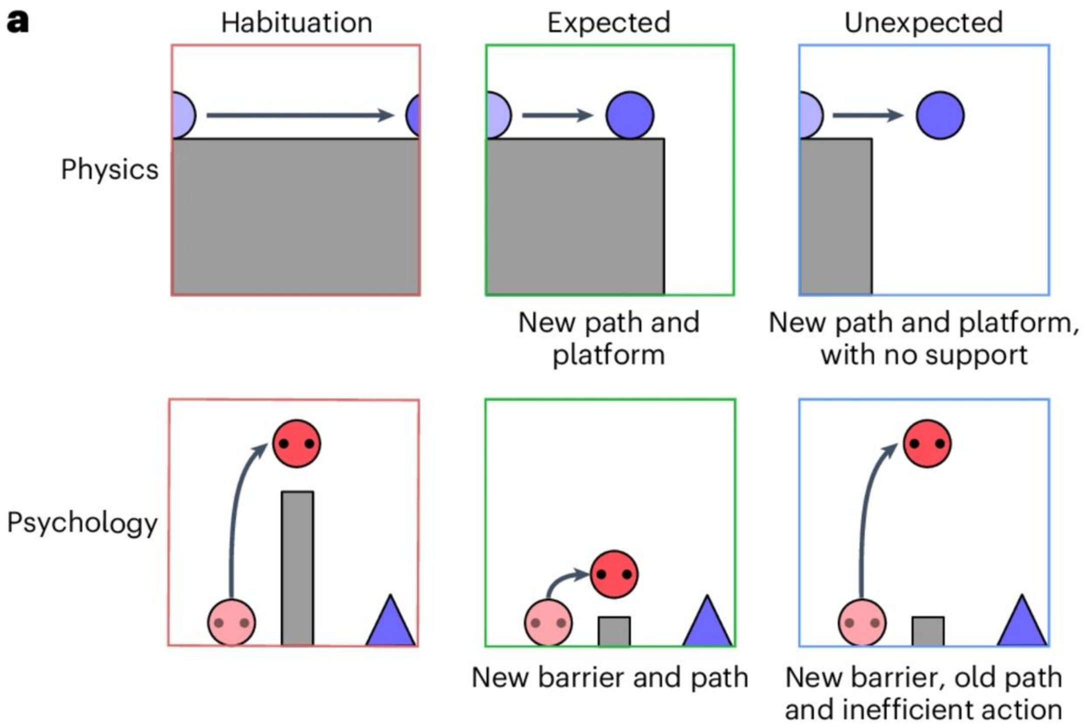


Fig. 1a, Kunin *et al.* (2024)
Nature Human Behavior

Domain-specific

Infants possess fragile, specific innate knowledge

This can be sophisticated knowledge but might break under seemingly trivial circumstances

Learning plays the critical role of optimizing these basic building blocks

Domain-general

Infants possess learning mechanisms that allow them to quickly acquire sophisticated knowledge

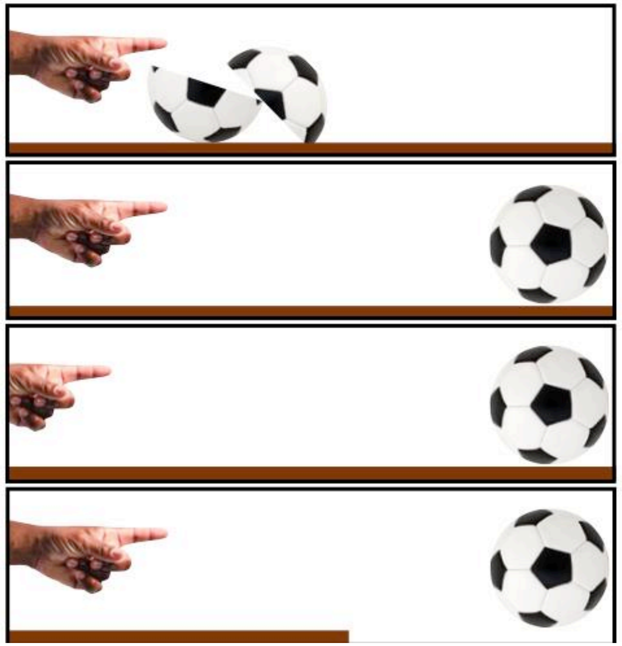
Experience is important but not necessary since infants can make inferences

These learning and reasoning mechanisms are innate

Core knowledge

Infants are born with knowledge for domains that are evolutionarily important for foundational skills

Elizabeth Spelke



Objects

Numbers

Places

Agents

Plus: Forms, social reasoning…

Statistical learning

A foundational, rapid cognitive mechanism enabling babies to detect structure, patterns, and probabilities in their environment

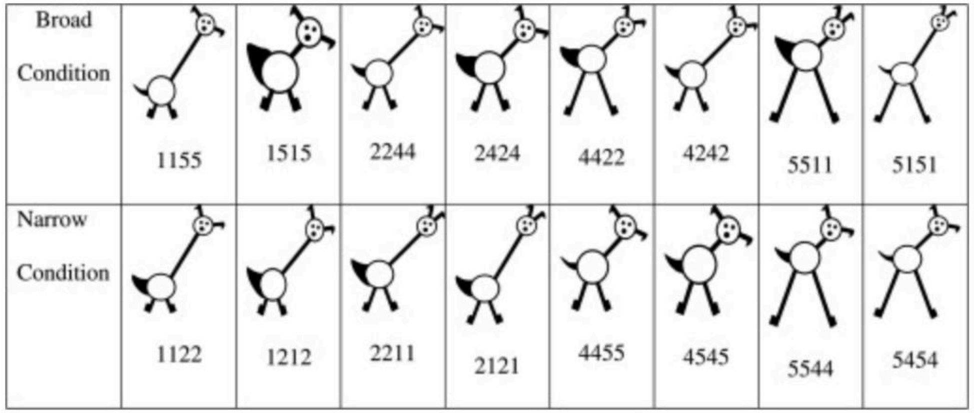
Jenny Saffran

Statistical Learning by 8-Month-Old Infants

Jenny R. Saffran, Richard N. Aslin, Elissa L. Newport

Learners rely on a combination of experience-independent and experience-dependent mechanisms to extract information from the environment. Language acquisition involves both types of mechanisms, but most theorists emphasize the relative importance of experience-independent mechanisms. The present study shows that a fundamental task of language acquisition, segmentation of words from fluent speech, can be accomplished by 8-month-old infants based solely on the statistical relationships between neighboring speech sounds. Moreover, this word segmentation was based on statistical learning from only 2 minutes of exposure, suggesting that infants have access to a powerful mechanism for the computation of statistical properties of the language input.

Science, 1996



Plunkett *et al.*, 2008

Learning an underlying distributional structure – labels can interact with the perceptual learning

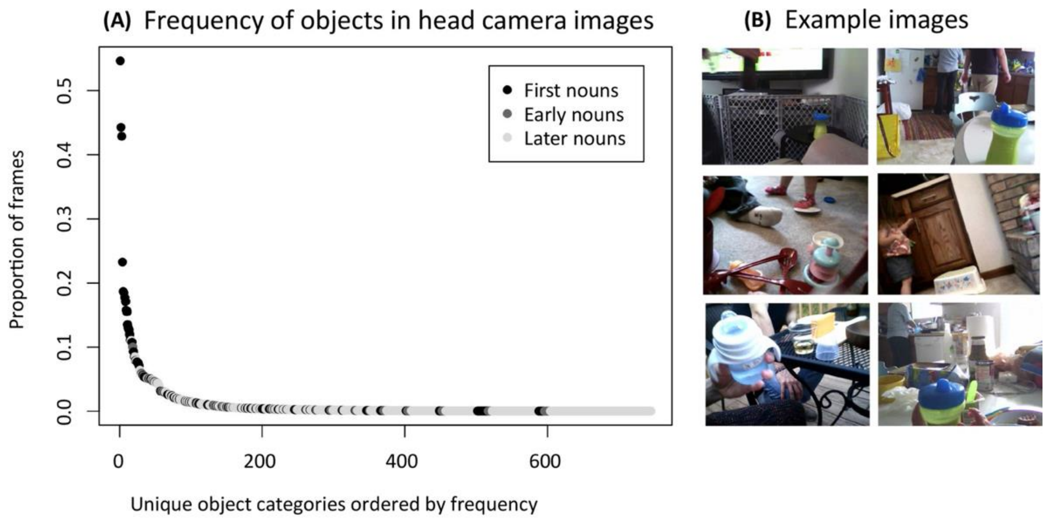
Unique perception in infants

Visual acuity and color perception are poor

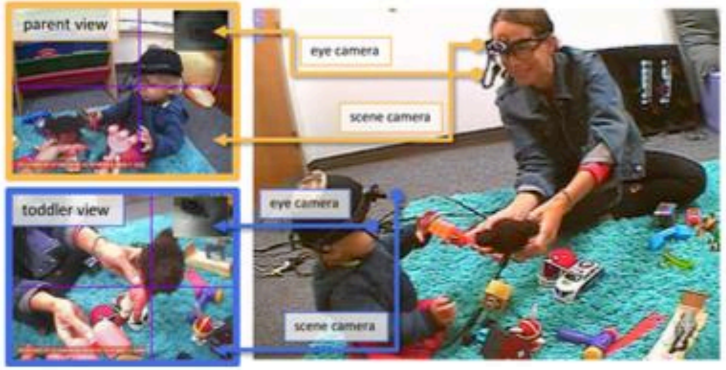
Unique perception in infants

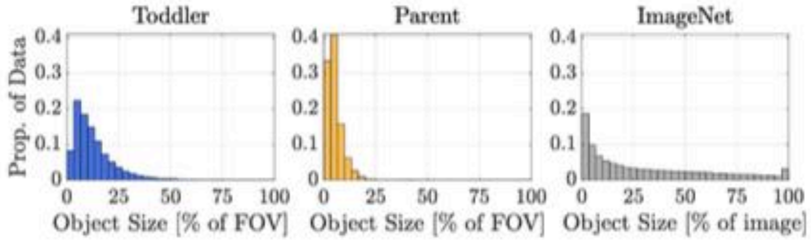
Visual acuity and color perception are poor

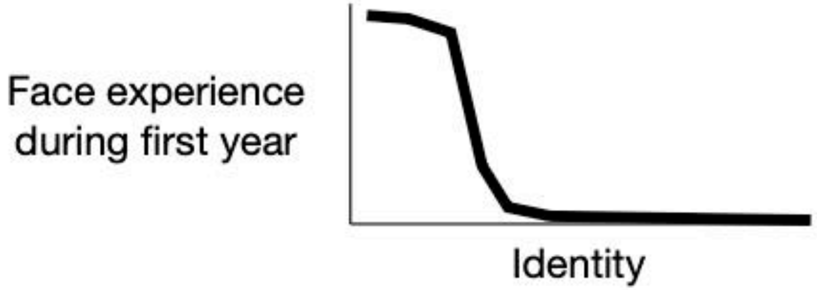
Infants see few objects/faces often



Smith et al., 2018







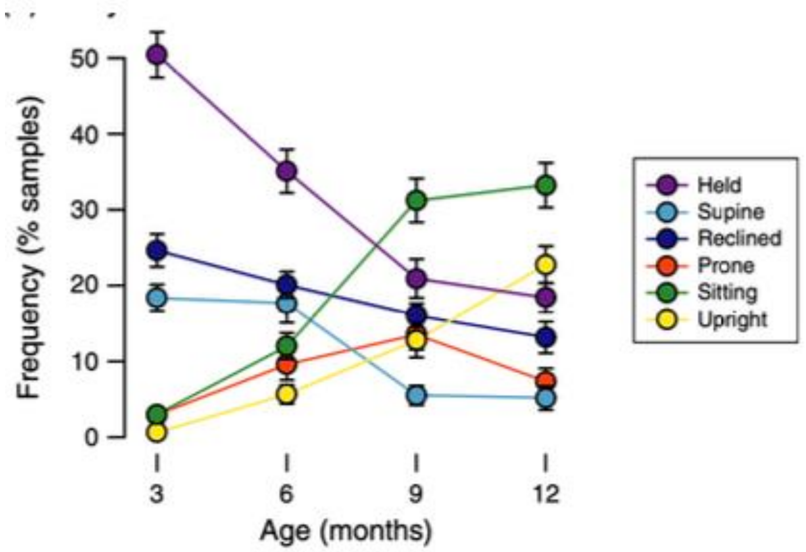
Bambach, et al., 2018; Jayaraman, et al., 2019

Unique perception in infants

Visual acuity and color perception are poor

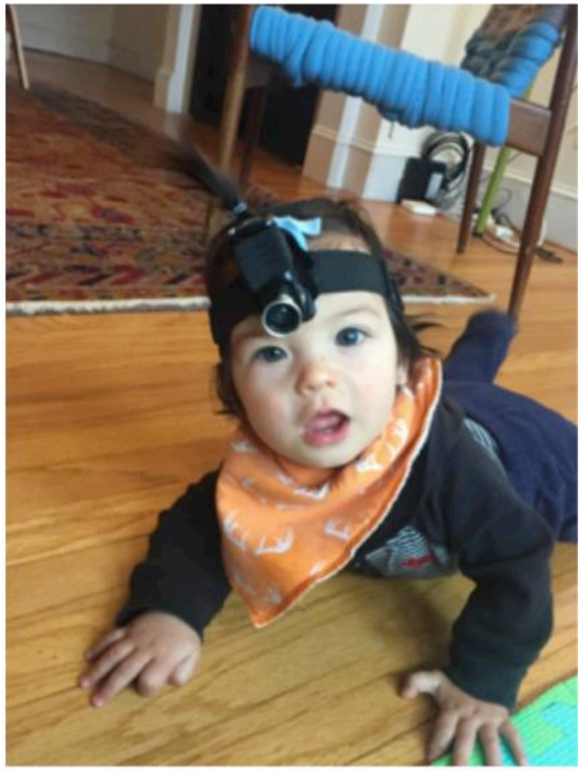
Infants see few objects/faces often

They are lying on their backs most of the time

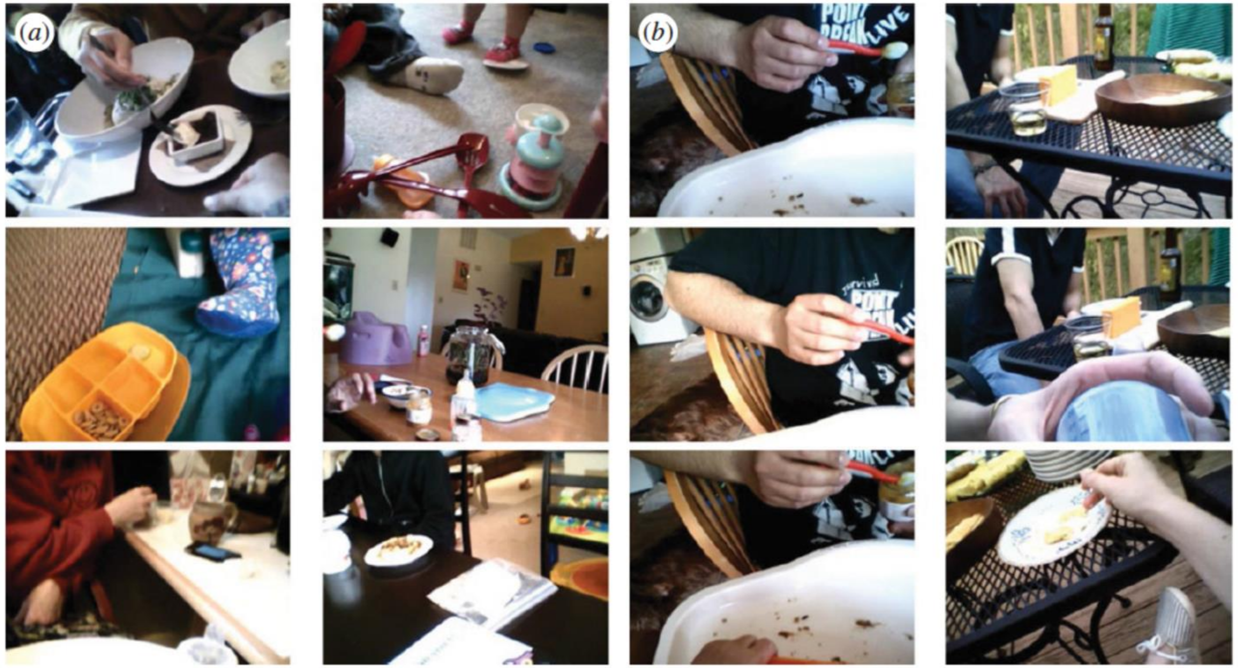


Franchak, et al., 2019

Headcam studies



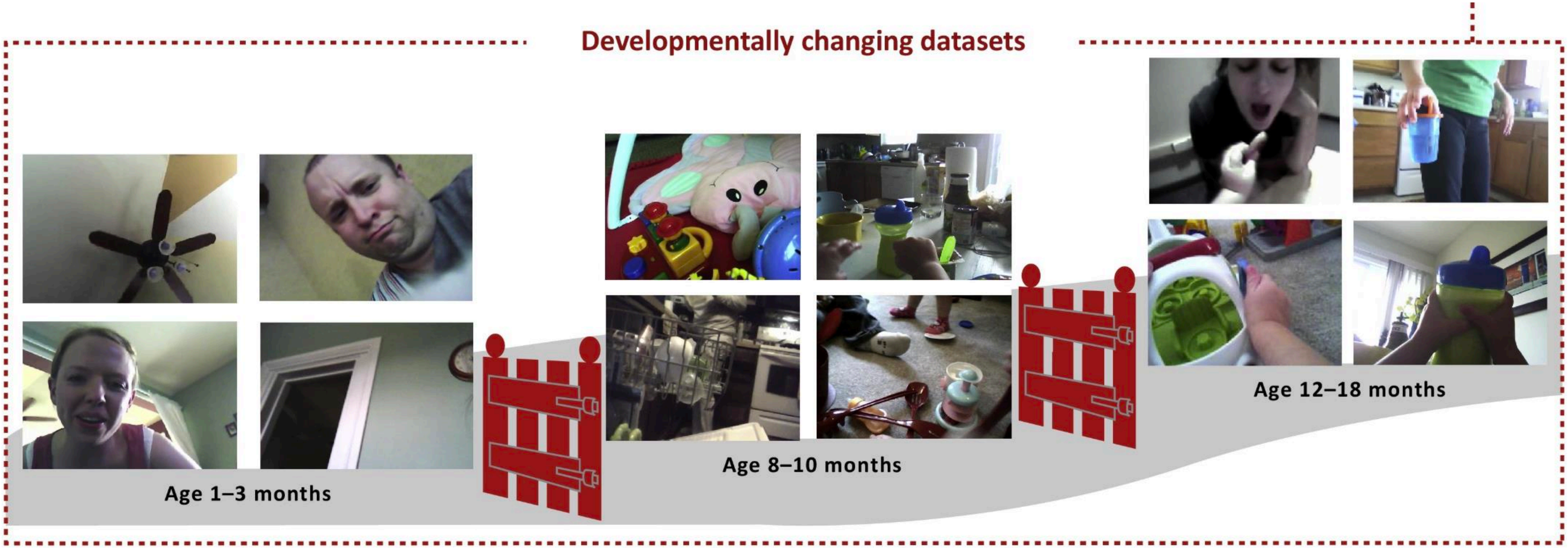
This is more like what real visual experience looks like:



Clerkin, Hart, Rehg, Yu, & Smith (2017)

The Developing Infant Creates a Curriculum for Statistical Learning

Linda B. Smith · Swapnaa Jayaraman · Elizabeth Clerkin · Chen Yu



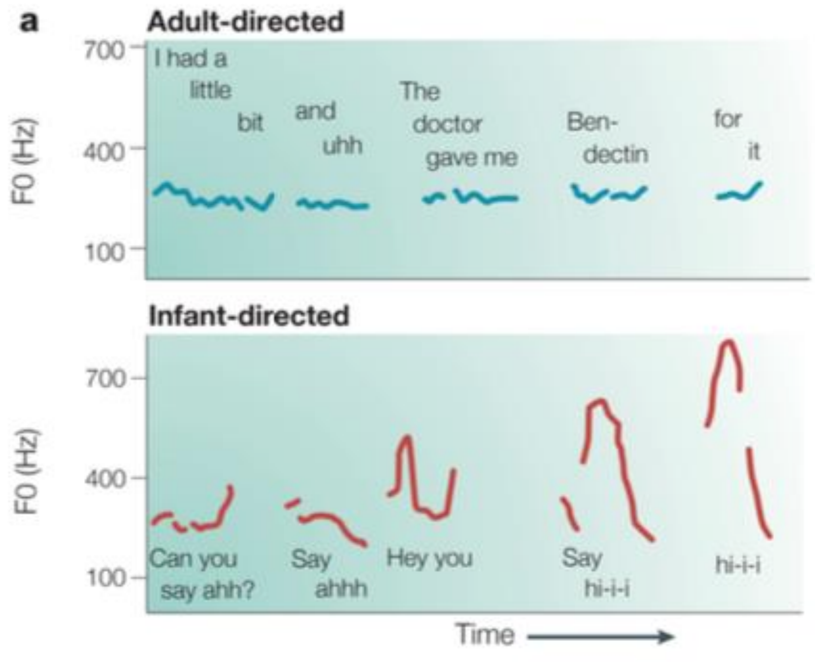
Unique perception in infants

Visual acuity and color perception are poor

Infants see few objects/faces often

They are lying on their backs most of the time

Infant directed speech is unique



Fernald & Kuhl, 1987

Domain-specific

Infants possess fragile, specific innate knowledge

This can be sophistica
but might break under
circumstances

Learning plays the critical role of optimizing these basic building blocks

We need to figure out the relative importance of each

Domain-general

Infants possess learning mechanisms
hem to quickly acquire
ed knowledge

is important but not
since infants can make

These learning and reasoning mechanisms are innate

Behavior is highly constraining of the brain.

But how does that behavior arise?

Domain-specific

Infants possess fragile, specific innate knowledge

This can be sophistica
but might break under
circumstances

Learning plays the critical role of optimizing these basic building blocks

D

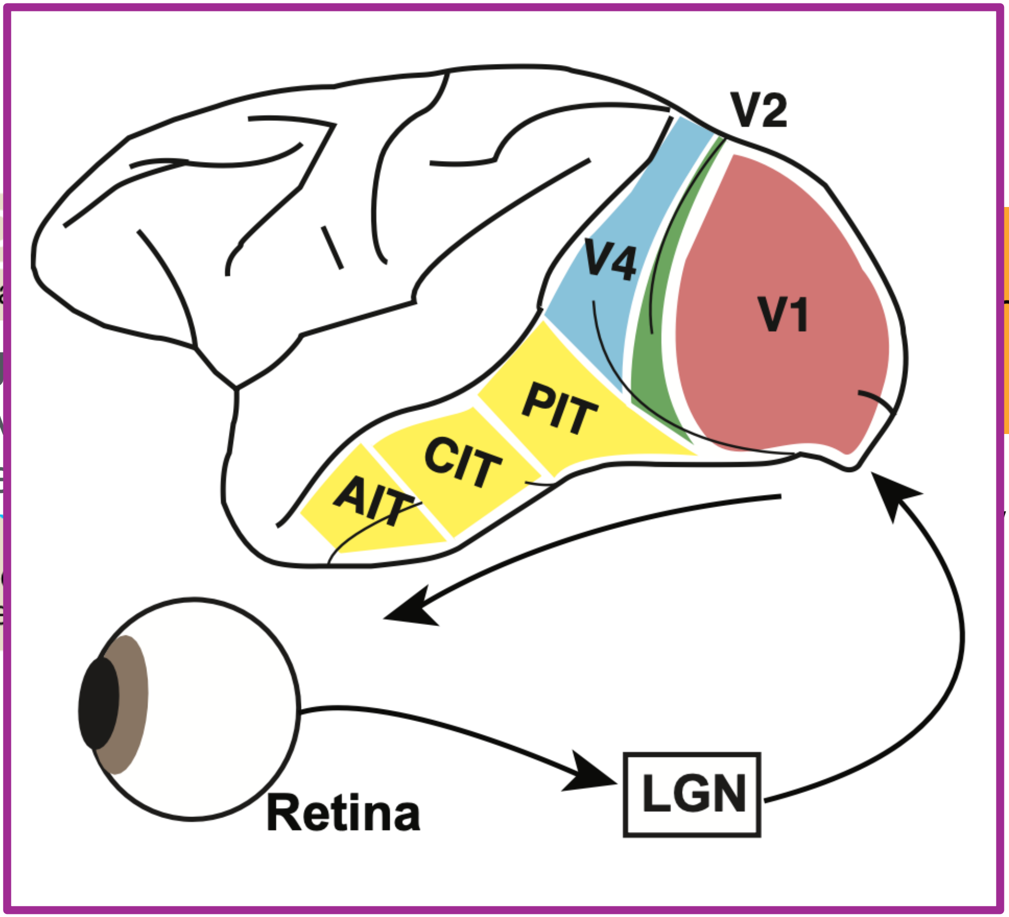
Infa

Th
me

We need to fig
out the relativ
importance of e

ighly constraining of the brain.

does that behavior arise?



