A generalist model for intracortical motor brain-computer interfaces

Preprint

The Team

Joel Ye
Fabio Rizzoglio
Adam Smoulder
Hongwei Mao
Xuan Ma
Patrick Marino
Raeed Chowdhury
Dalton Moore
Gary Blumenthal
William Hockeimer
Nicolas G. Kunigk
J. Patrick Mayo
Adam Rouse
Aaron Batista
Steven Chase
Charles Greenspon
Lee E. Miller
Nicholas Hatsopoulos
Andrew Schwartz
Jennifer L. Collinger
Leila Wehbe
Robert A. Gaunt

DNNs excel at
single tasks.

DNNs features
transfer across settings.