Developmental Cognitive Neuroscience

Infant neuroimaging

EEG

- Good temporal resolution
- Poor spatial resolution

MEG

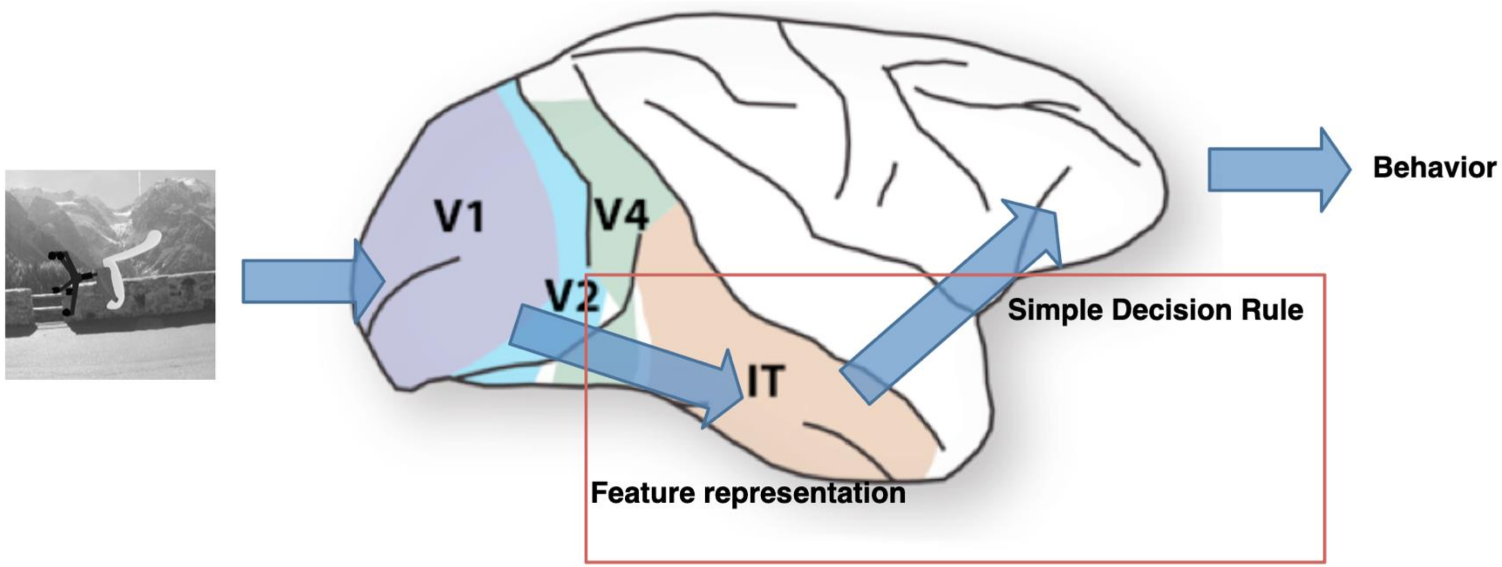
- Temporal + spatial
- Emerging technology in OPM-MEG

Fig. 1c, Corvilain *et al.* (2025) Imaging Neurosci

fNIRS

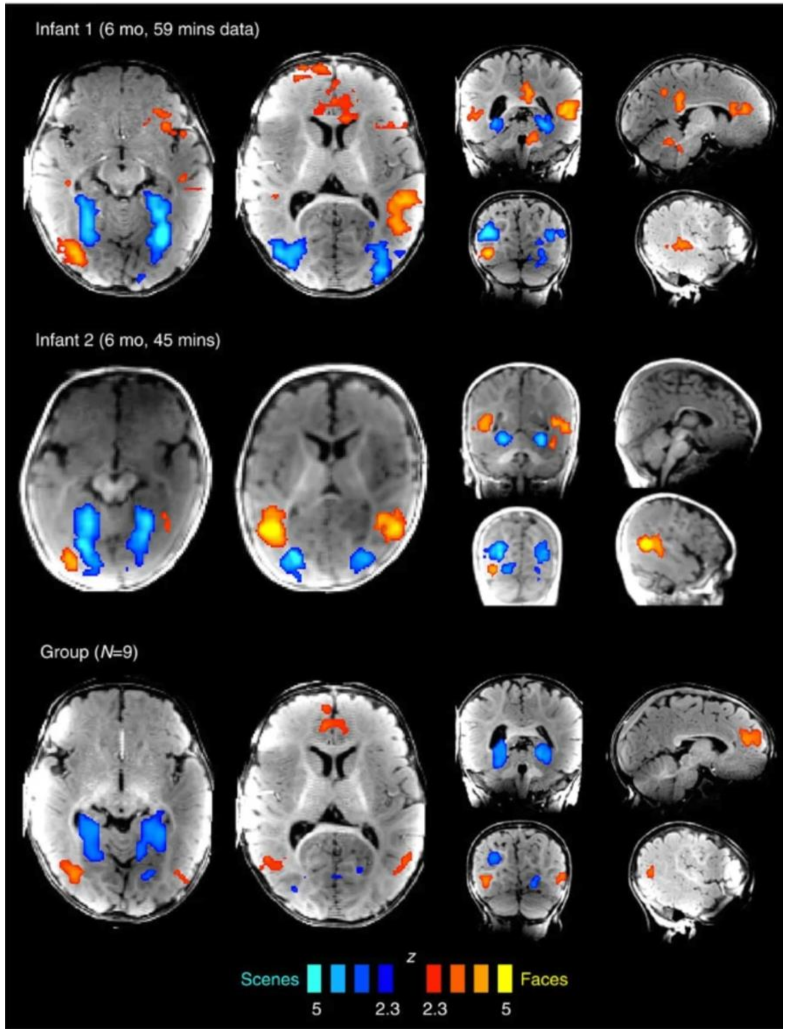
- Better tolerated, easier to use in naturalistic setting
- Localisation not great

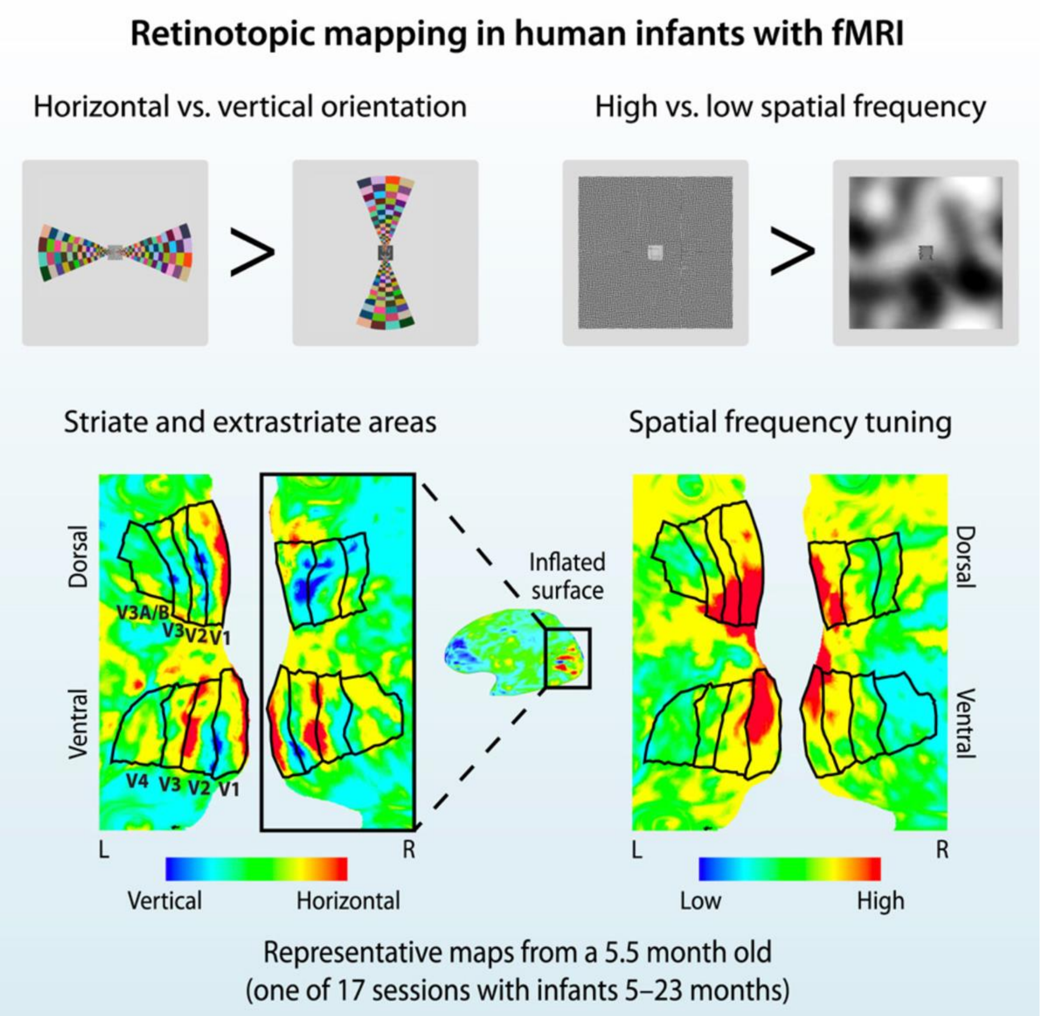
Fig. 2c, Gervain *et al.* (2023) Neurophotonics



These methods make it difficult to study representations on the ventral surface of the brain

- But is MRI feasible in awake, behaving infants?





Deen *et al.*, 2017
Kosakowski *et al.*, 2022

Ellis *et al.*, 2020
Ellis *at al.*, 2021

EEG

- Good temporal resolution

MEG

- Temporal + spatial
- Emerging technology in OPM-MEG

Fig. 1c, Corvilain *et al.* (2025)
Imaging Neurosci

fNIRS

- Better tolerated, easier to use in naturalistic setting

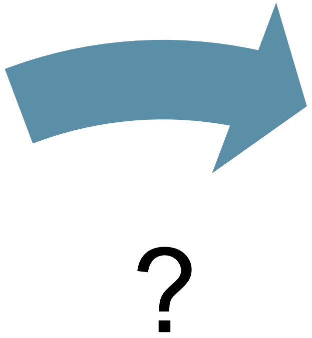
Fig. 2c, Gervain *et al.* (2023)
Neurophotonics

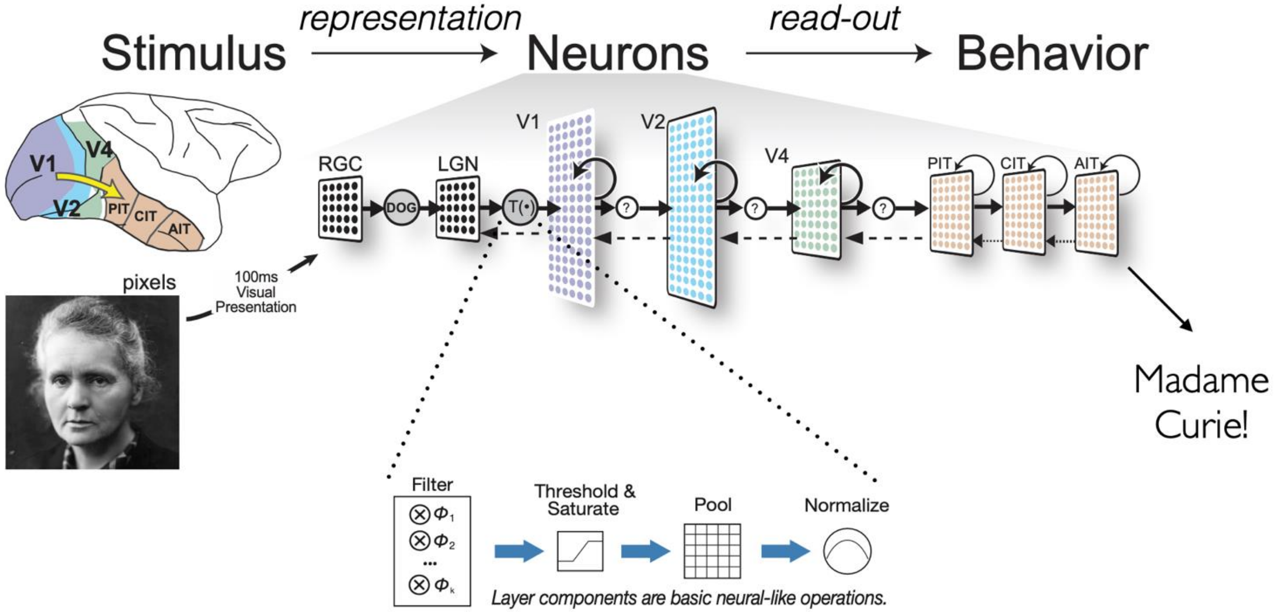
MRI

- High spatial resolution and access to deep brain structures

Brain

Model


?


Stimulus
representation
Neurons
read-out
Behavior
V1
V4
V2
PIT
CIT
AIT
pixels
100ms
Visual
Presentation
RGC
DOG
LGN
T(•)
V1
V2
V4
PIT
CIT
AIT
Madame
Curie!
Filter
Threshold &
Saturate
Pool
Normalize
Layer components are basic neural-like operations.

Overview

- Why should we model development?
- How to study infants? What can we do in early life?
- Recent advances in Developmental NeuroAI.

Is it fair to say that AI is really like a baby?

Overview

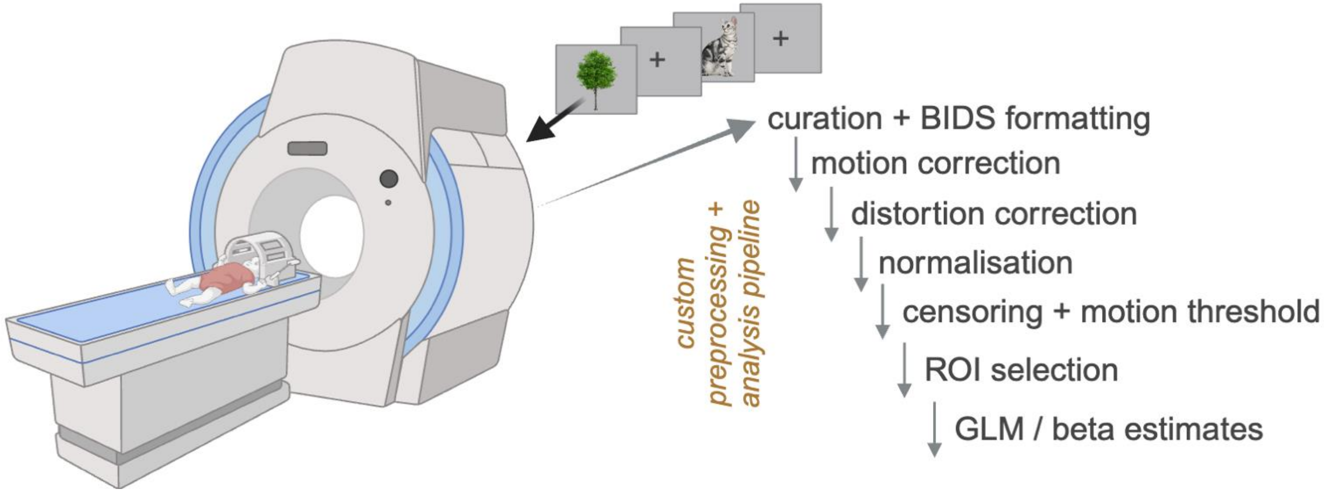
- Why should we model development?
- How to study infants? What can we do in early life?
- Recent advances in Developmental NeuroAI.*** abridged version

Is it fair to say that AI is really like a baby?

Longitudinal awake infant fMRI

Foundations of Cognition
www.foundcog.org

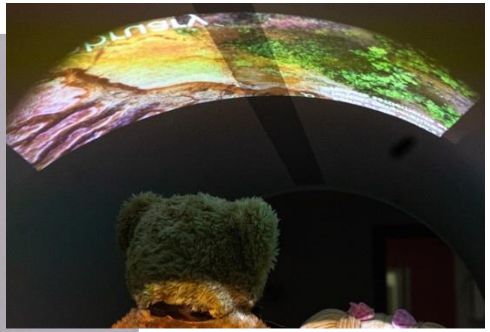
- 2-months (n=130)
- 9-months (n=65)
- adults (n=18)



O'Doherty *et al.*, (2026) Nature Neuroscience

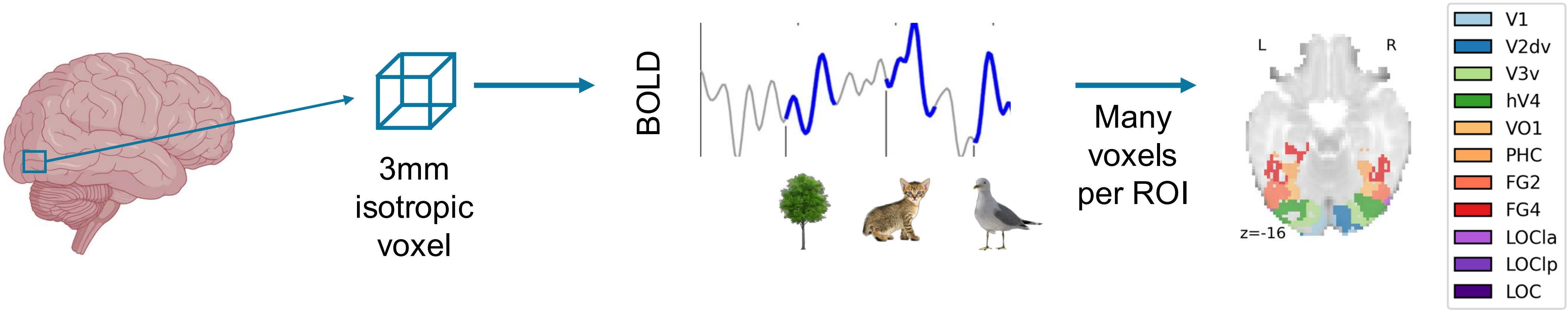
Infant scan setup

Flexible Task switching

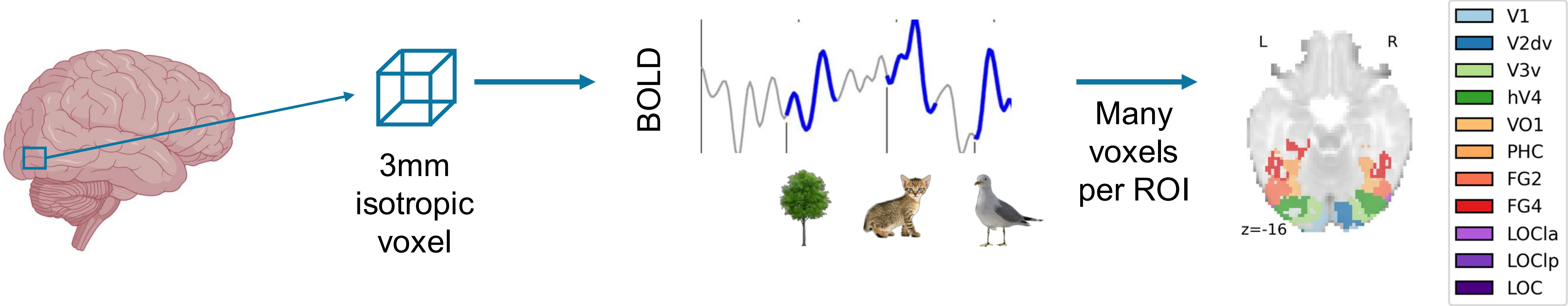


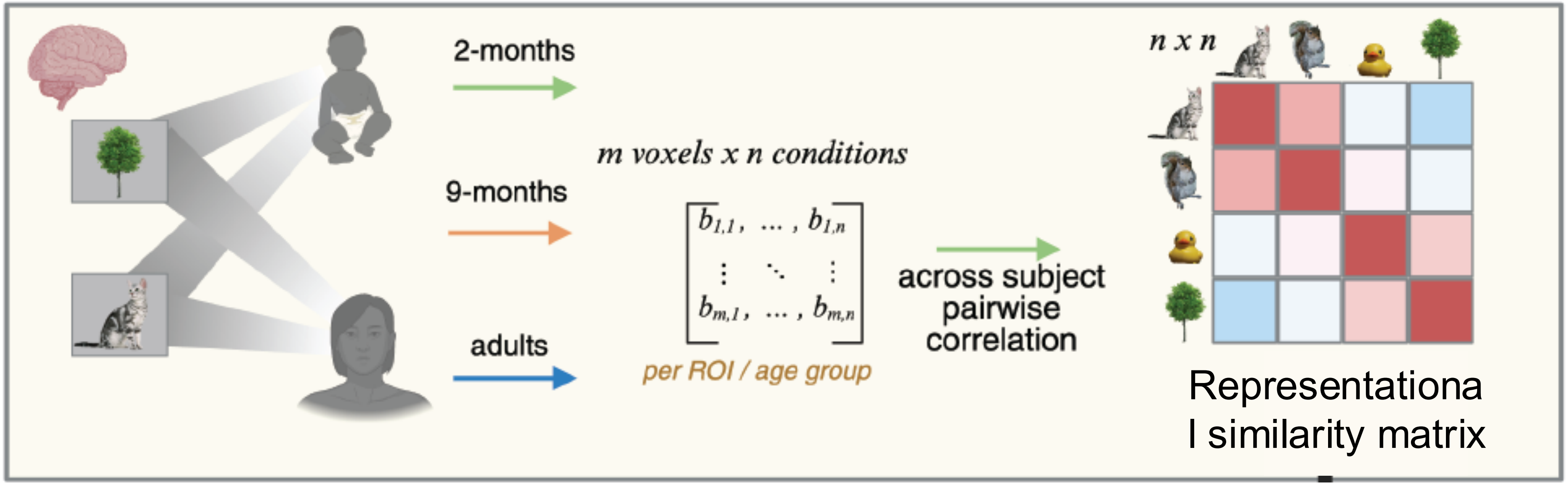
Facial camera recording
- real time monitoring
- retrospective tagging of attentive state

Inspired by Deen *et al.*, 2017; Ellis *et al.*, 2020



3mm
isotropic
voxel
BOLD
Many
voxels
per ROI
L
R
z=-16
V1
V2dv
V3v
hV4
VO1
PHC
FG2
FG4
LOCla
LOClp
LOC

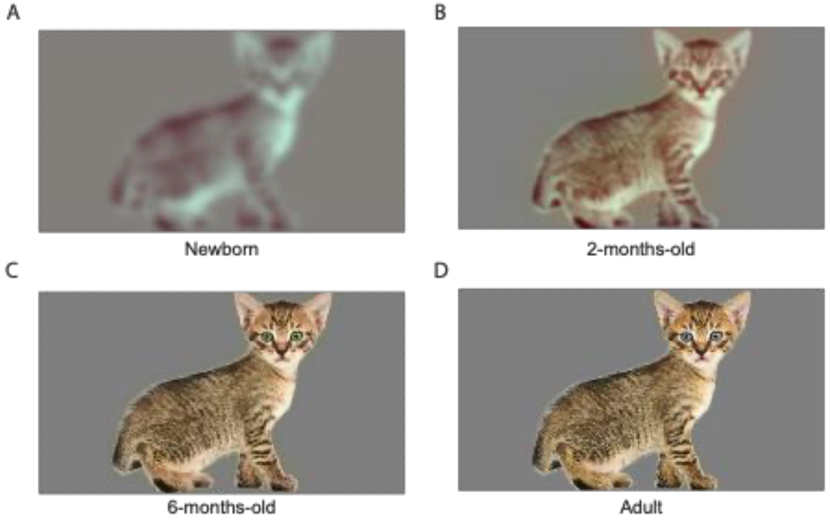




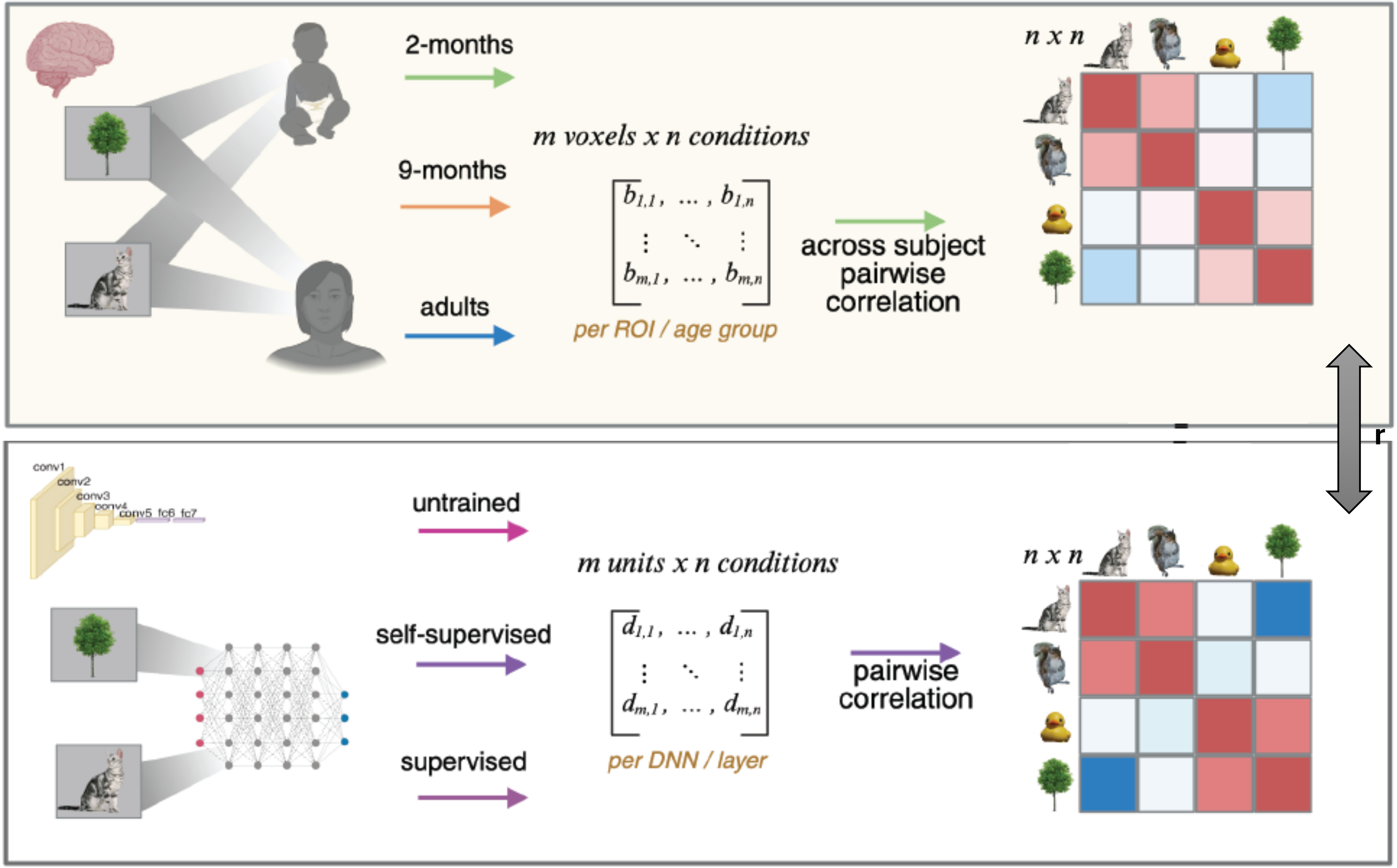
O'Doherty *et al.*, (2026) Nature Neuroscience

Pictures task for MVPA

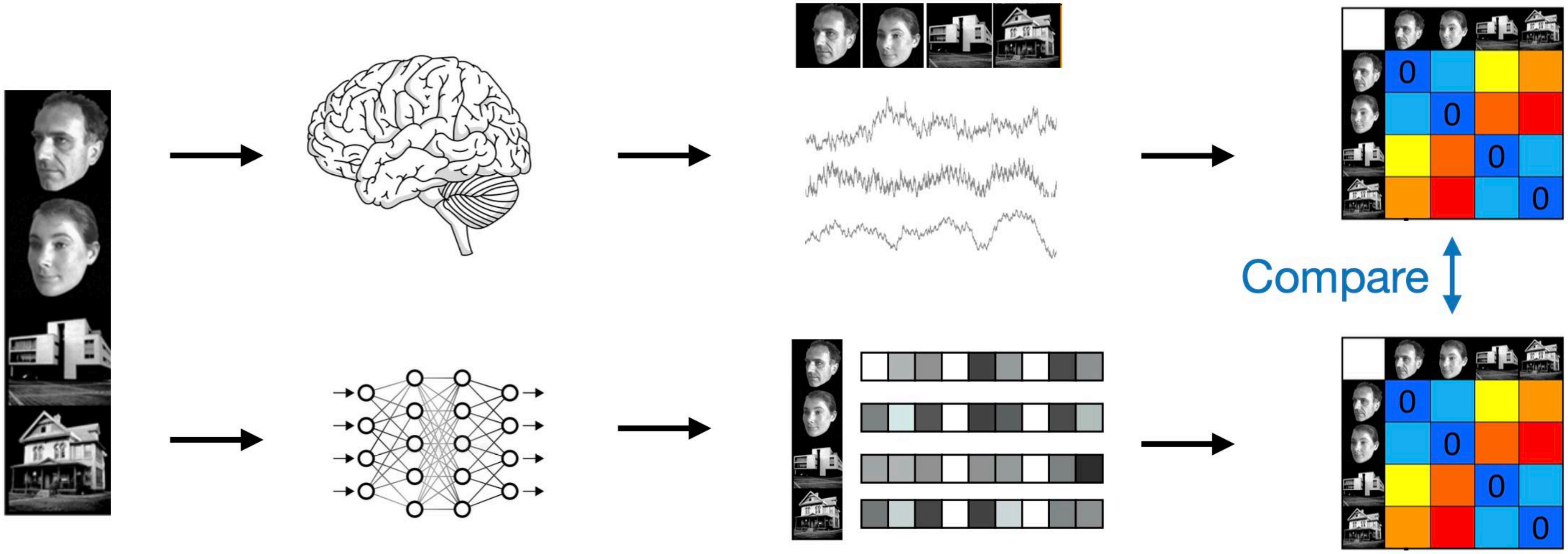
- 36 images
- 12 categories (x3 exemplar each)
    - Chosen across a variety of viewpoints
    - Each relates to video *contexts* in another task
- x4 categories per animate, inanimate small, inanimate large
- Images loom towards the infant
- 3 s presentation, separated by jittered fixation



Simulation using our display parameters (Dineen *et al.*, in prep)
- Tiny Eyes, Alex Wade University of York
- https://github.com/wadelab/VischeckTinyeyes



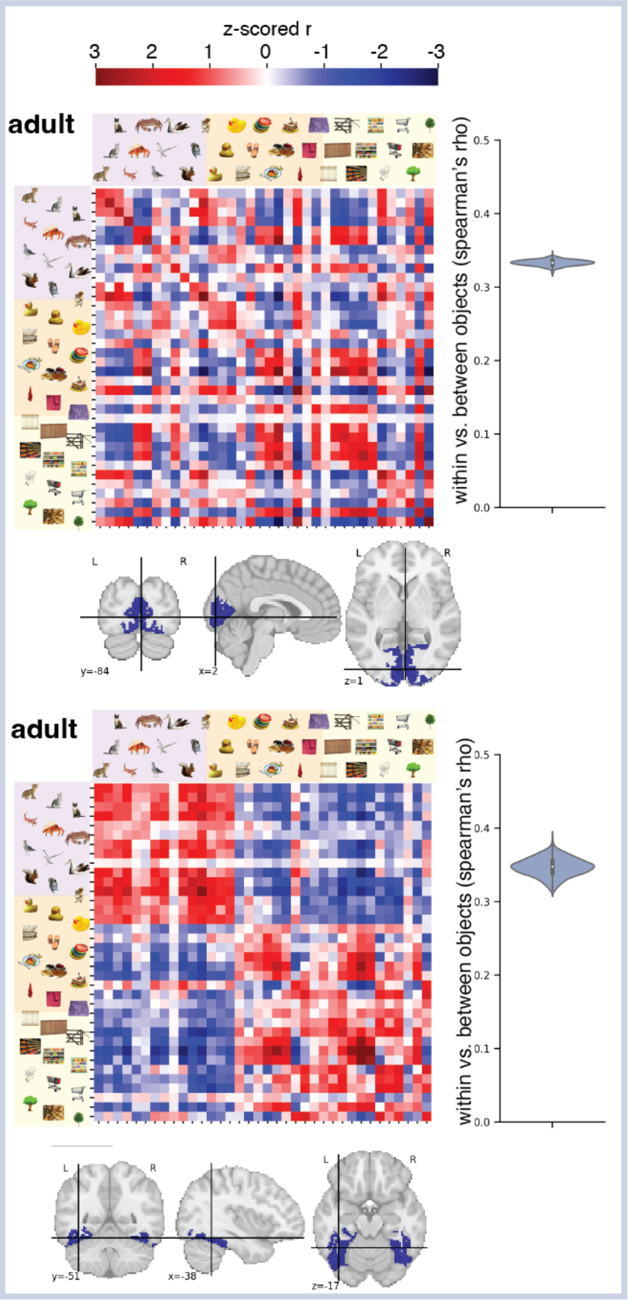
O'Doherty *et al.*, (2026) Nature Neuroscience



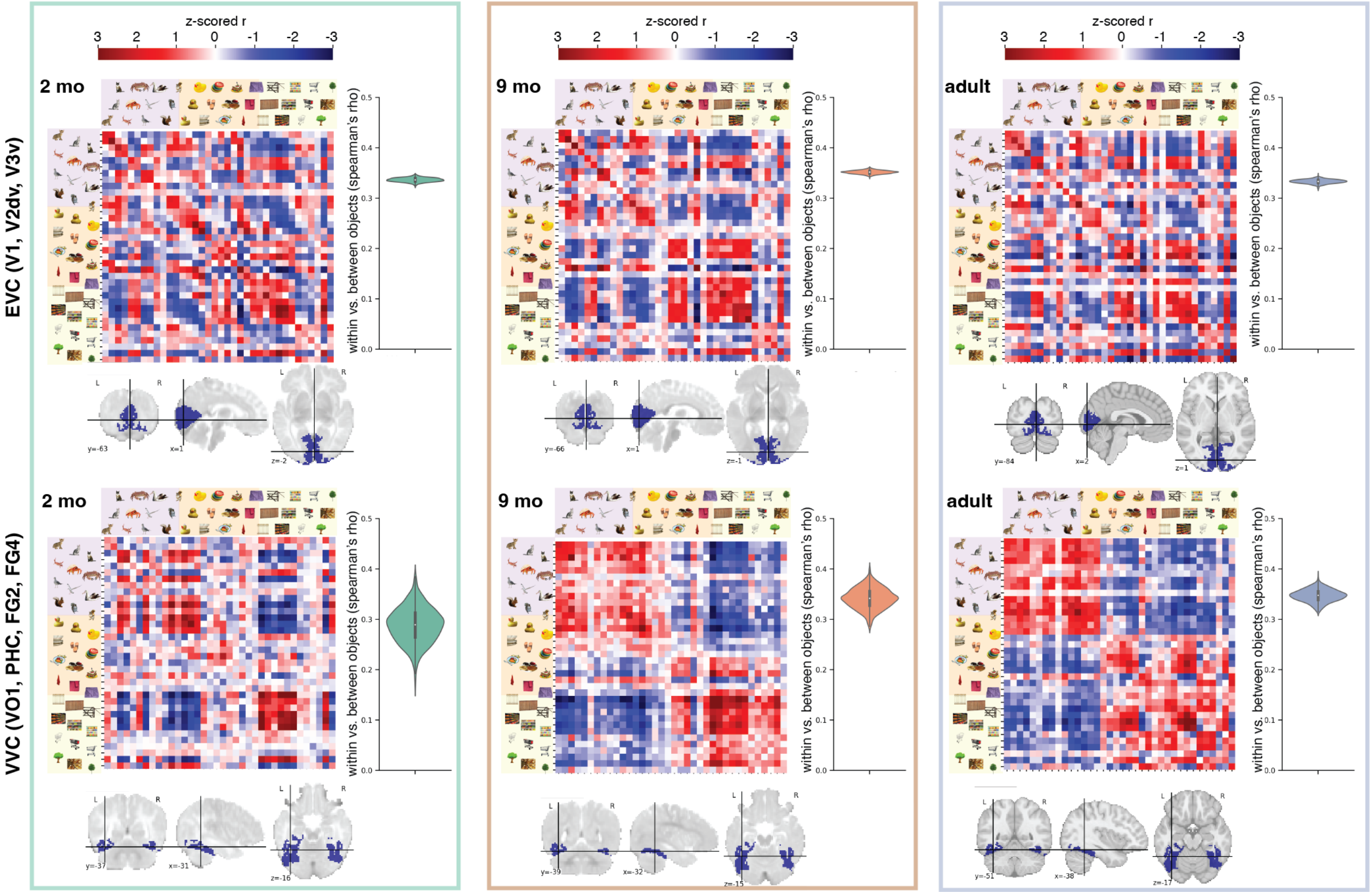
Compare
0
0
0
0
0
0
0
0

EVC (V1, V2dv, V3v)

VVC (VO1, PHC, FG2, FG4)



O'Doherty *et al.*, (2026) Nature Neuroscience



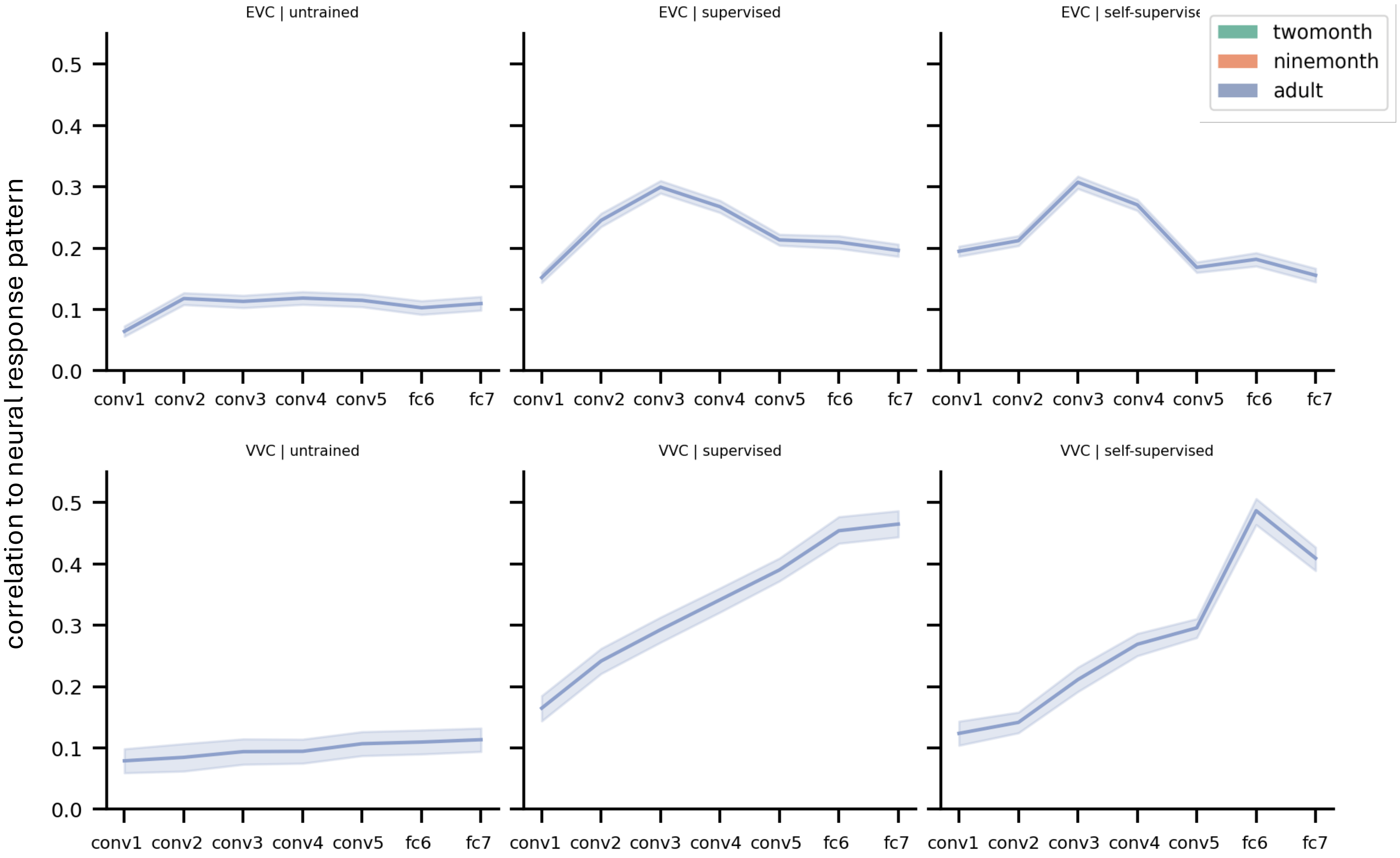
O'Doherty *et al.*, (2026) Nature Neuroscience

DNN Modelling

AlexNet trained on ImageNet

Random initialisation, supervised and self-supervised contrastive learning

Expected hierarchical correspondence between layer and visual hierarchy



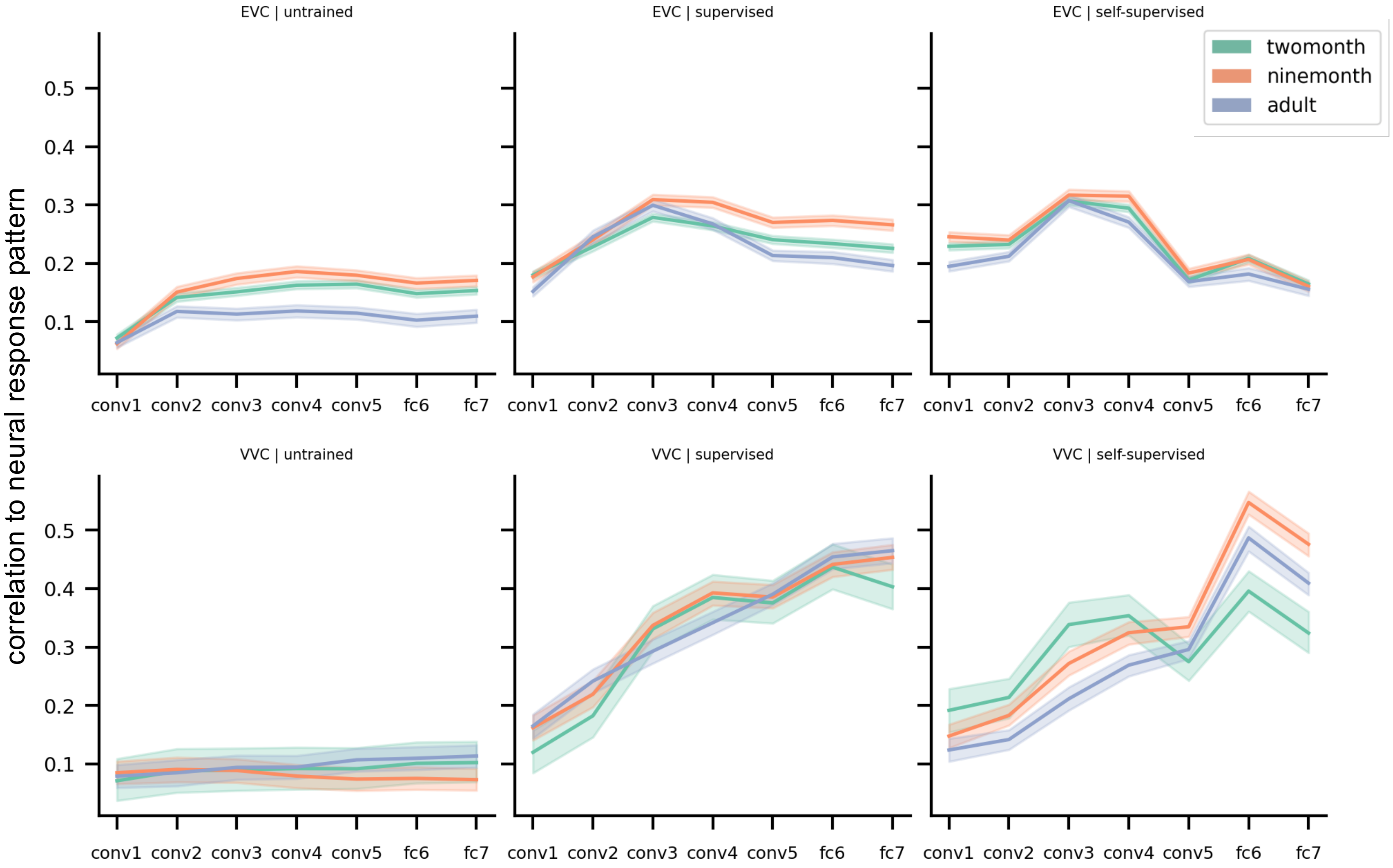
O'Doherty *et al.*, (2026) Nature Neuroscience

DNN Modelling

AlexNet trained on ImageNet

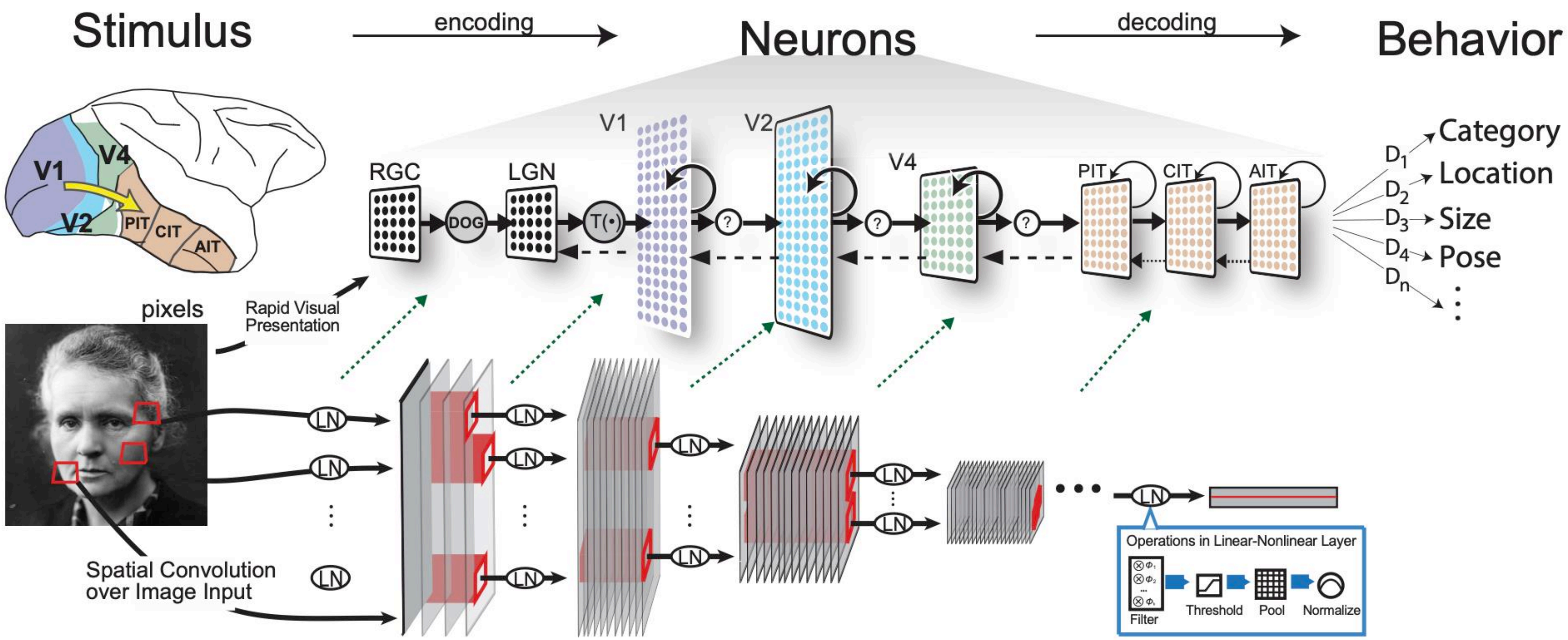
Random initialisation, supervised and self-supervised contrastive learning

Expected hierarchical correspondence between layer and visual hierarchy

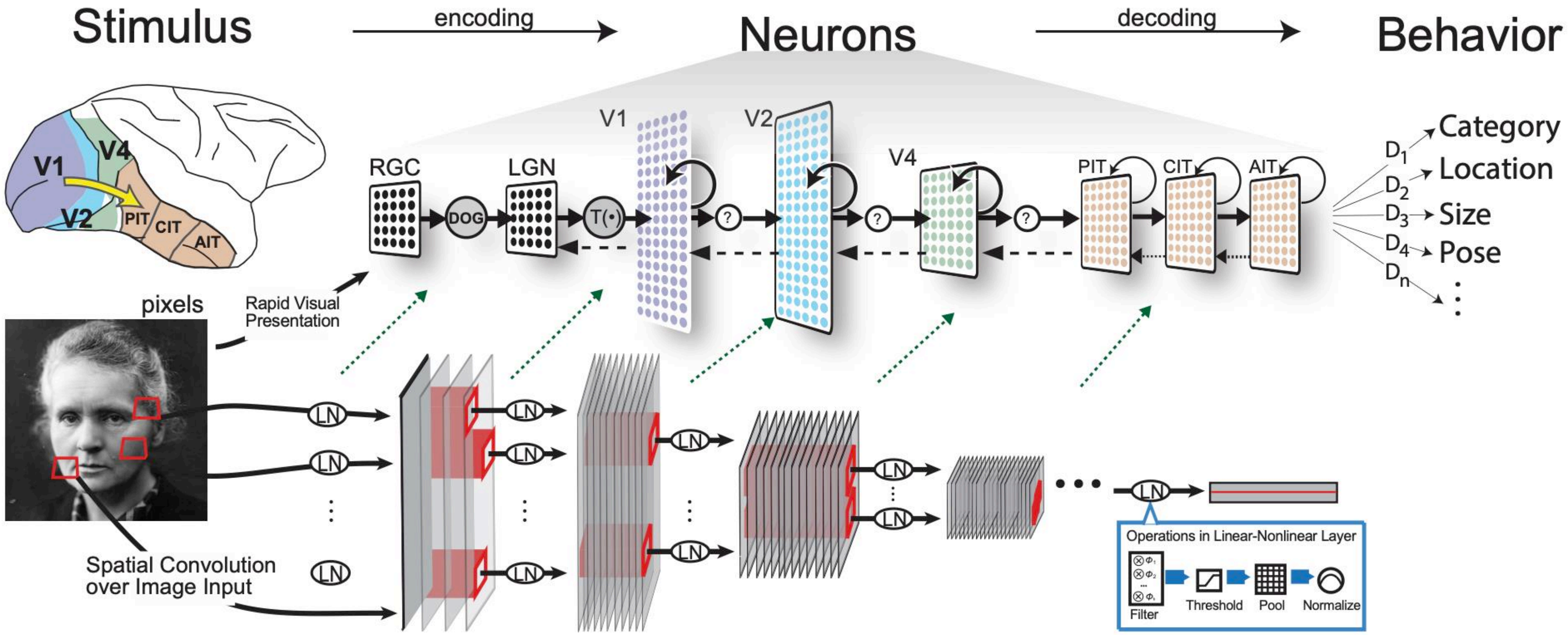


O'Doherty *et al.*, (2026) Nature Neuroscience

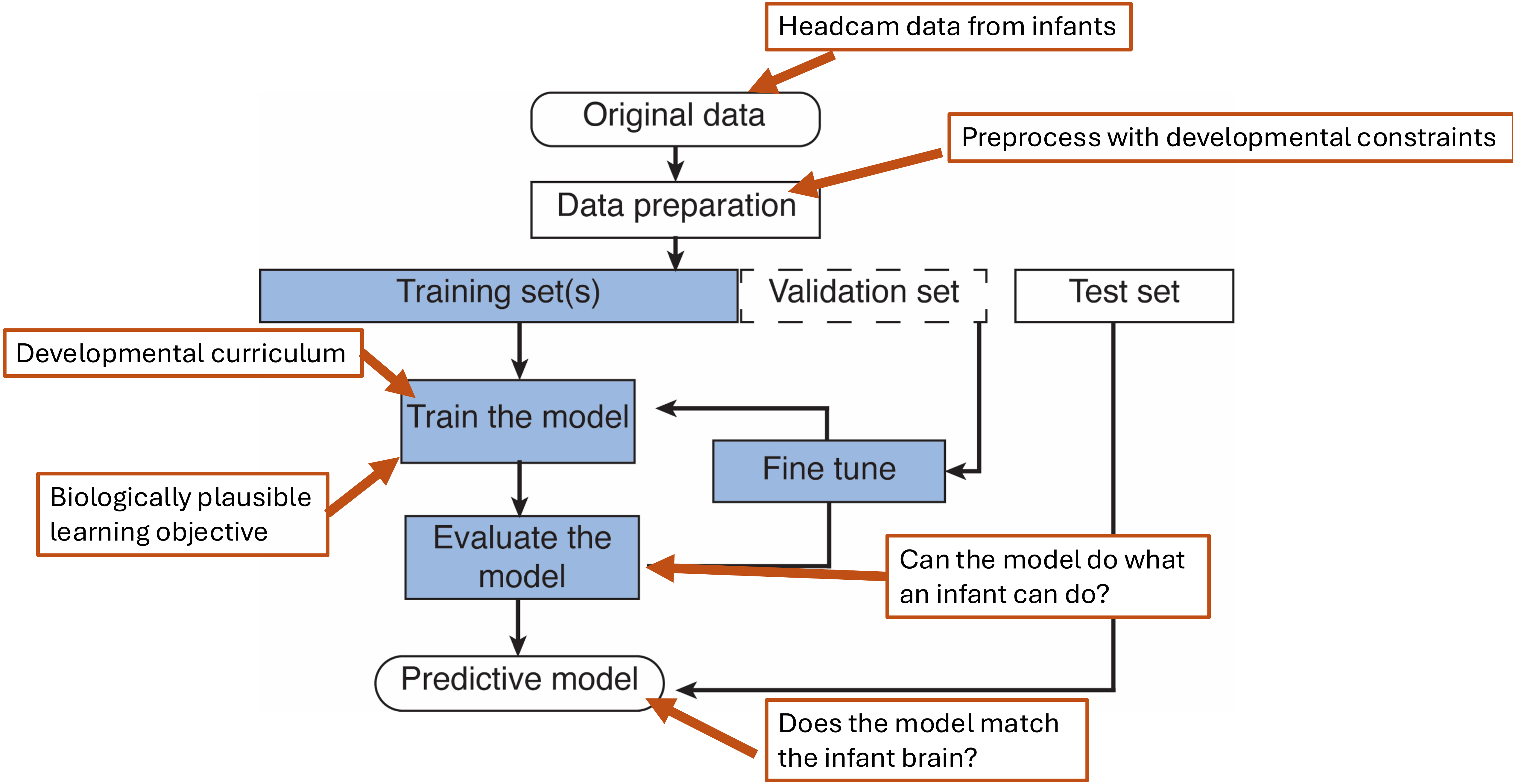
Large-Scale Neural Network Models for Neuroscience

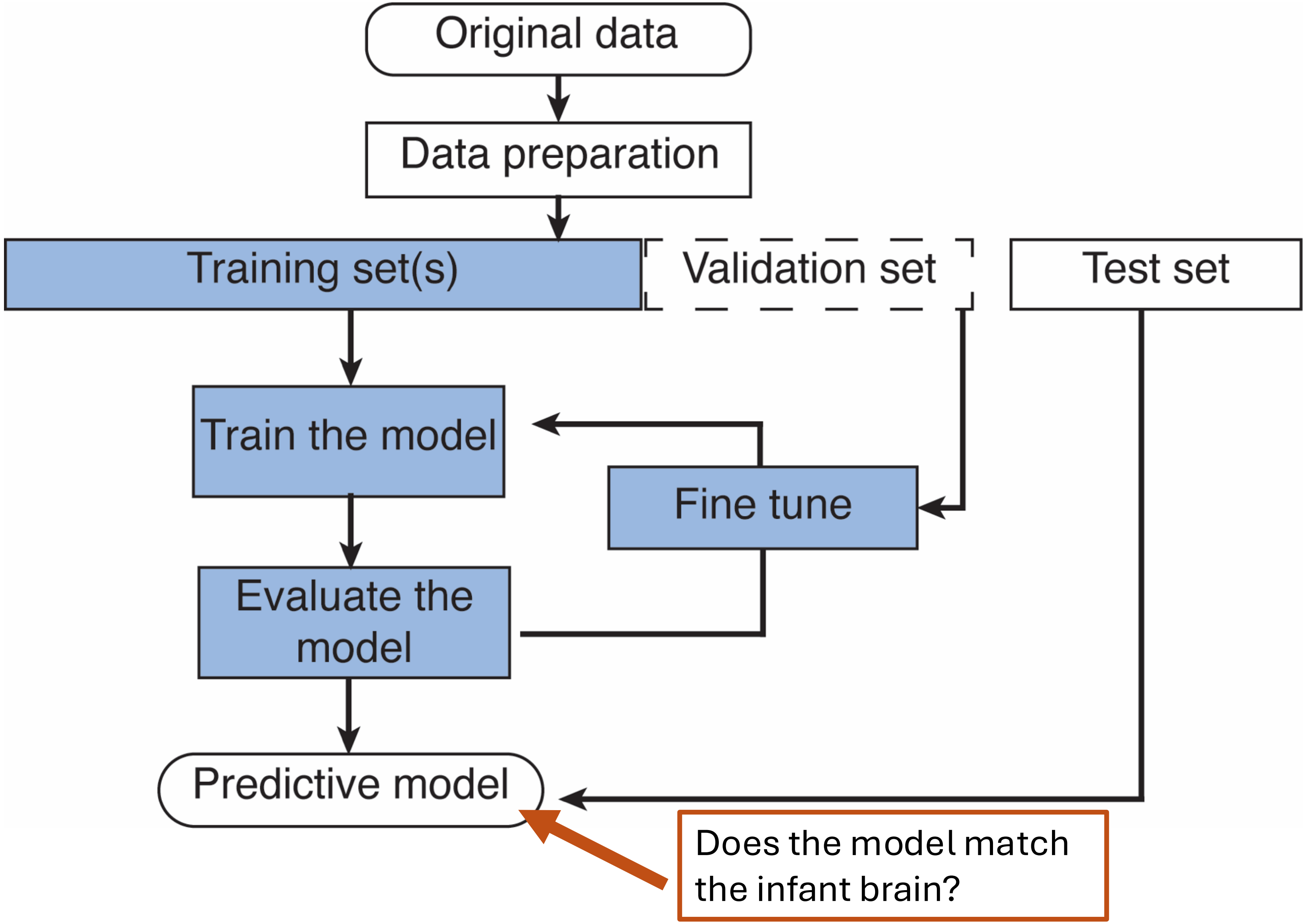


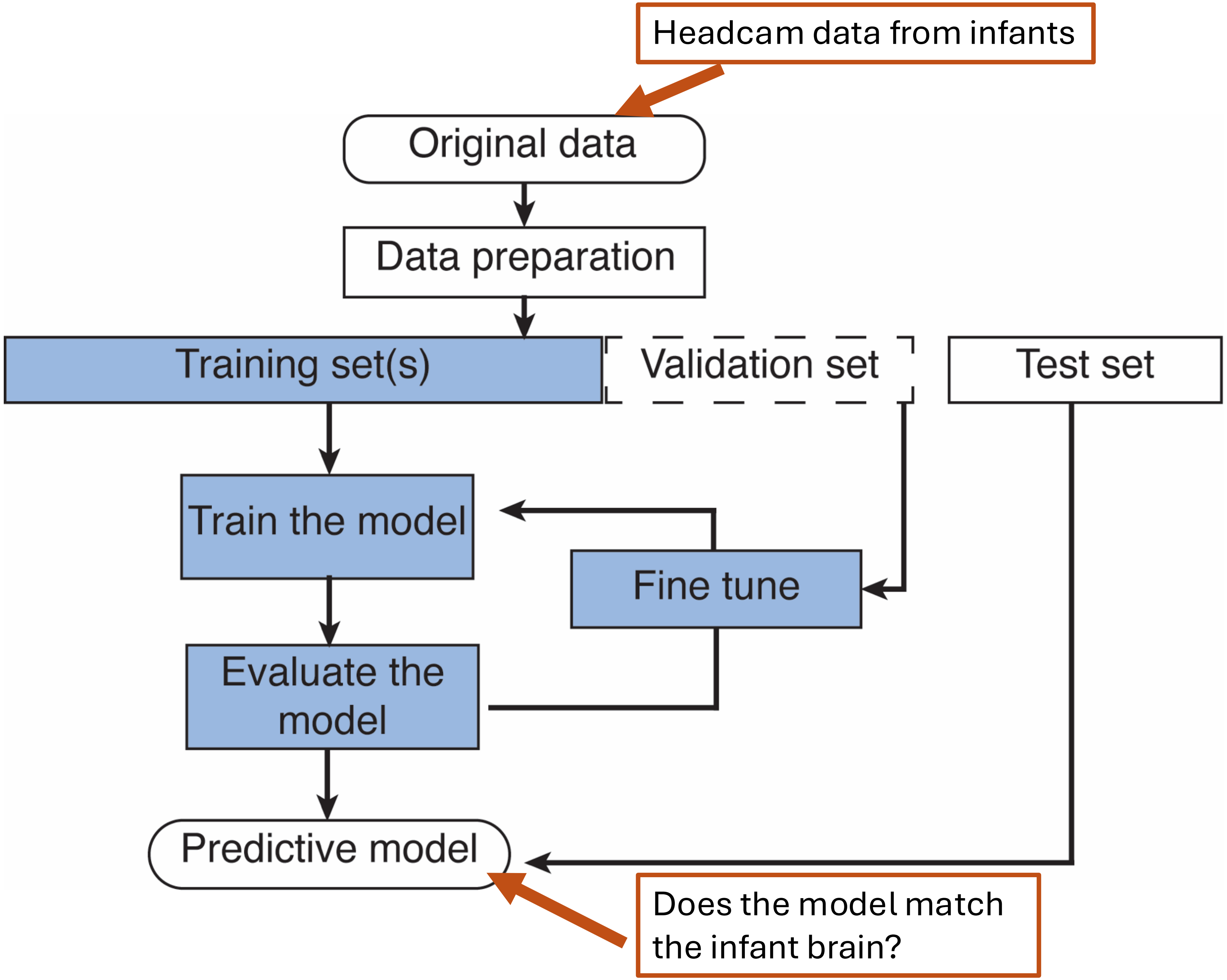
Large-Scale Neural Network Models for Neuroscience



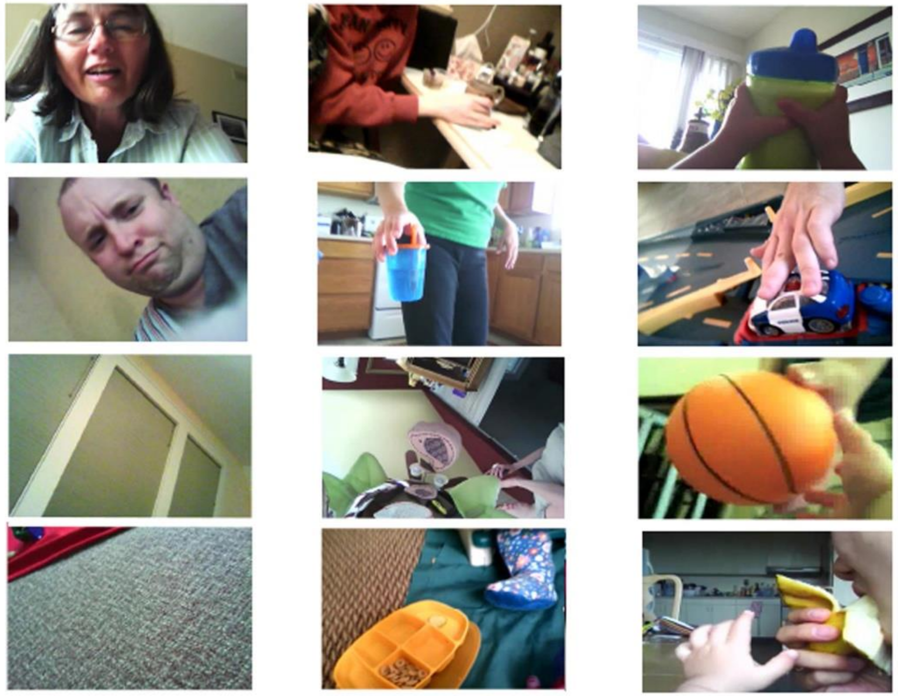
✅ 👍 Works for babies too! 👶


Headcam data from infants
Original data
Preprocess with developmental constraints
Data preparation
Training set(s)
Validation set
Test set
Developmental curriculum
Train the model
Fine tune
Biologically plausible learning objective
Evaluate the model
Can the model do what an infant can do?
Predictive model
Does the model match the infant brain?


Original data
Data preparation
Training set(s)
Validation set
Test set
Train the model
Fine tune
Evaluate the model
Predictive model
Does the model match the infant brain?


Headcam data from infants
Original data
Data preparation
Training set(s)
Validation set
Test set
Train the model
Fine tune
Evaluate the model
Predictive model
Does the model match the infant brain?

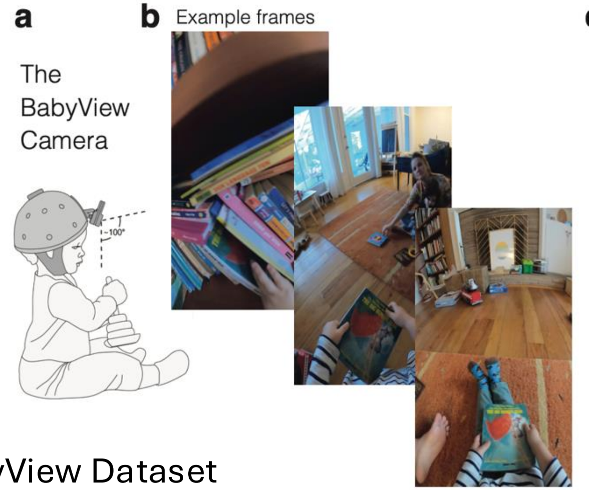
Infants generate interesting datasets for visual learning



1 to 3 months | 8 to 10 months | 12 to 24 months

Smith & Slone (2017)

SAYCam
Sullivan *et al.* (2021)



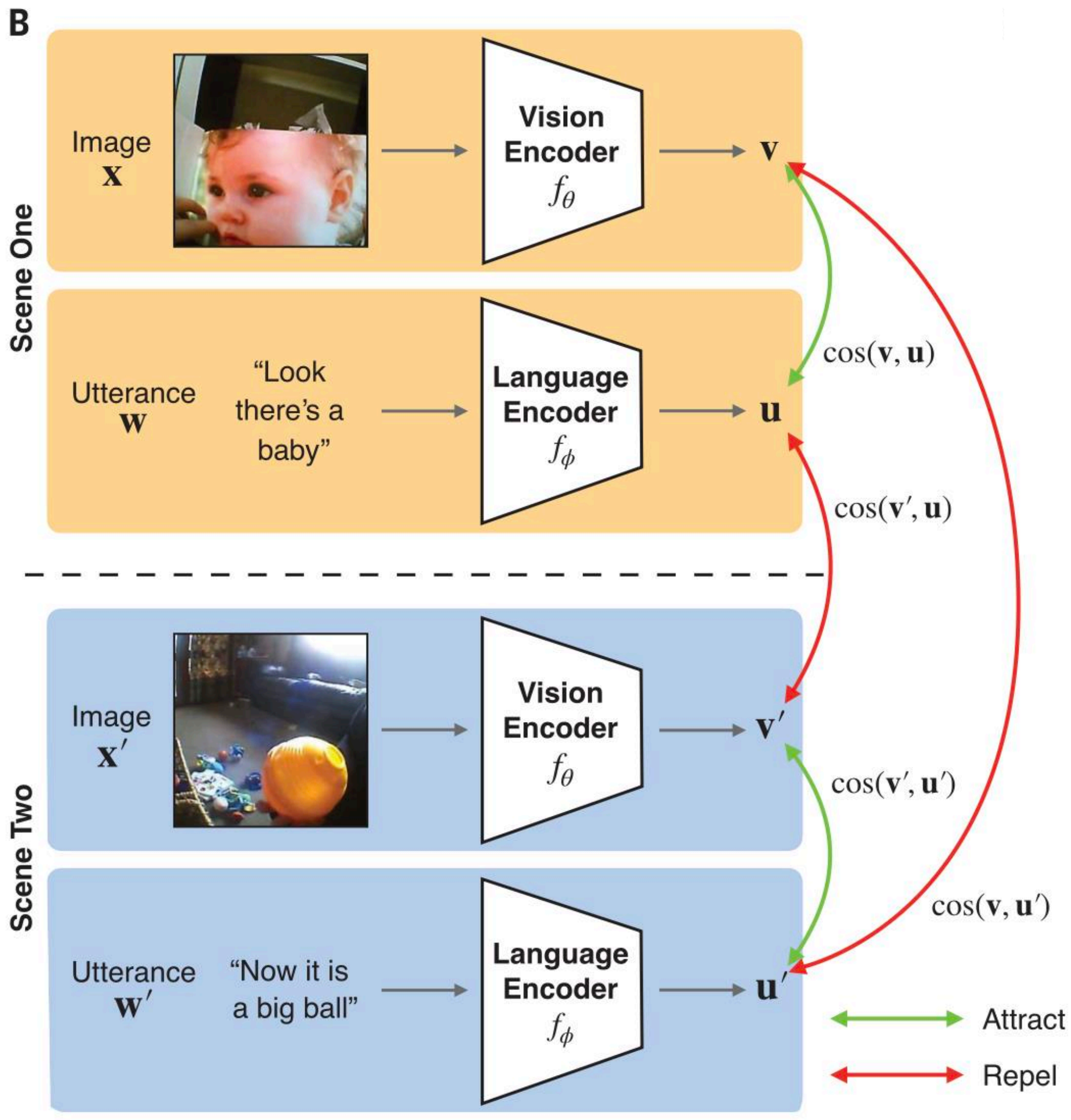
BabyView Dataset
Long *et al.* (2025)

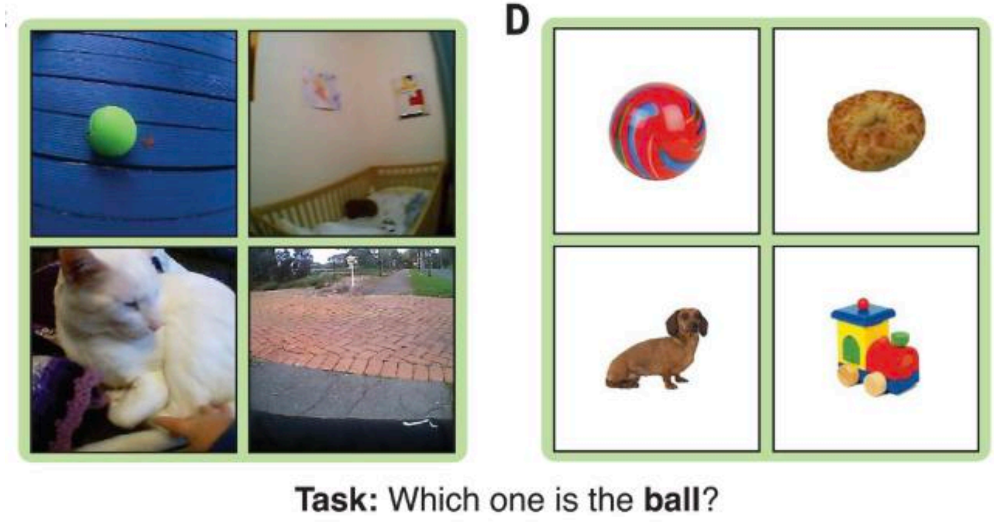
MACHINE LEARNING

Grounded language acquisition through the eyes and ears of a single child

Wai Keen Vong[1]*, Wentao Wang[1], A. Emin Orhan[1], Brenden M. Lake[1,2]

- Image-text model
- ViT encoder head
- Pretrained on headcam data





Assessed on a more dev psych type task

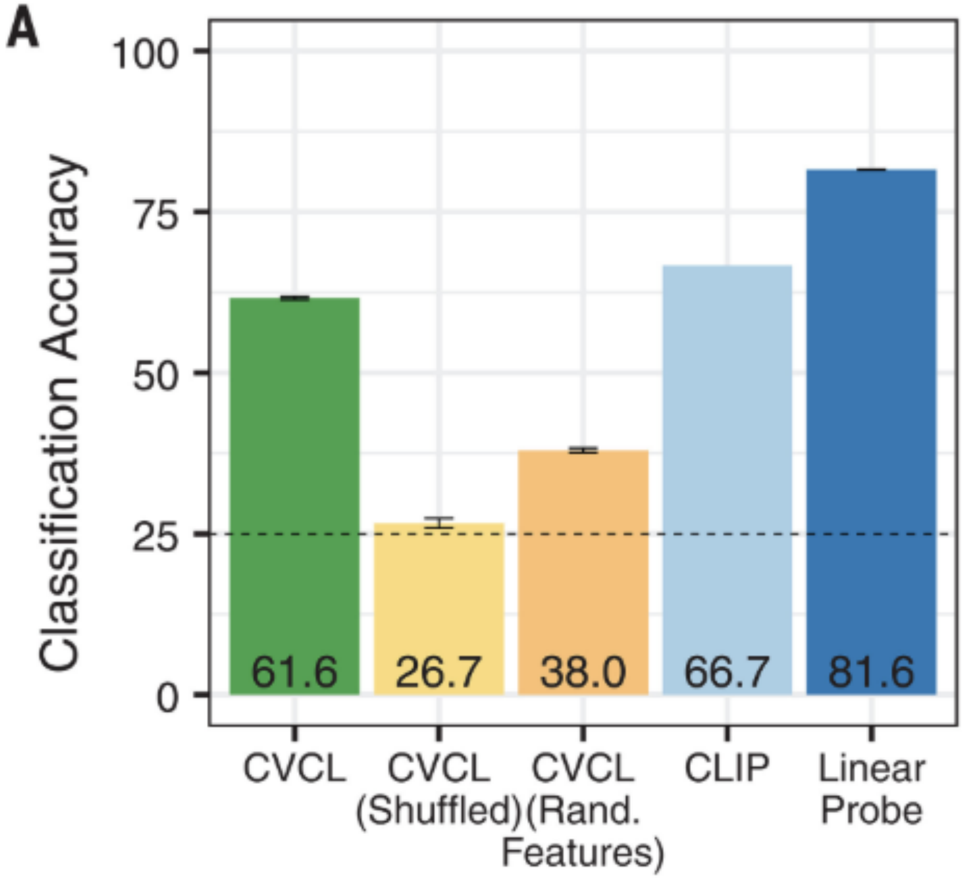
CVCL

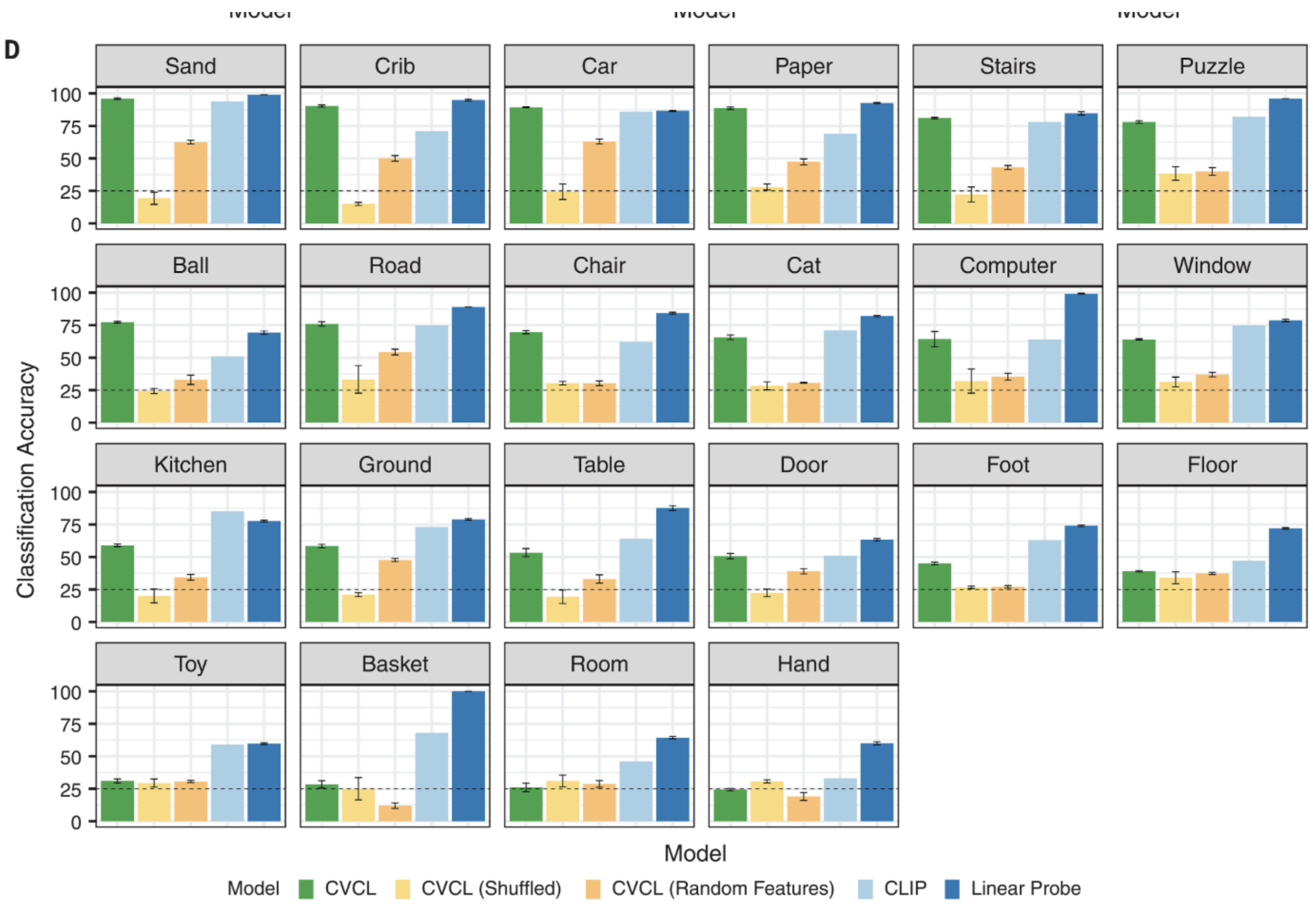
600,000 video frames paired with 37,500 transcribed utterances (extracted from 61 hours of video

CLIP

400 million image-text pairs from the web

Can word-referent mappings be learned with this amount of data?


A
Classification Accuracy
100
75
50
25
0
61.6
26.7
38.0
66.7
81.6
CVCL
CVCL (Shuffled)
CVCL (Rand. Features)
CLIP
Linear Probe


D
Sand
Crib
Car
Paper
Stairs
Puzzle
Ball
Road
Chair
Cat
Computer
Window
Kitchen
Ground
Table
Door
Foot
Floor
Toy
Basket
Room
Hand
Classification Accuracy
Model
Model
CVCL
CVCL (Shuffled)
CVCL (Random Features)
CLIP
Linear Probe

Limitations

The model is unrealistic:

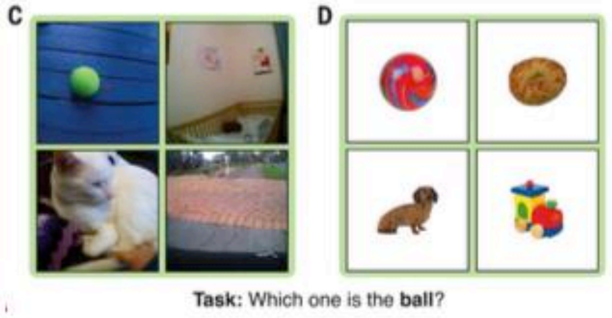
Using transcribed speech

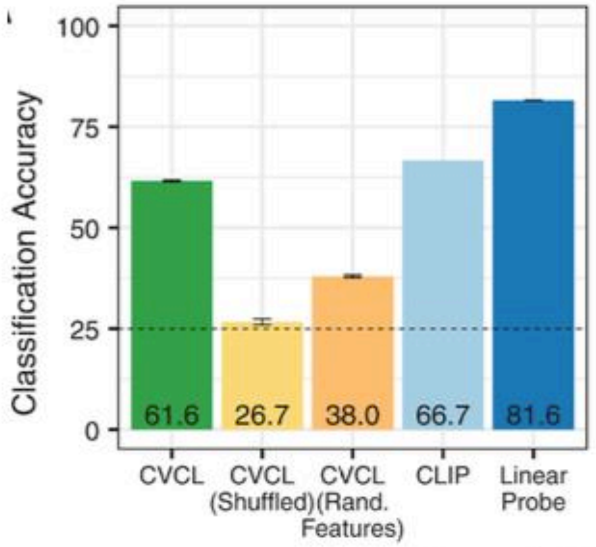
Training is not continuous

This is just testing noun learning, not grammar or non-nouns

They show that learning from domain-general mechanisms is possible, they don't show that it is sufficient

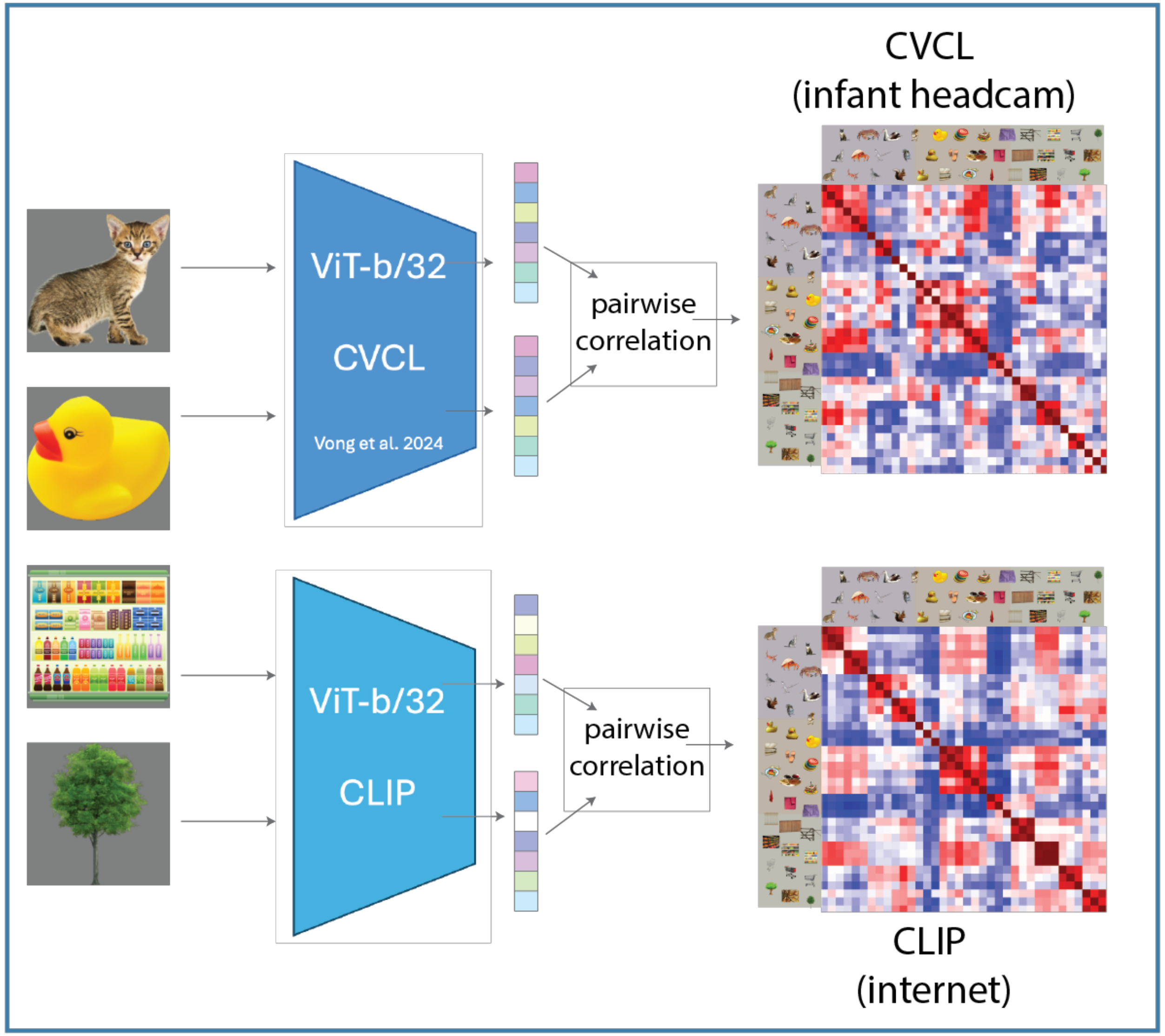
Future studies must match children input and outputs





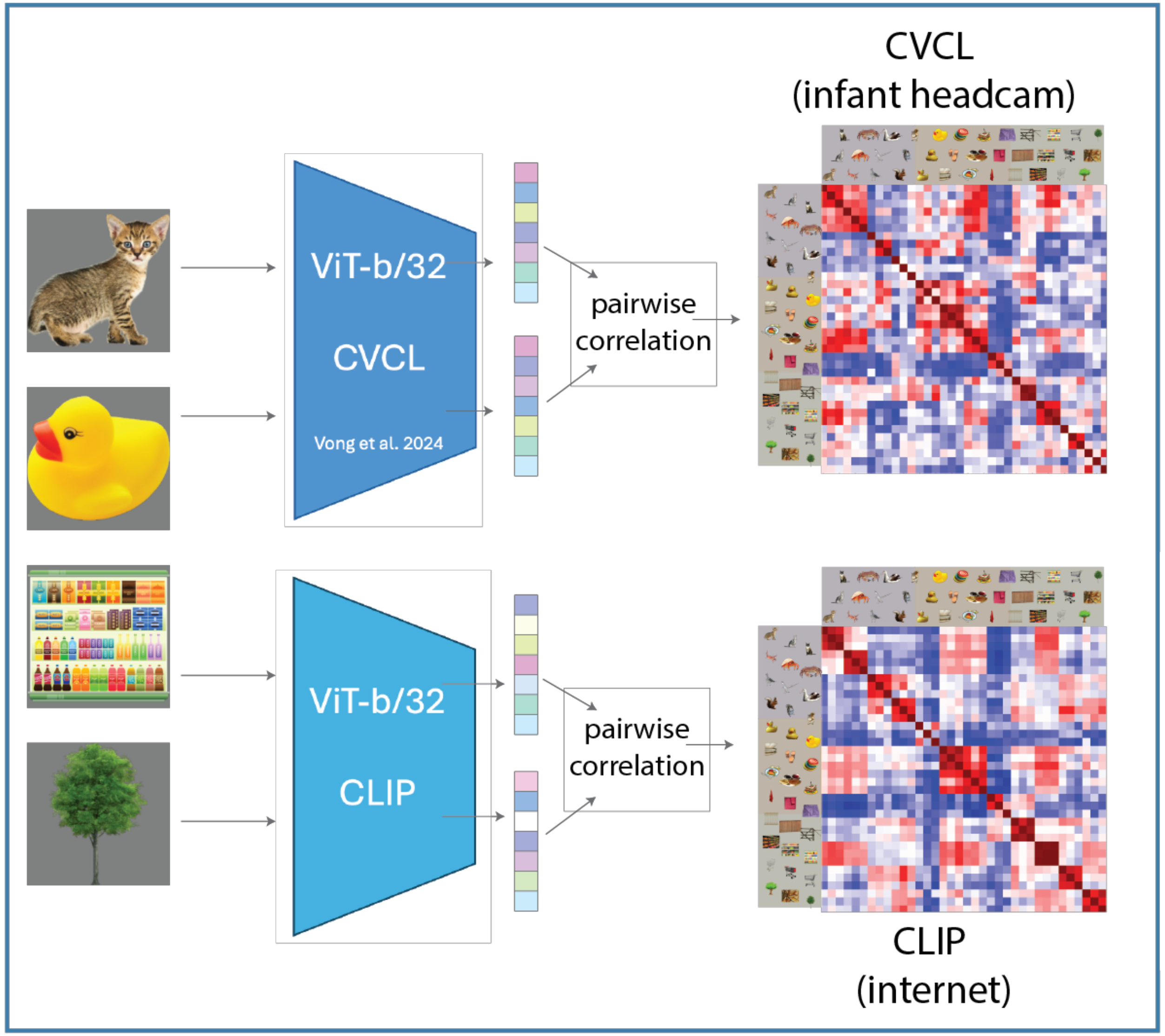
Vong et al., 2024

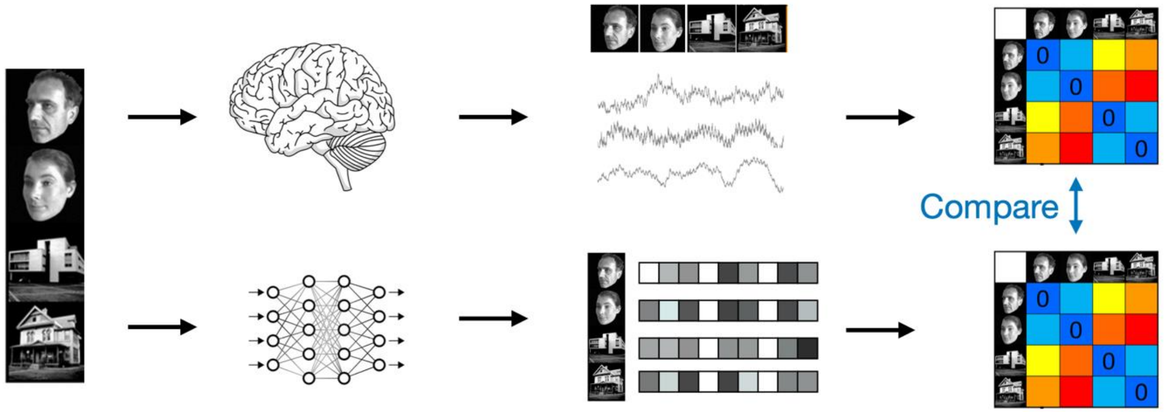
Developmental DNNs for modelling the developing brain



O'Doherty *et al.*, (in prep)

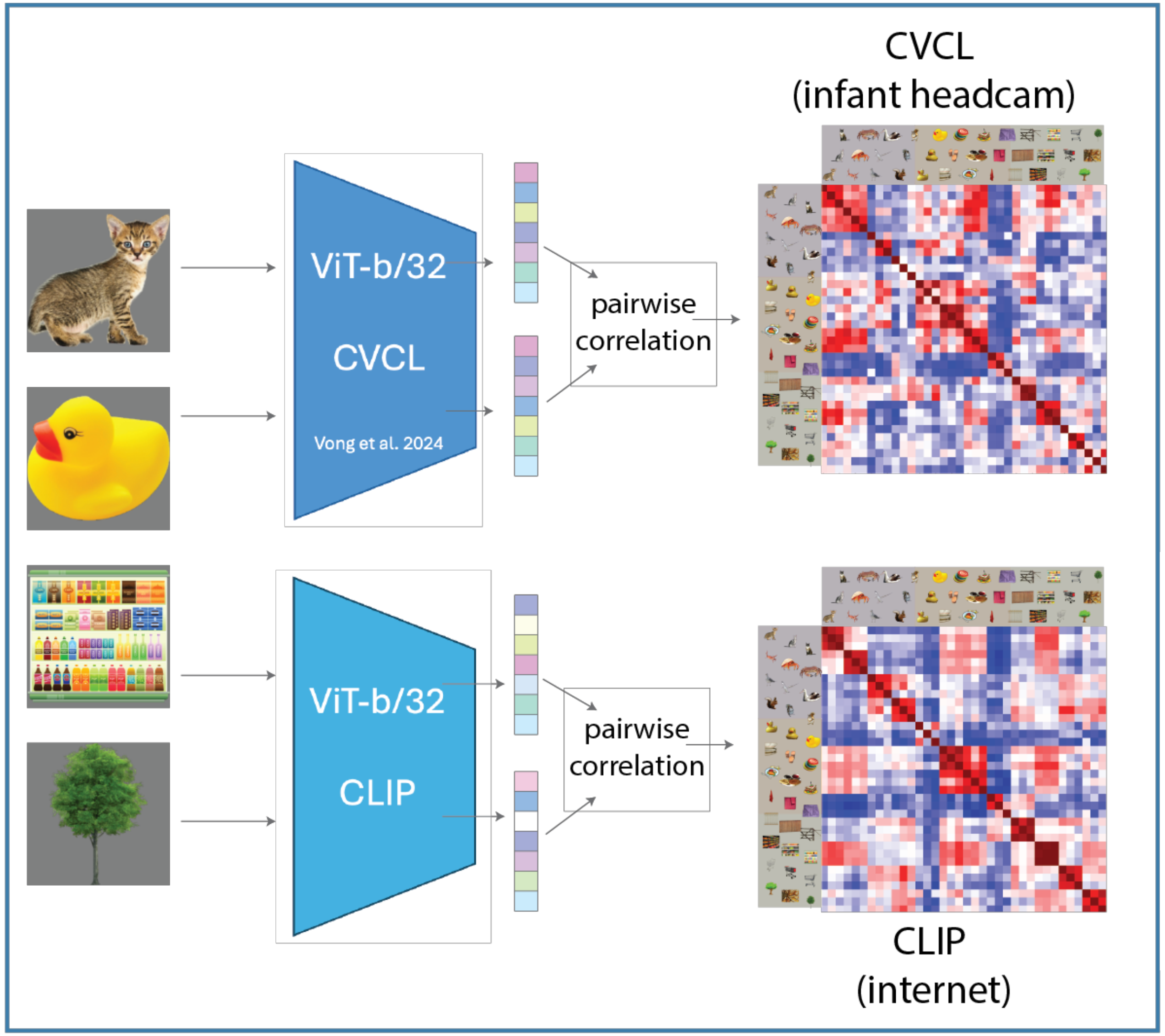
Developmental DNNs for modelling the developing brain

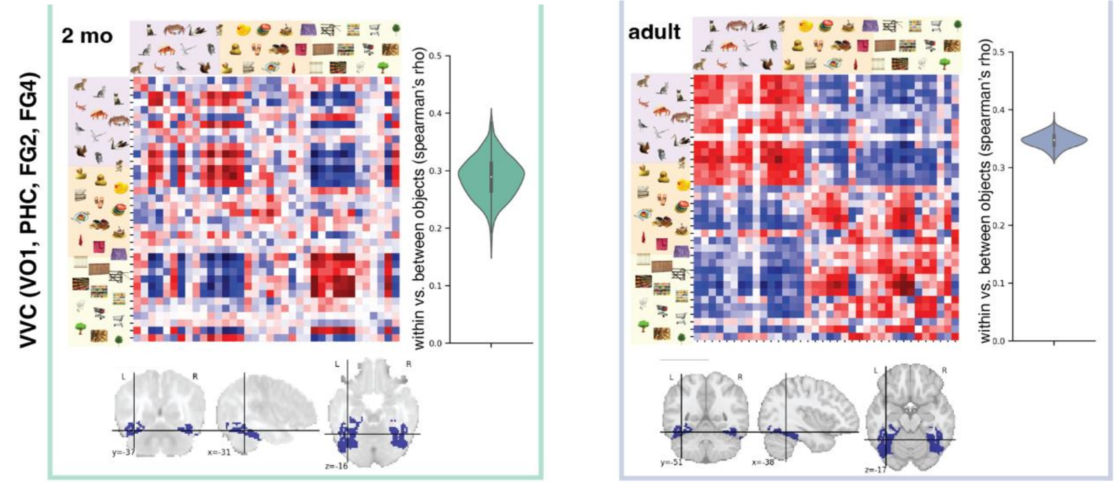


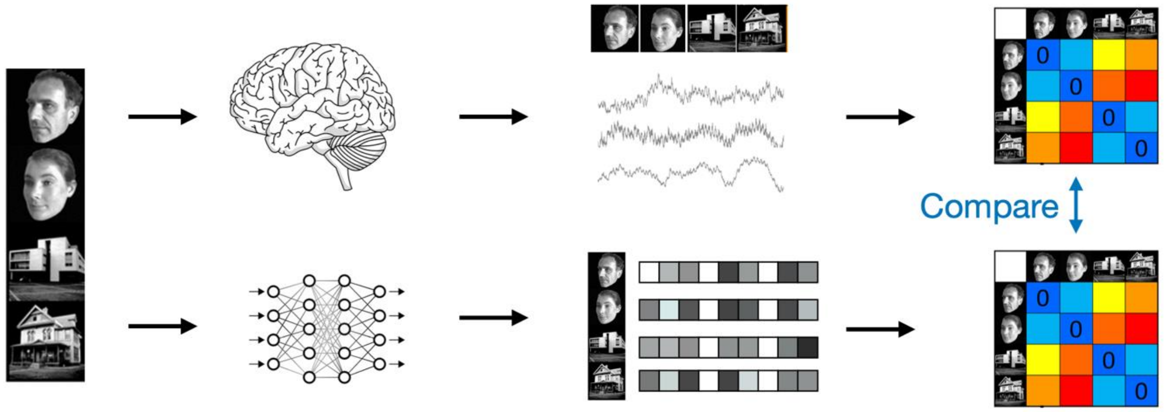


O'Doherty *et al.*, (in prep)

Developmental DNNs for modelling the developing brain

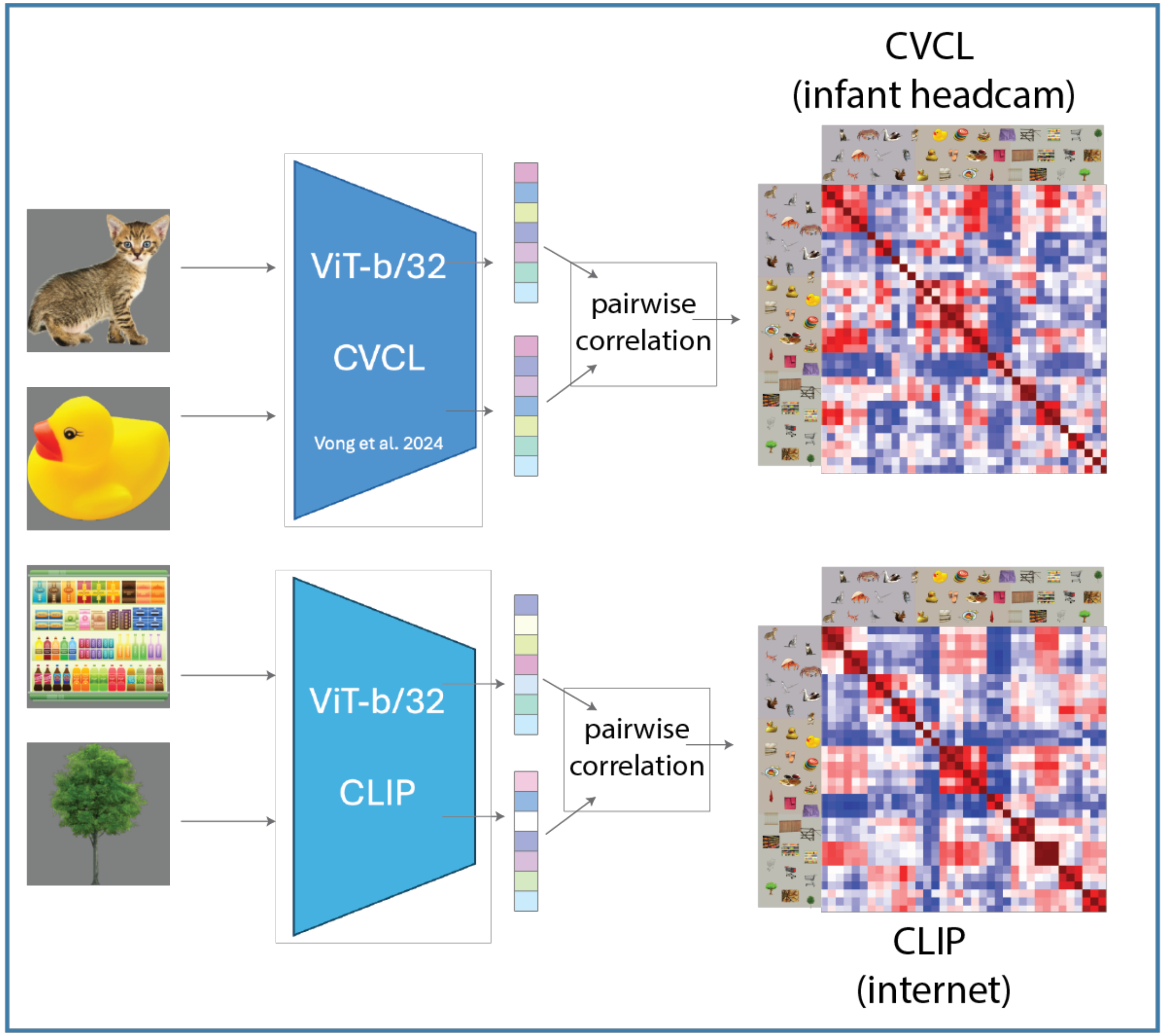


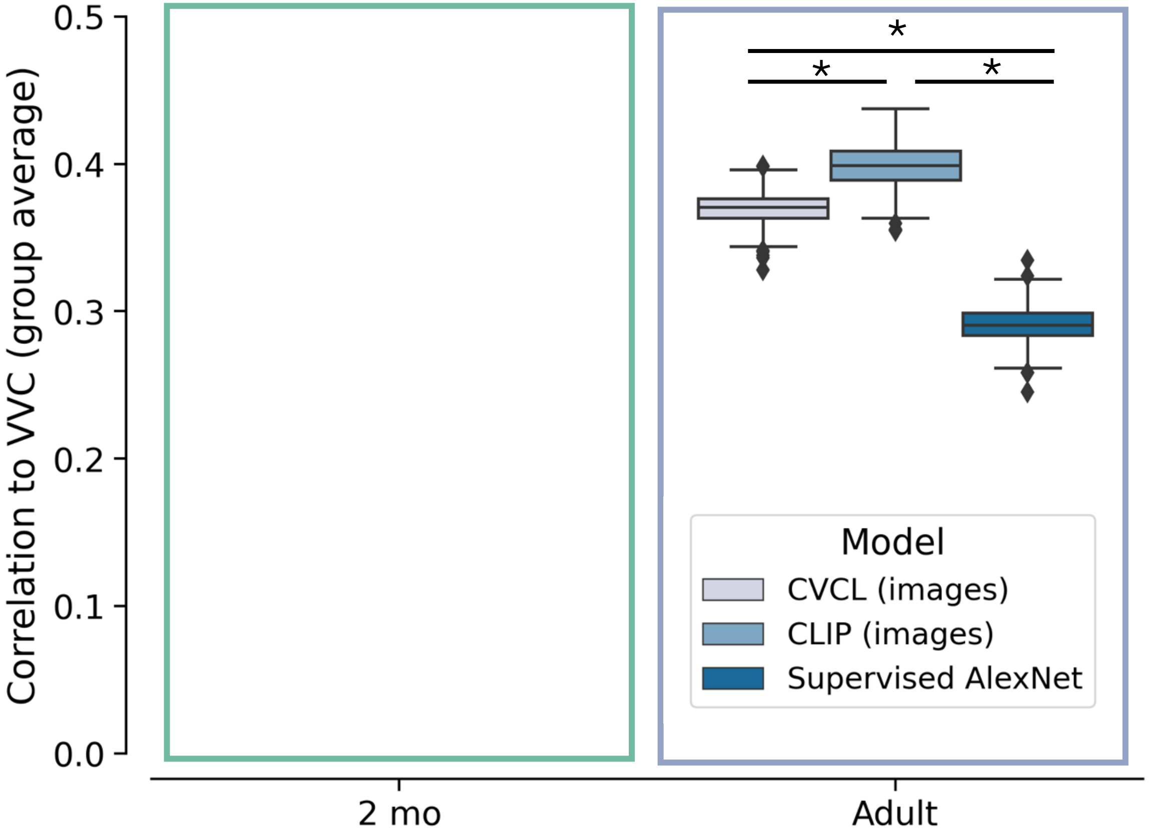




O'Doherty *et al.*, (in prep)

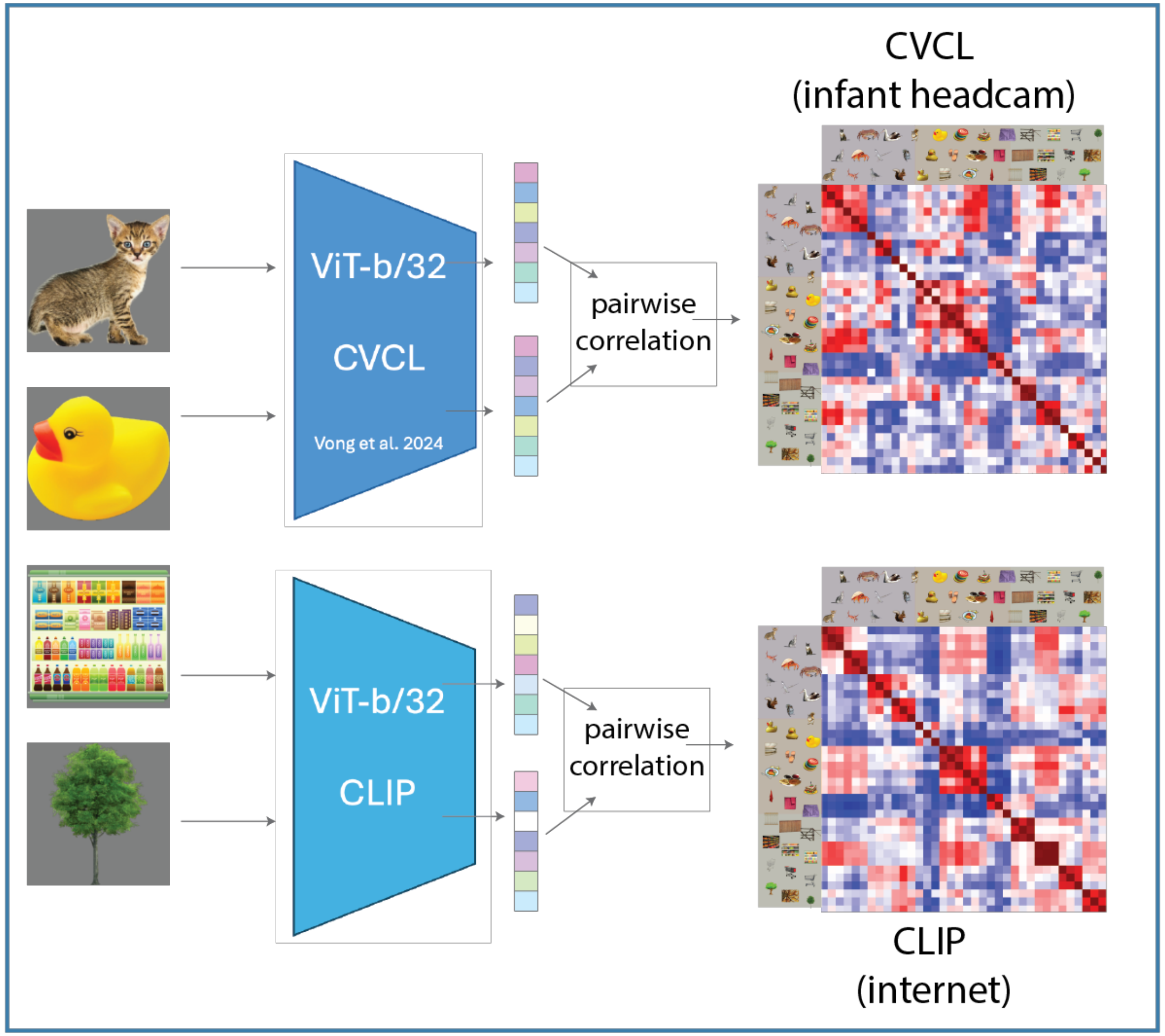
Developmental DNNs for modelling the developing brain

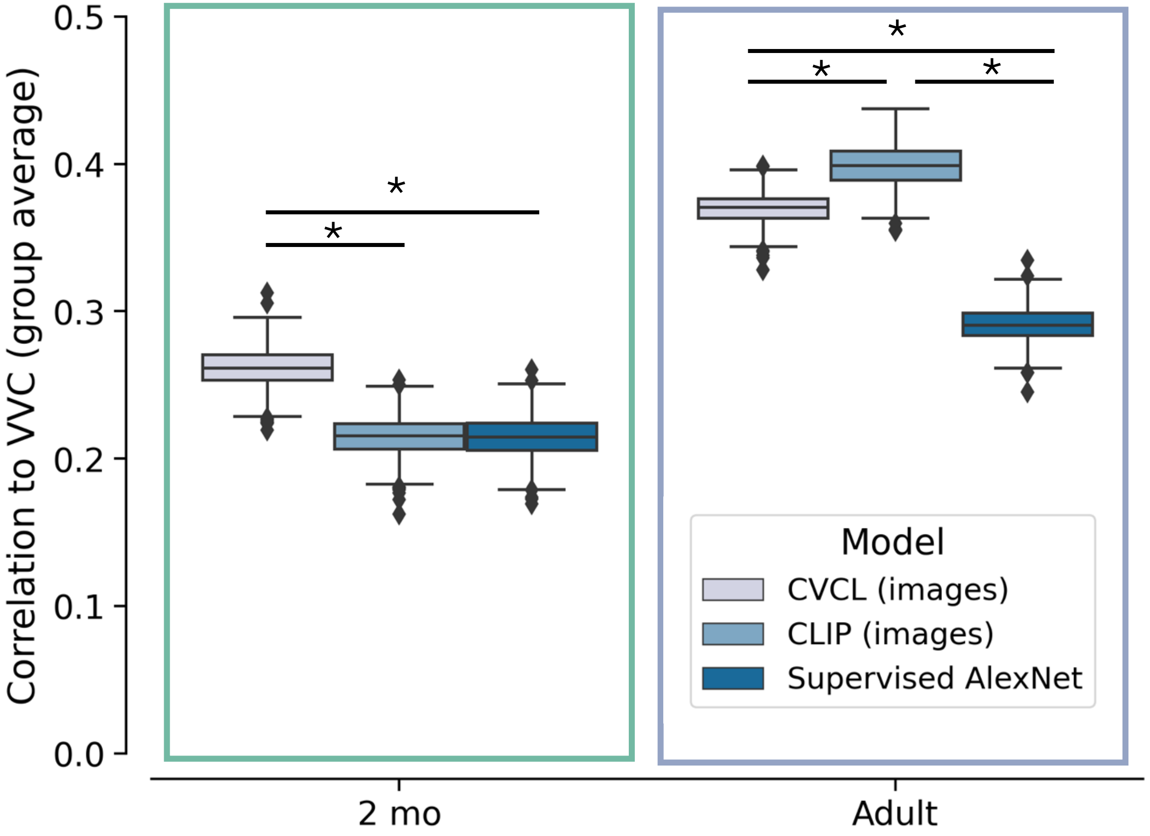




O'Doherty *et al.*, (in prep)

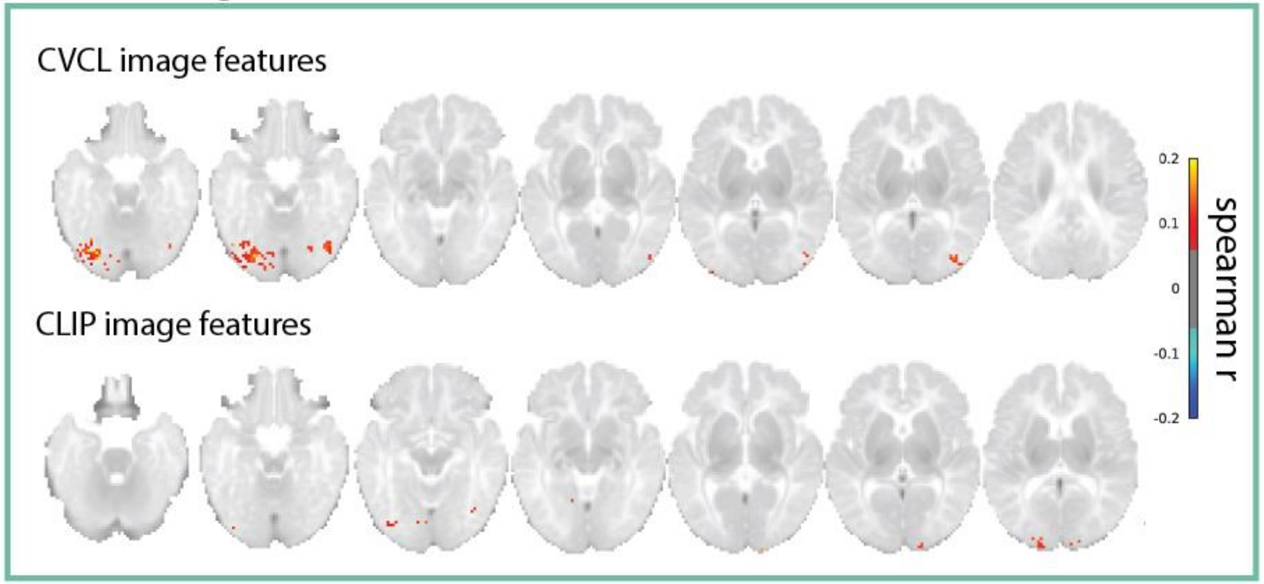
Developmental DNNs for modelling the developing brain



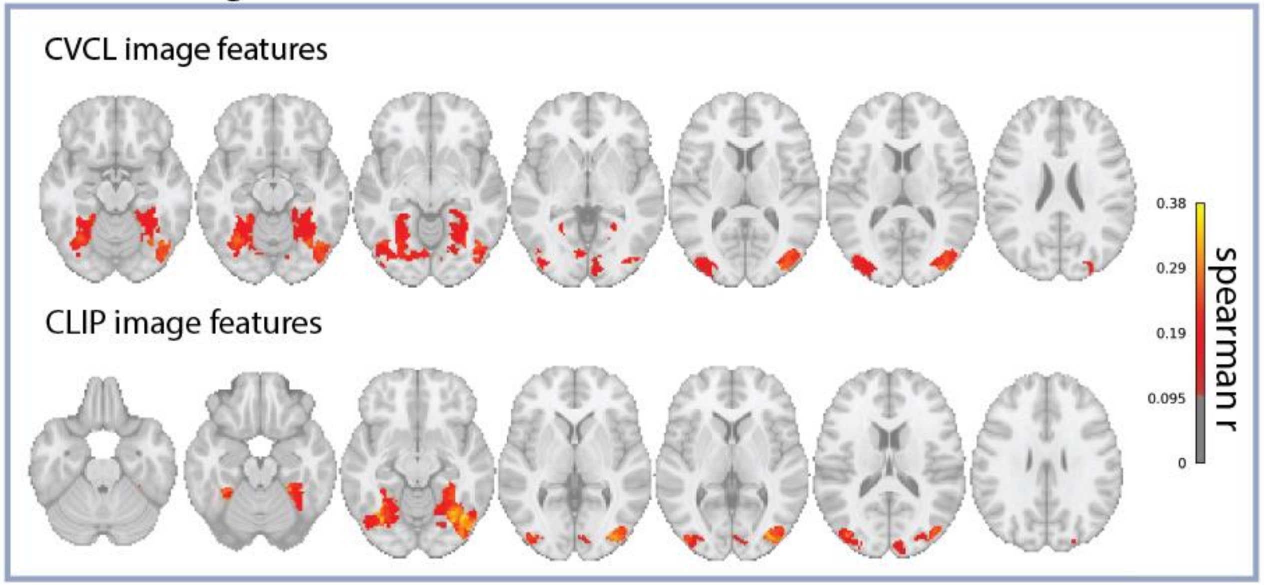


O'Doherty *et al.*, (in prep)

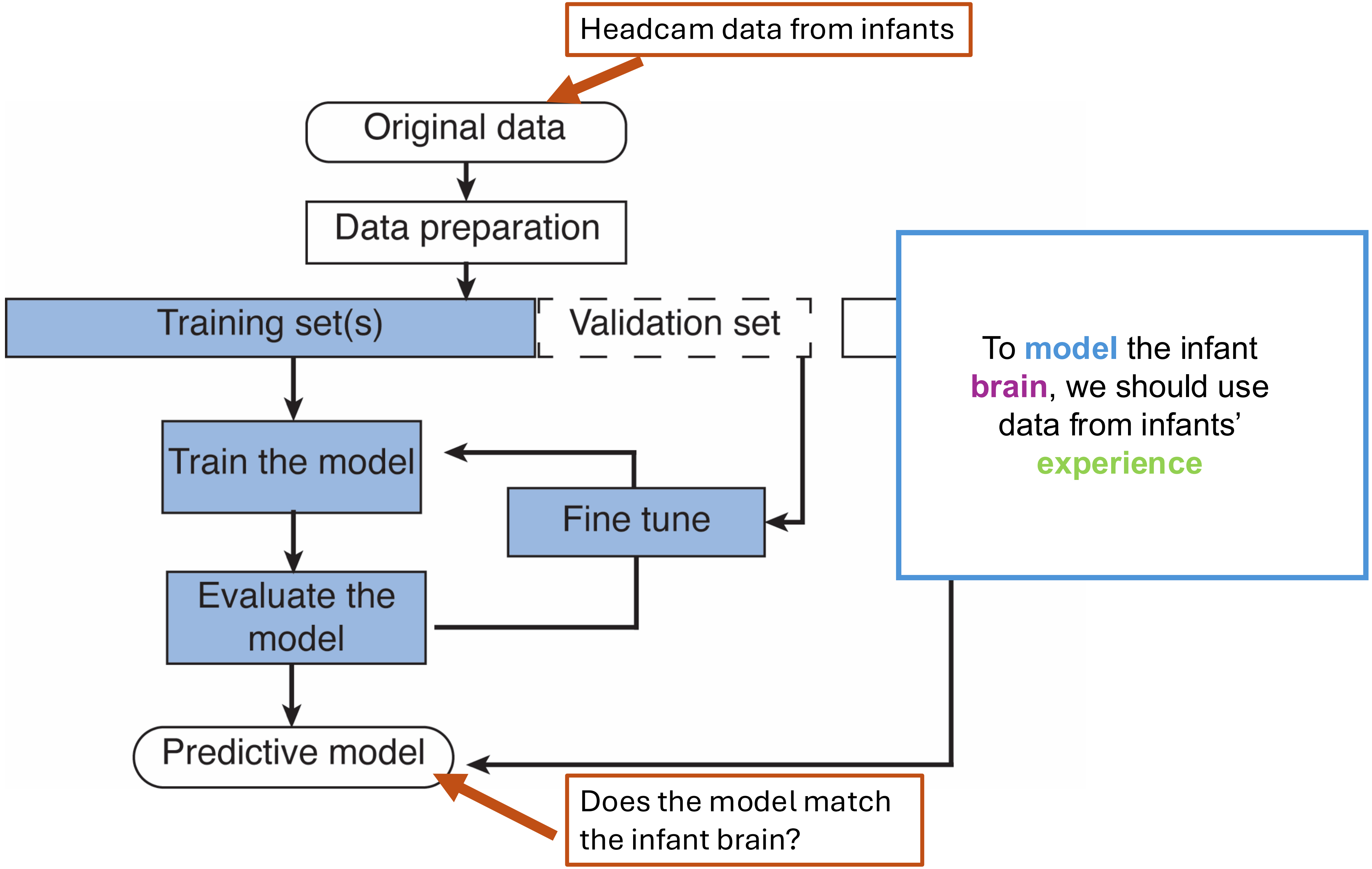
2 mo Searchlight

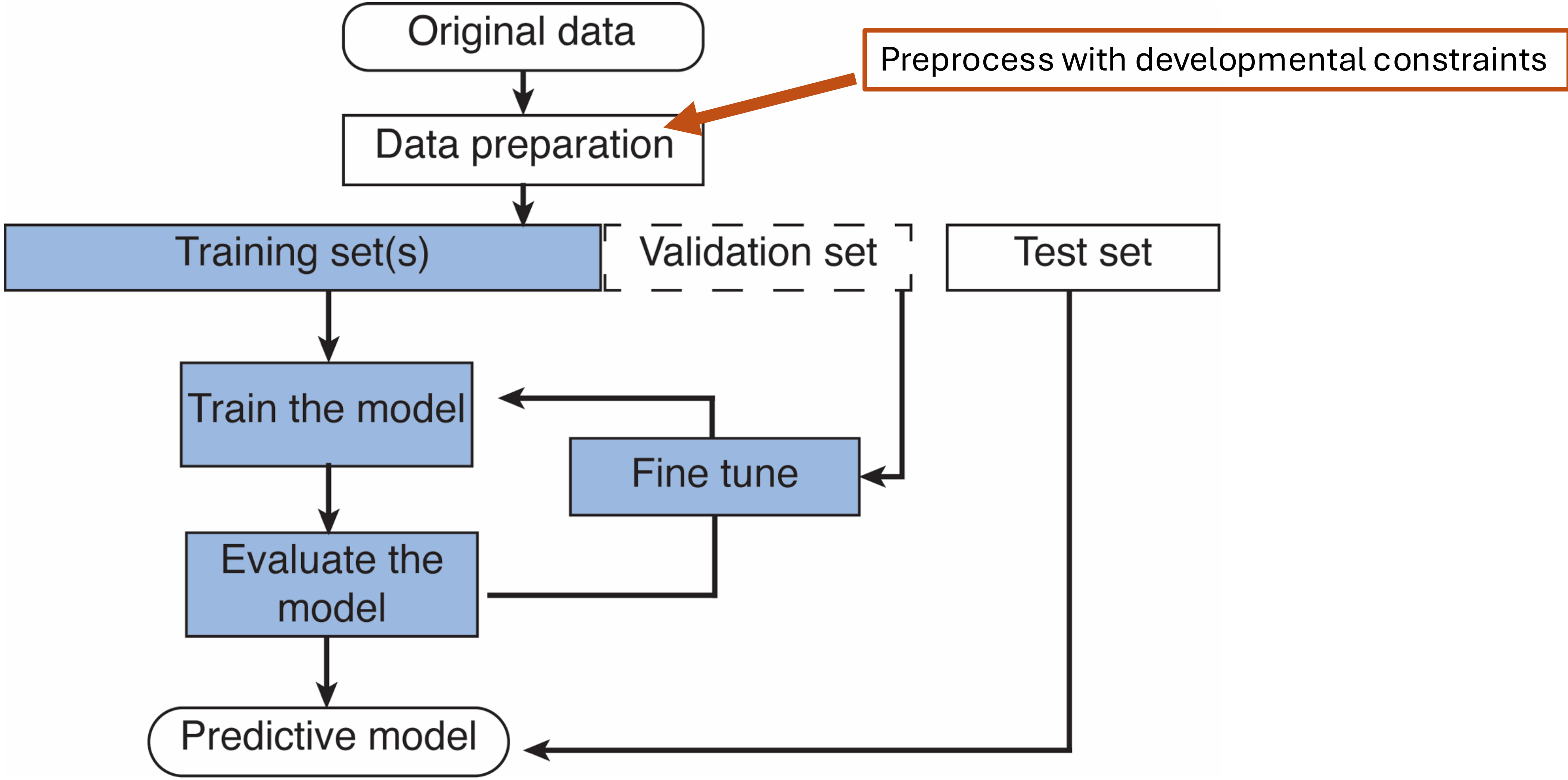


Adult Searchlight



O'Doherty *et al.*, (in prep)

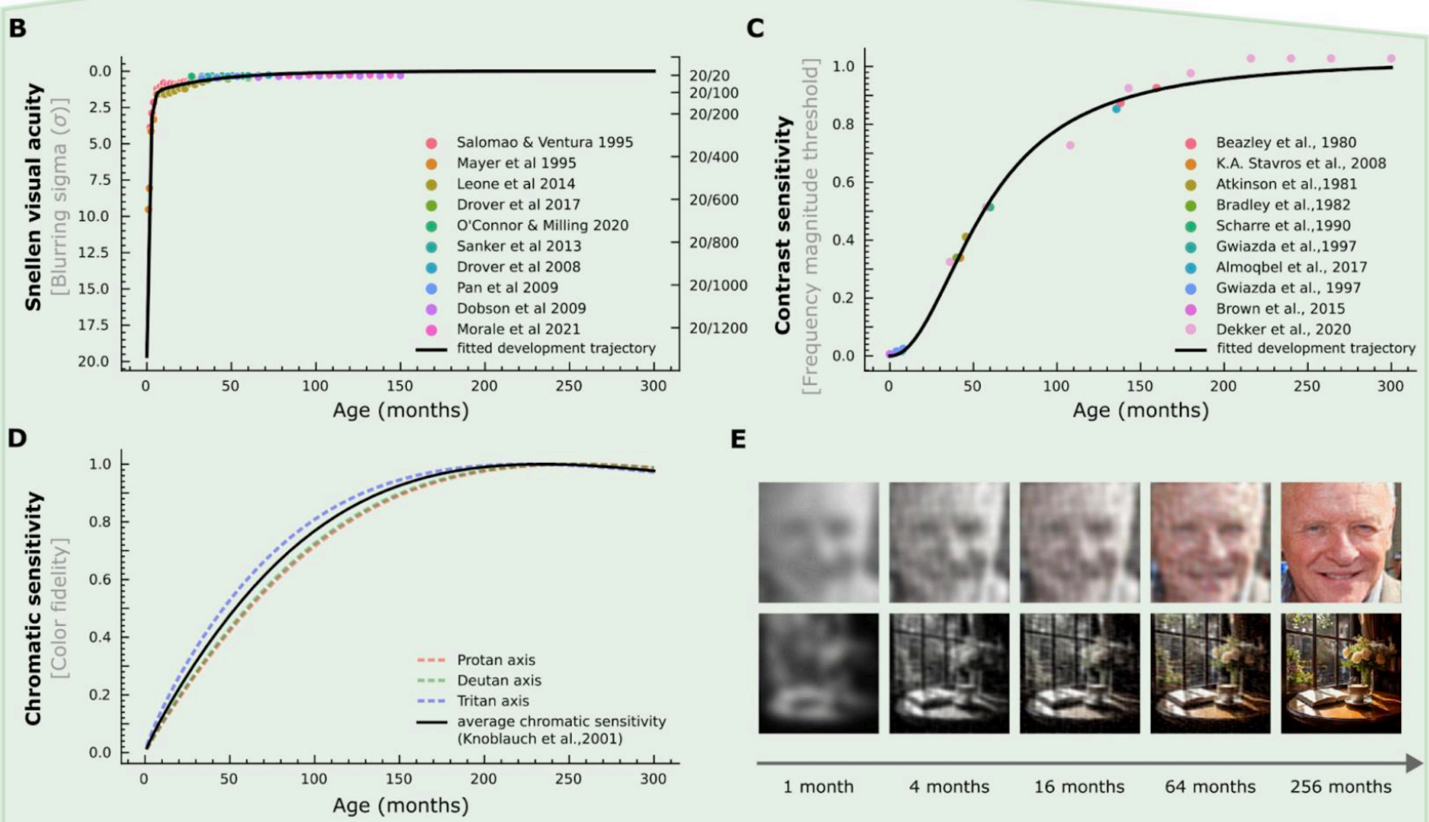

Headcam data from infants
Original data
Data preparation
Training set(s)
Validation set
Train the model
Fine tune
Evaluate the model
Predictive model
To model the infant brain, we should use data from infants' experience
Does the model match the infant brain?

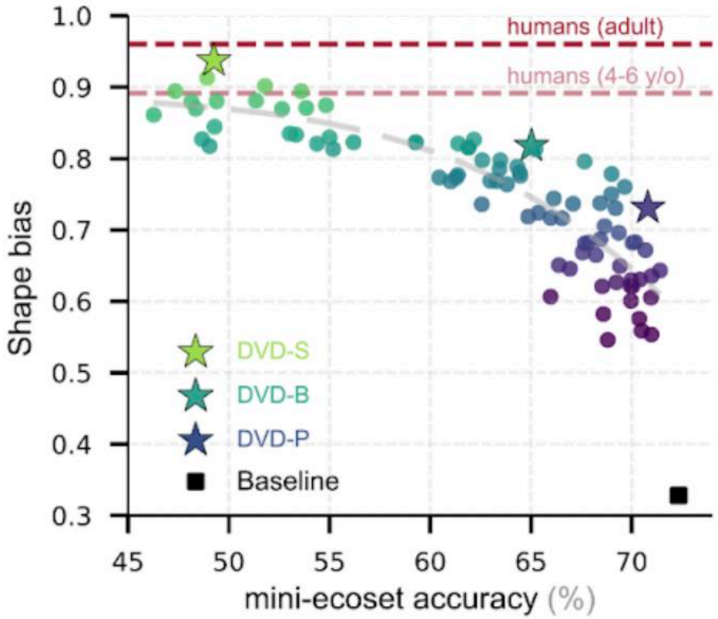

Original data
Data preparation
Preprocess with developmental constraints
Training set(s)
Validation set
Test set
Train the model
Fine tune
Evaluate the model
Predictive model

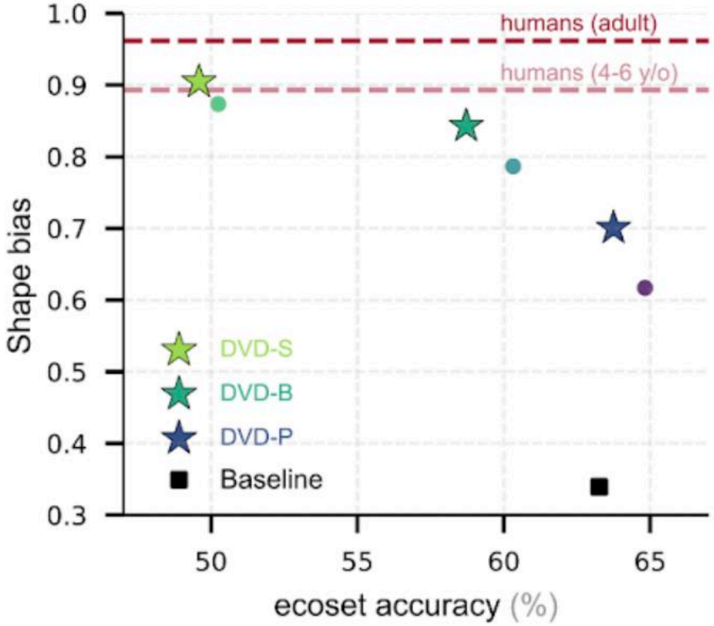
Neural networks are biased towards texture, not shape like humans

What interventions will correct for this shape bias?

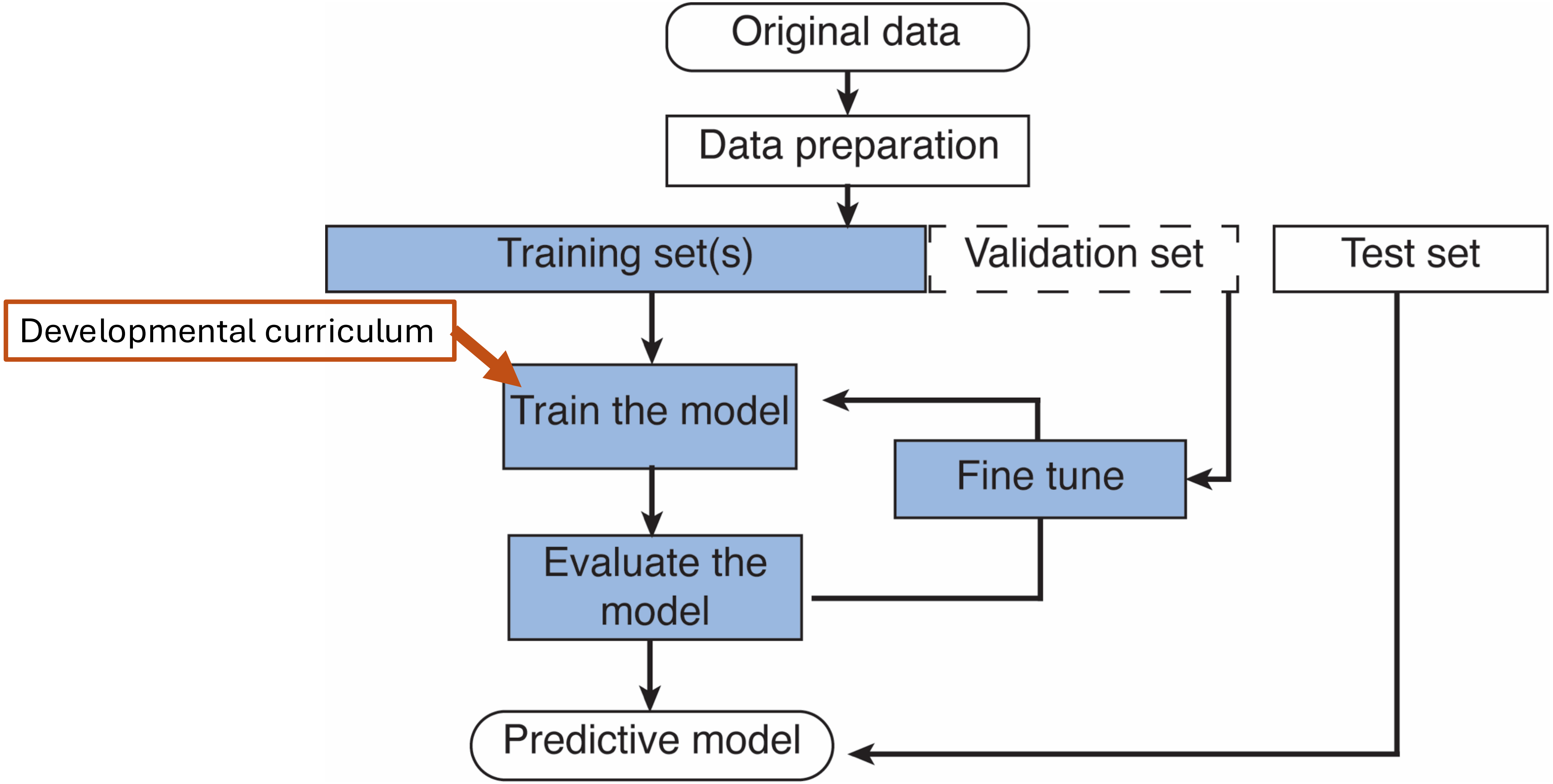
Lu *et al., arXiv preprint* 2025







Lu *et al.*, *arXiv preprint* 2025


Original data
Data preparation
Training set(s)
Validation set
Test set
Developmental curriculum
Train the model
Fine tune
Evaluate the model
Predictive model

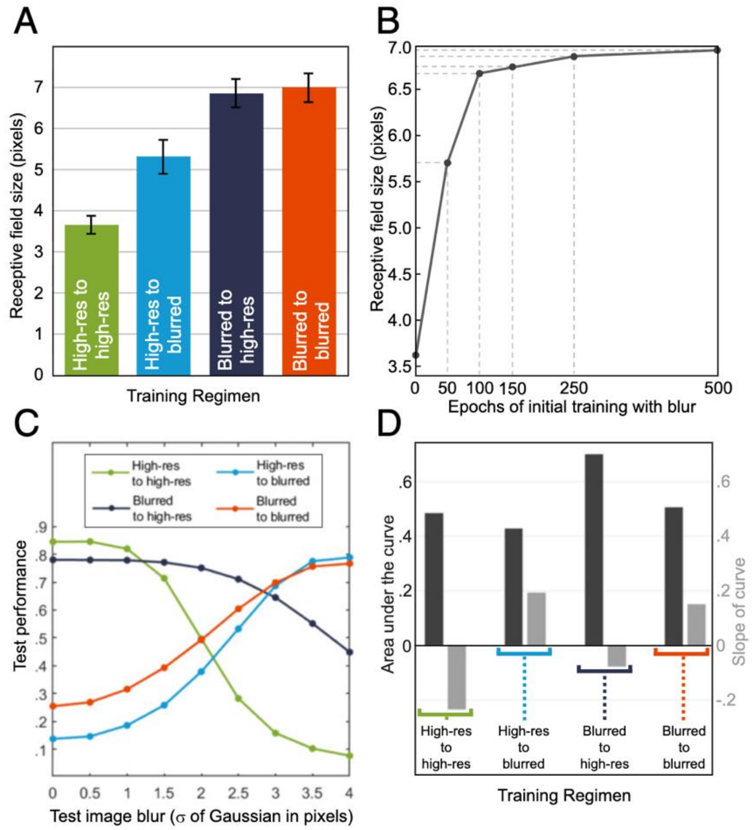
Children who are born with cataracts and then sight is later restored

- Visual acuity recovers
- Configural processing of faces is impacted
- Although some resilience in ventrotemporal cortex Mattioni *et al.* 2025

Potential downside of high initial visual acuity

Lukas Vogelsang[a,b,1], Sharon Gilad-Gutnick[a,1], Evan Ehrenberg[a], Albert Yonas[c], Sidney Diamond[a], Richard Held[a,2], and Pawan Sinha[a,3]

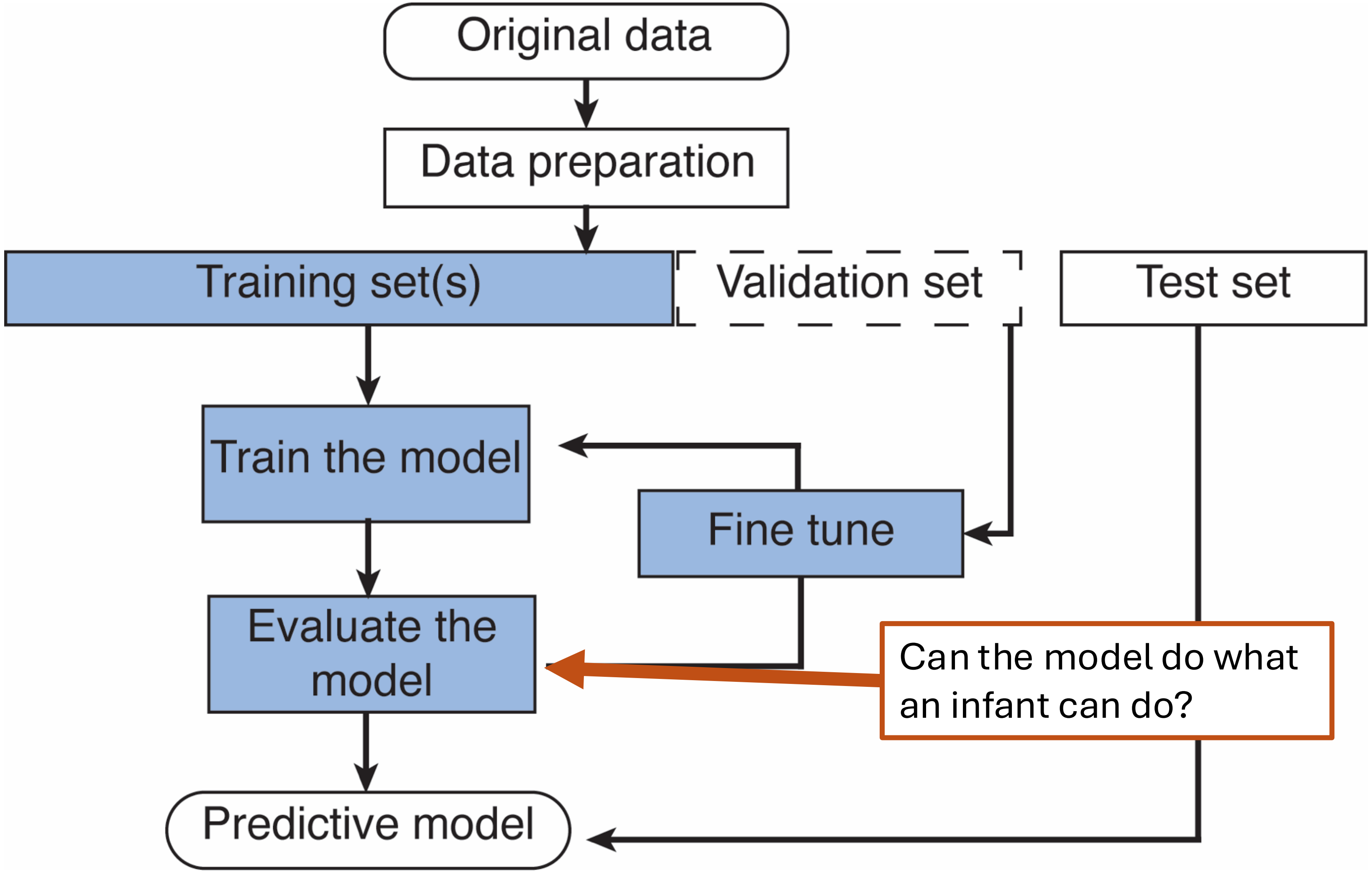
[a]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Institute of Neuroinformatics, University of Zurich/ETH Zurich, 8057 Zurich, Switzerland; and [c]Department of Psychology, Arizona State University, Tempe, AZ 85281



Is blurry vision adaptive?

Initial blurry exposure allows for larger receptive fields – ability to integrate information over a larger visual area

Even limited exposure to blur stabilises receptive fields and allows for better face classification


Original data
Data preparation
Training set(s)
Validation set
Test set
Train the model
Fine tune
Evaluate the model
Can the model do what an infant can do?
Predictive model

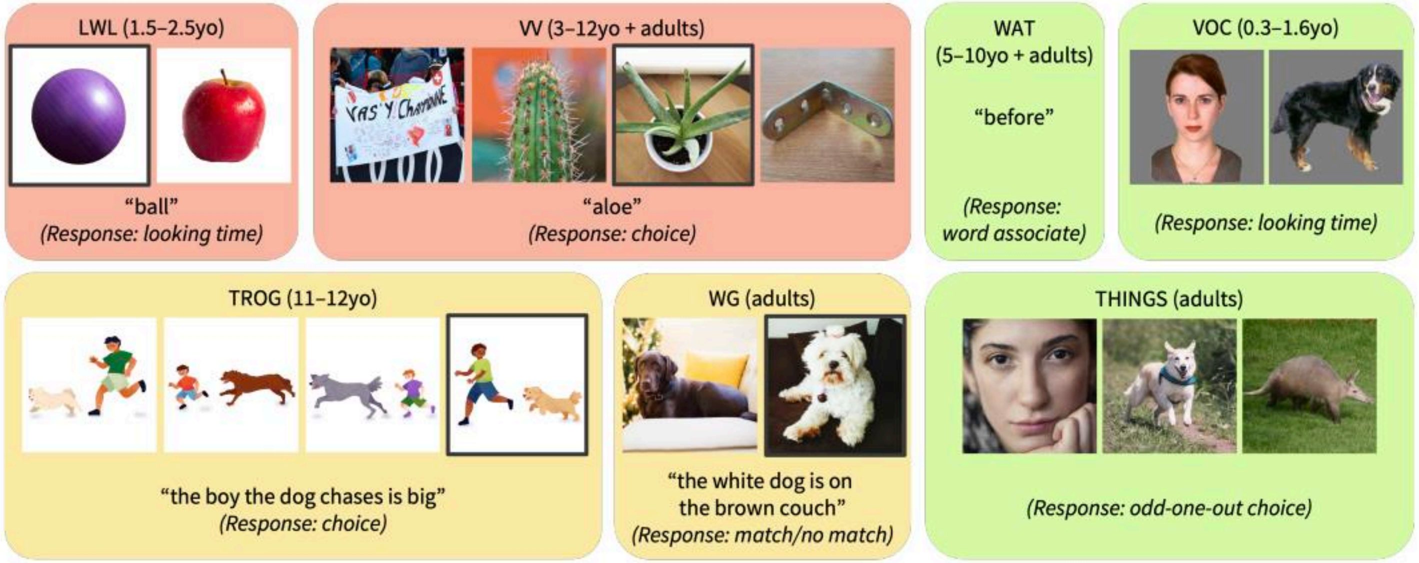
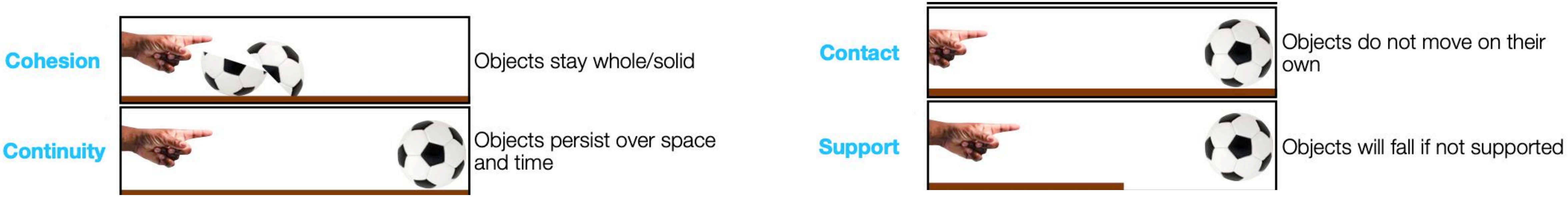


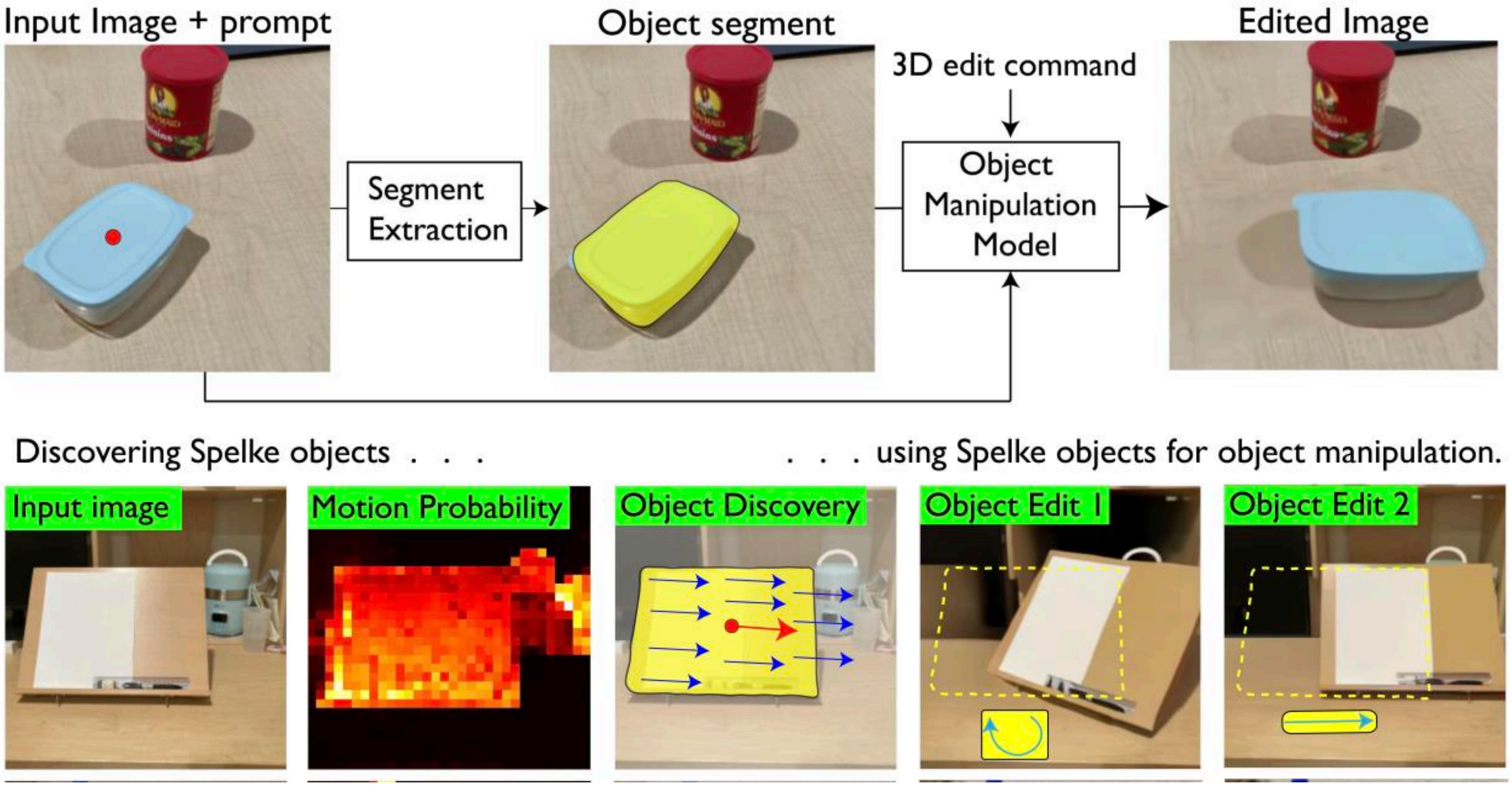
Table 2: Model characteristics and performance across all tasks, demonstrating variation across models. Arrows indicate the direction of better performance (i.e., lower is better vs. higher is better). Bolded results indicate most human-like result on a task.

			Lexicon		Syntax		Semantics		
Model	# params	# images	LWL (↓)	VV (↓)	TROG (↓)	WG (↓)	WAT (↓)	VOC (↑)	THINGS (↑)
CLIP-base [48]	149M	400M	0.014	0.205	0.732	0.256	0.495	-0.081	**0.397**
CLIP-large [48]	428M	400M	0.013	**0.179**	0.692	0.256	0.495	0.005	0.246
ViLT [49]	87M	4.1M	0.009	0.326	0.682	0.252	0.495	-0.053	0.127
FLAVA [50]	350M	70M	0.013	0.197	0.912	0.254	0.495	-0.042	0.189
BLIP [51]	252M	14M	0.010	0.193	**0.576**	0.259	0.495	-0.104	0.185
BridgeTower [52]	333M	4M	**0.008**	0.265	0.584	**0.250**	0.495	-0.095	0.345
OpenCLIP-H [53]	1.0B	32B	0.012	0.188	0.683	0.255	**0.495**	0.031	0.227
SigLIP [54]	800M	9B	0.067	0.612	0.888	0.258	0.495	-0.028	0.192
CVCL [4]	26M	600K	0.060	0.740	0.911	0.258	0.495	**0.138**	0.175
Human			0.010	0.091	0.028			0.251	
Random (OpenCLIP)	1.0B	0	0.087	0.740	0.908	0.258	0.495	0.246	0.054

DevBench

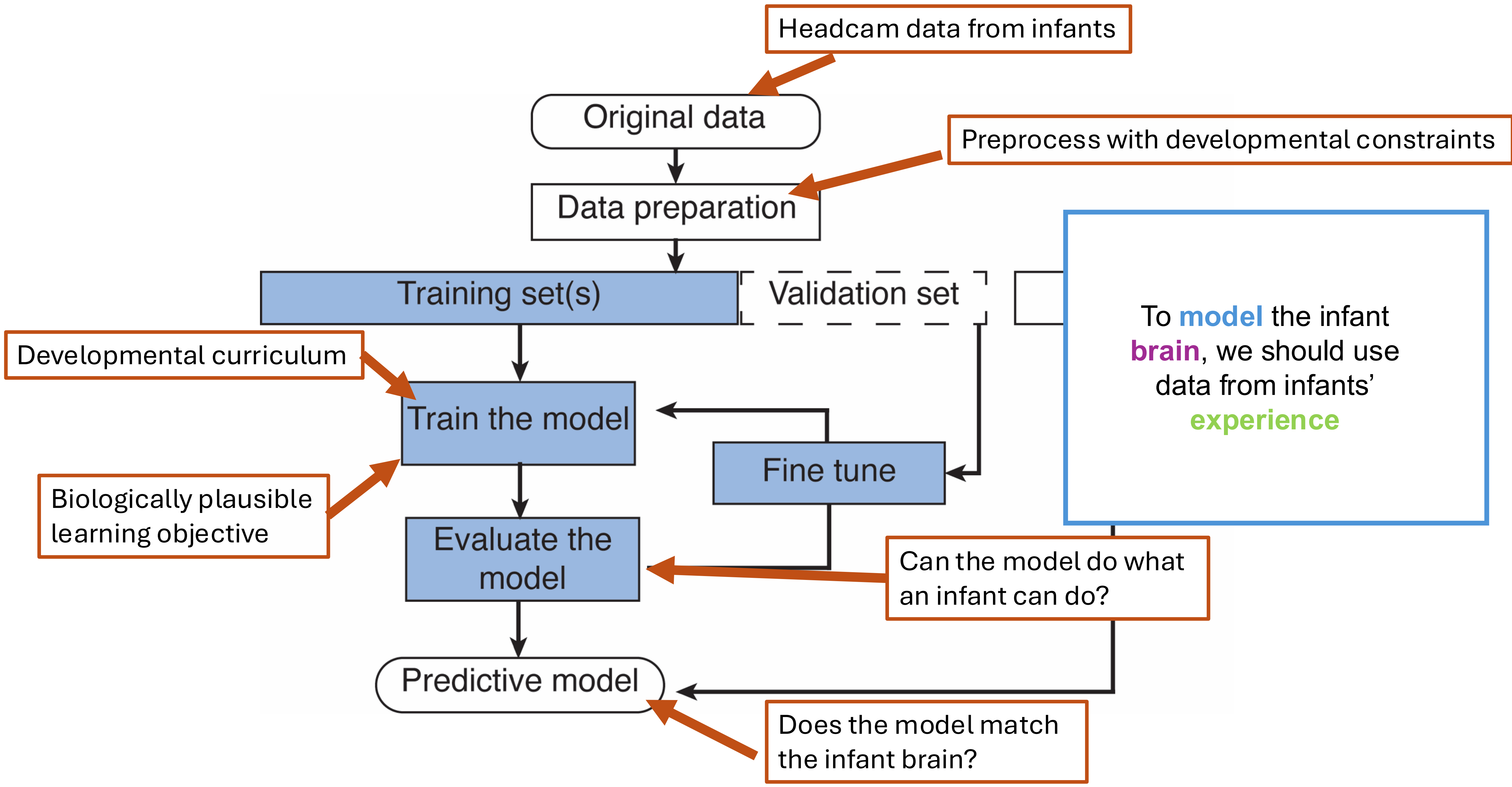
Tan *et al. NeurIPS 2024*

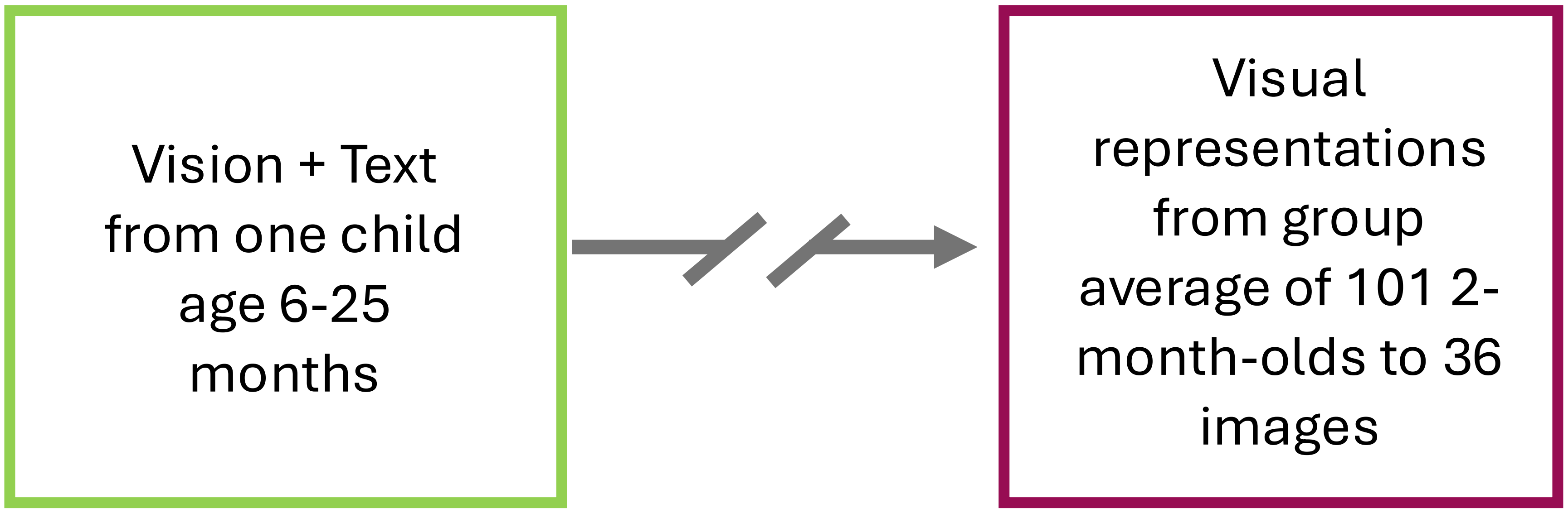


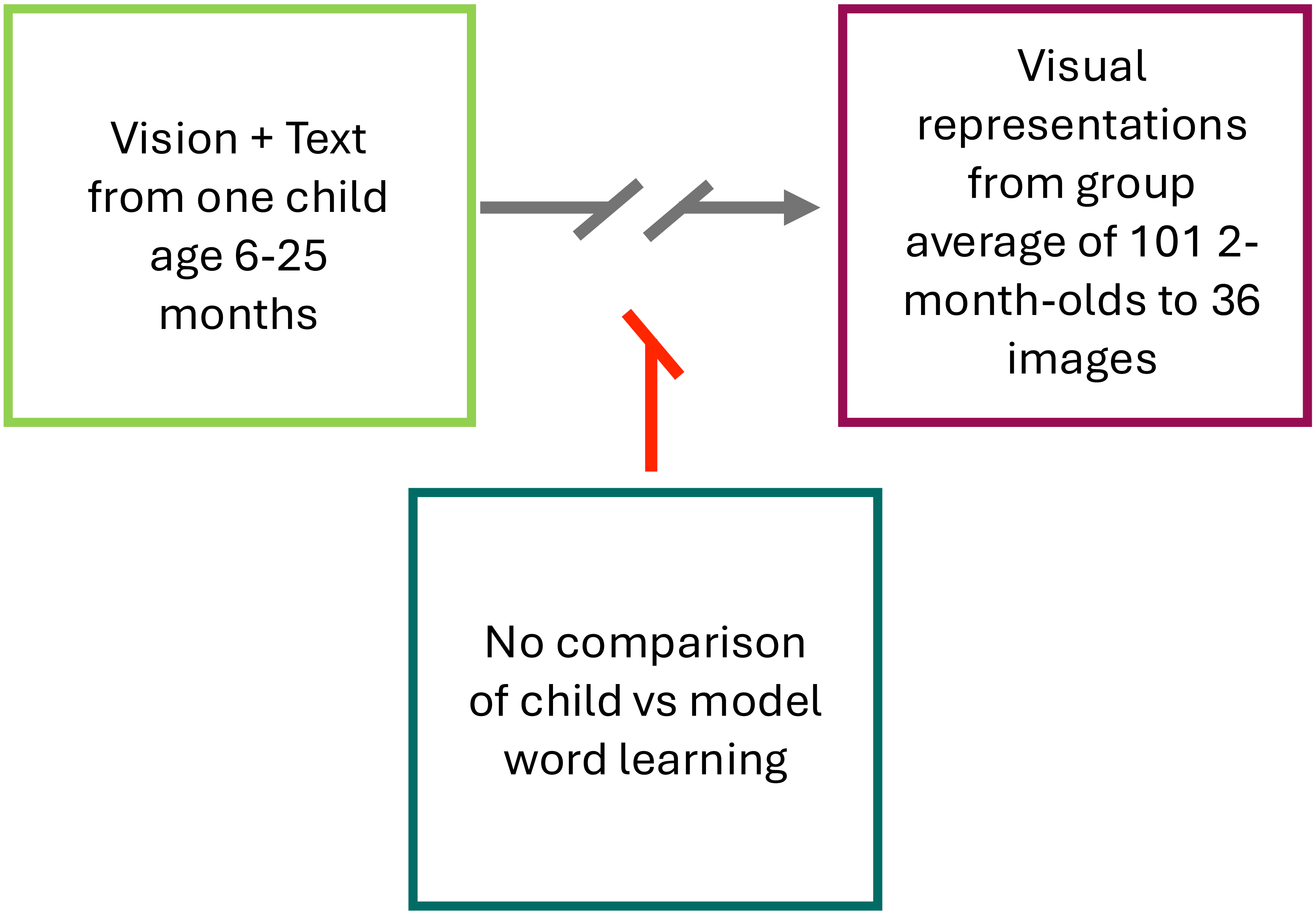


SpelkeBench

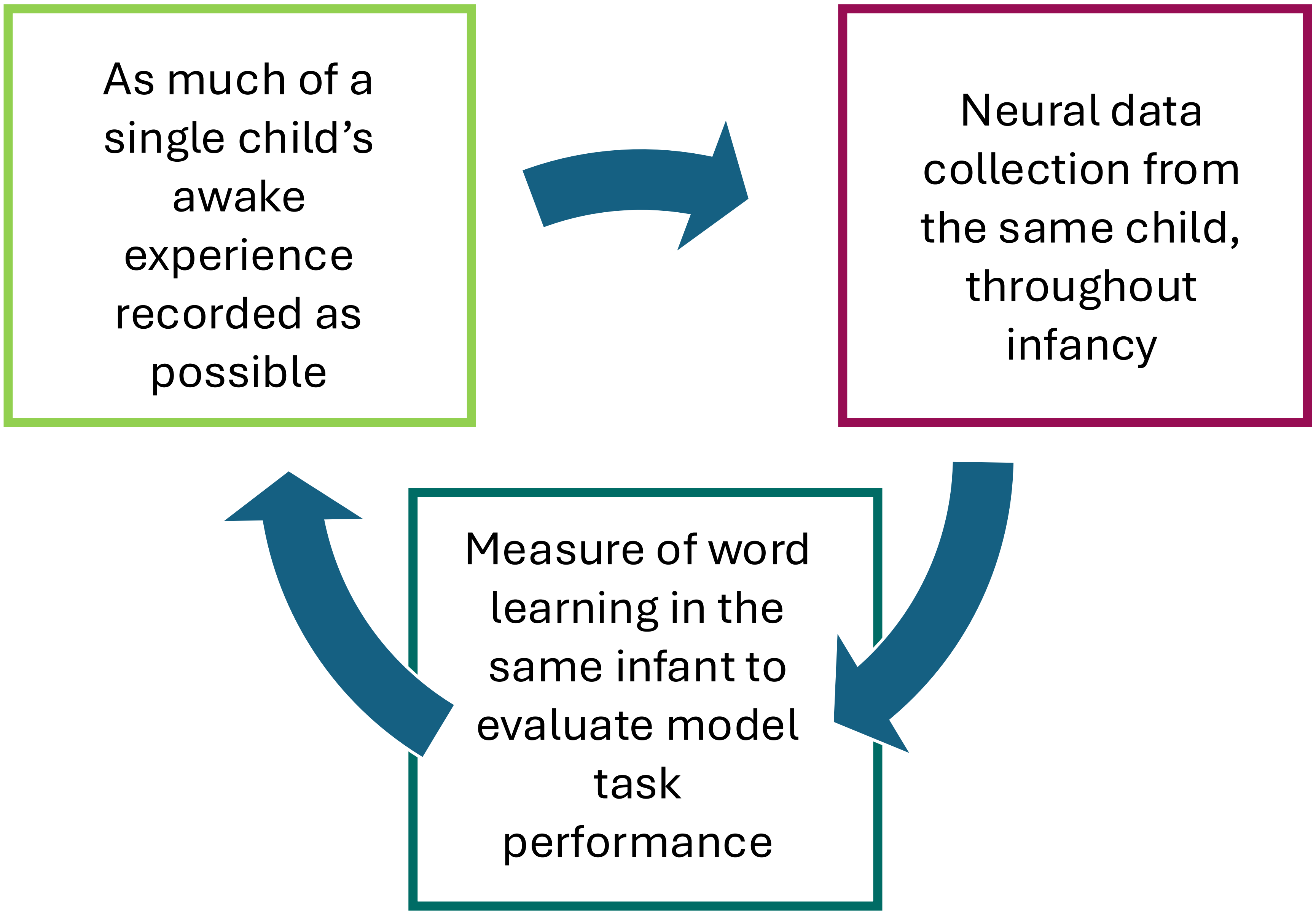
Vankatesh *et al. arXiv 2025*


Headcam data from infants
Original data
Preprocess with developmental constraints
Data preparation
Training set(s)
Validation set
To model the infant brain, we should use data from infants' experience
Developmental curriculum
Train the model
Fine tune
Biologically plausible learning objective
Evaluate the model
Can the model do what an infant can do?
Predictive model
Does the model match the infant brain?


Vision + Text from one child age 6-25 months
Visual representations from group average of 101 2-month-olds to 36 images

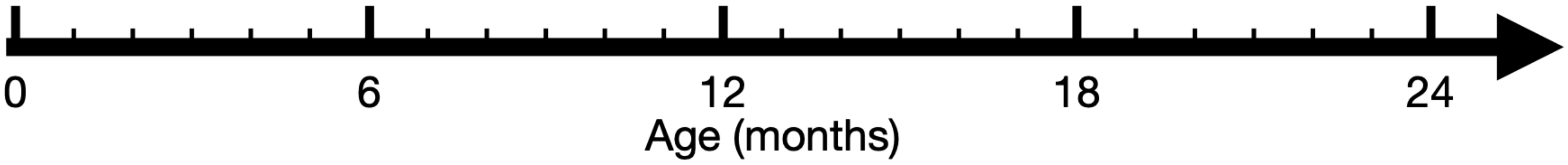


Vision + Text from one child age 6-25 months
Visual representations from group average of 101 2-month-olds to 36 images
No comparison of child vs model word learning


As much of a single child's awake experience recorded as possible
Neural data collection from the same child, throughout infancy
Measure of word learning in the same infant to evaluate model task performance

Project Simsom

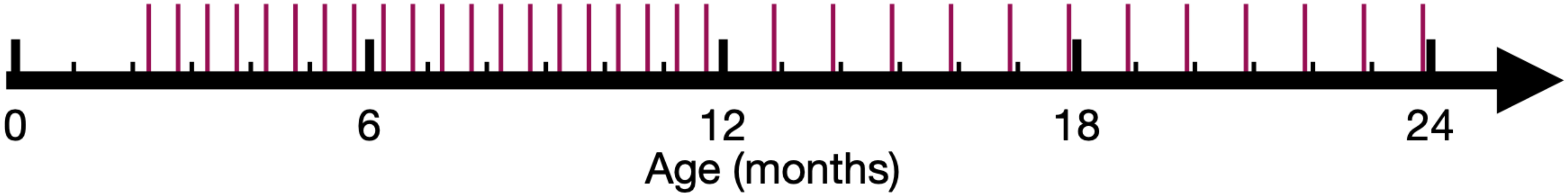
Project Simsom



Project Simsom

Brain

Dense sample of the
brain's
response to audiovisual
static
and dynamic stimuli



Project Simsom

Experience

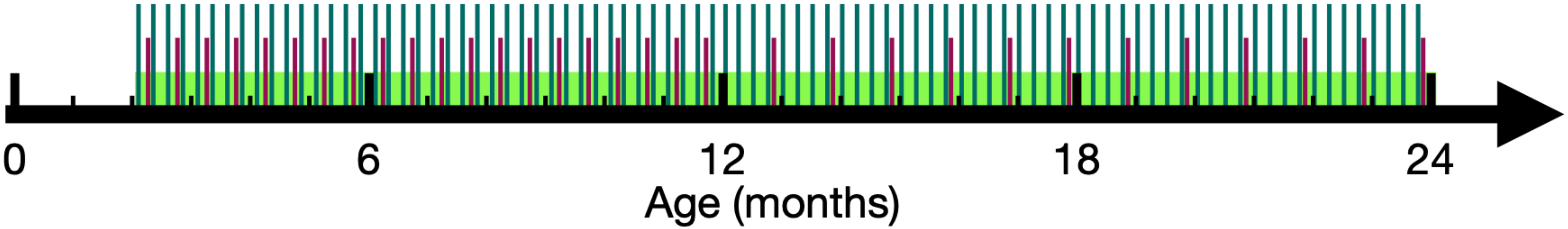
Dense sample of visual and
auditory experiences, plus minor interventions to change
experiences experimentally

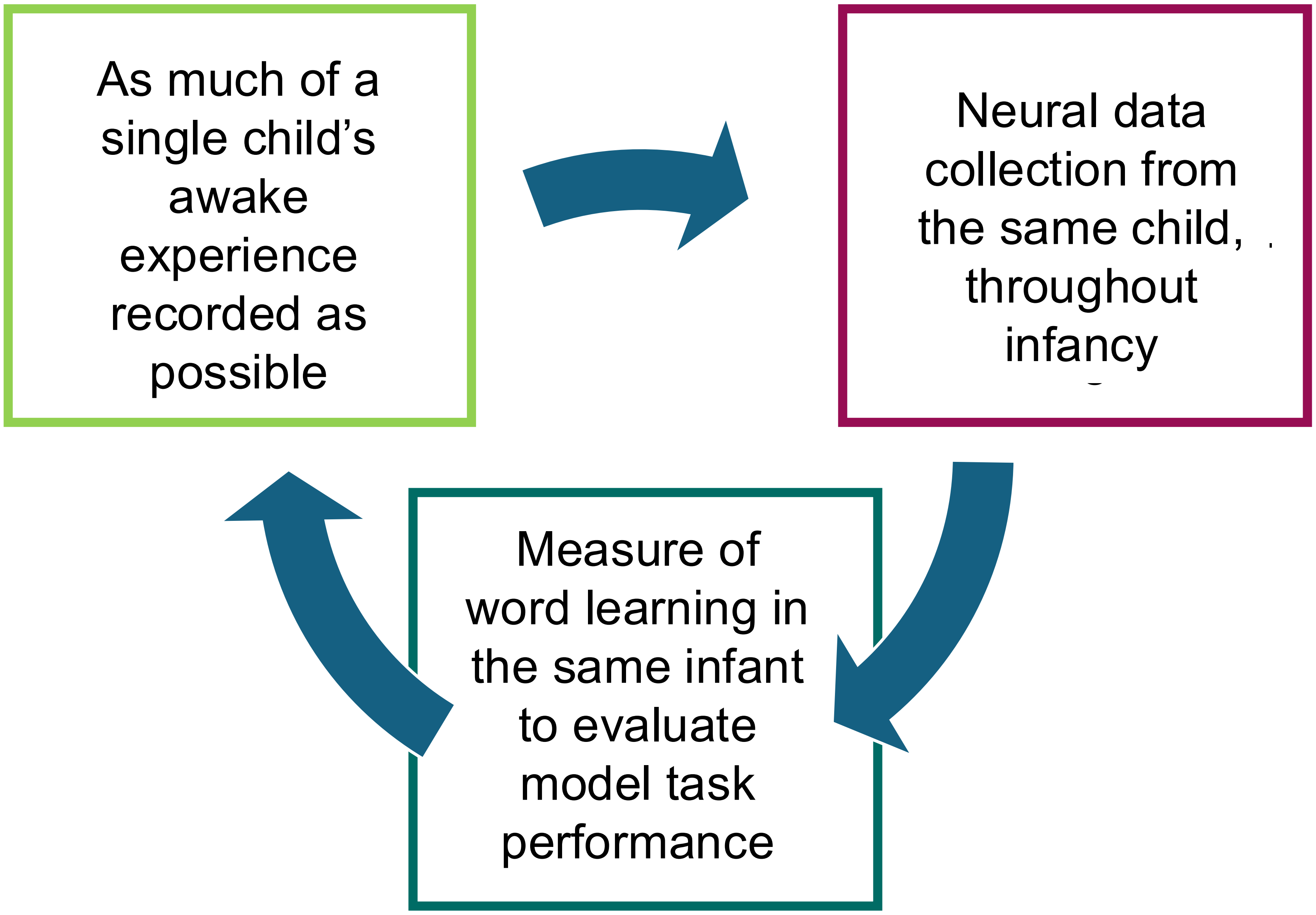
Brain

Dense sample of the brain's
response to audiovisual static
and dynamic stimuli

Behavior

Dense sample of noun-label knowledge that is expected
to show longitudinal change




As much of a single child's awake experience recorded as possible
Neural data collection from the same child, throughout infancy
Measure of word learning in the same infant to evaluate model task performance

References

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. Neuron, 93(3), 480-490.

Lewkowicz, D. J. (2011). The biological implausibility of the nature–nurture dichotomy and what it means for the study of infancy. Infancy, 16(4), 331-367.

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. Nature neuroscience, 20(10), 1404-1412.

Frank, M. C. (2023). Openly accessible LLMs can help us to understand human cognition. Nature Human Behaviour, 7(11), 1825-1827.

Kanwisher, N., Khosla, M., & Dobs, K. (2023). Using artificial neural networks to ask 'why'questions of minds and brains. Trends in Neurosciences, 46(3), 240-254.

Orhan, A. E., & Lake, B. M. (2024). Learning high-level visual representations from a child's perspective without strong inductive biases. Nature Machine Intelligence, 6(3), 271-283.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nature neuroscience, 19(3), 356-365

Kunin, L., Piccolo, S. H., Saxe, R., & Liu, S. (2024). Perceptual and conceptual novelty independently guide infant looking behaviour: a systematic review and meta-analysis. Nature Human Behaviour, 8(12), 2342-2356.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. science, 274(5294), 1926-1928.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. Cognition, 106(2), 665-681.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. Trends in cognitive sciences, 22(4), 325-336.

Bambach, S., Crandall, D., Smith, L., & Yu, C. (2018). Toddler-inspired visual object learning. Advances in neural information processing systems, 31.

References

Jayaraman, S., & Smith, L. B. (2019). Faces in early visual environments are persistent not just frequent. Vision research, 157, 213-221.
Franchak, J. M. (2019). Changing opportunities for learning in everyday life: Infant body position over the first year. Infancy, 24(2), 187-209
Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1711), 20160055.
Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. Infant behavior and development, 10(3), 279-293.
Kosakowski, H. L., Cohen, M. A., Takahashi, A., Keil, B., Kanwisher, N., & Saxe, R. (2022). Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. Current Biology, 32(2), 265-274
Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., ... & Saxe, R. (2017). Organization of high-level visual cortex in human infants. Nature communications, 8(1), 13995
Ellis, C. T., Skalaban, L. J., Yates, T. S., Bejjanki, V. R., Córdova, N. I., & Turk-Browne, N. B. (2020). Re-imagining fMRI for awake behaving infants. Nature Communications, 11(1), 4523
Ellis, C. T., Yates, T. S., Skalaban, L. J., Bejjanki, V. R., Arcaro, M. J., & Turk-Browne, N. B. (2021). Retinotopic organization of visual cortex in human infants. Neuron, 109(16), 2616-2626.
Corvilain, P., Wens, V., Bourguignon, M., Capparini, C., Fourdin, L., Ferez, M., ... & Bertels, J. (2025). Pushing the boundaries of MEG based on optically pumped magnetometers towards early human life. Imaging Neuroscience, 3, imag_a_00489
Gervain, J., Minagawa, Y., Emberson, L., & Lloyd-Fox, S. (2023). Using functional near-infrared spectroscopy to study the early developing brain: future directions and new challenges. Neurophotonics, 10(2), 023519-023519.
O'Doherty, C., Dineen, Á.T., Truzzi, A., King, G., Zaadnoordijk, L., Harrison, K., ... & Cusack, R. (2026). Infants have rich visual categories in ventrotemporal cortex at 2 months of age. Nature Neuroscience, 10.1038/s41593-025-02187-8.

References

Smith, L. B., & Slone, L. K. (2017). A developmental approach to machine learning?. Frontiers in psychology, 8, 296143.
Sullivan, J., Mei, M., Perfors, A., Wojcik, E., & Frank, M. C. (2021). SAYCam: A large, longitudinal audiovisual dataset recorded from the infant’s perspective. Open mind, 5, 20-29.
Long, B., Sparks, R. Z., Xiang, V., Stojanov, S., Yin, Z., Keene, G. E., ... & Frank, M. C. (2024). The BabyView dataset: High-resolution egocentric videos of infants' and young children's everyday experiences. arXiv preprint arXiv:2406.10447.
Vong, W. K., Wang, W., Orhan, A. E., & Lake, B. M. (2024). Grounded language acquisition through the eyes and ears of a single child. Science, 383(6682), 504-511.
Lu, Z., Thorat, S., Cichy, R. M., & Kietzmann, T. C. (2025). Adopting a human developmental visual diet yields robust, shape-based AI vision. arXiv preprint arXiv:2507.03168.
Vogelsang, L., Gilad-Gutnick, S., Ehrenberg, E., Yonas, A., Diamond, S., Held, R., & Sinha, P. (2018). Potential downside of high initial visual acuity. *Proceedings of the National Academy of Sciences*, *115*(44), 11333-11338.
Tan, A., Yu, C., Long, B., Ma, W., Murray, T., Silverman, R., ... & Frank, M. C. (2024). Devbench: A multimodal developmental benchmark for language learning. *Advances in Neural Information Processing Systems*, *37*, 77445-77467.
Venkatesh, R., Kotar, K., Chen, L. N., Kim, S., Wheeler, L. T., Watrous, J., ... & Yamins, D. (2025). Discovering and using Spelke segments. *arXiv preprint arXiv:2507.16038*.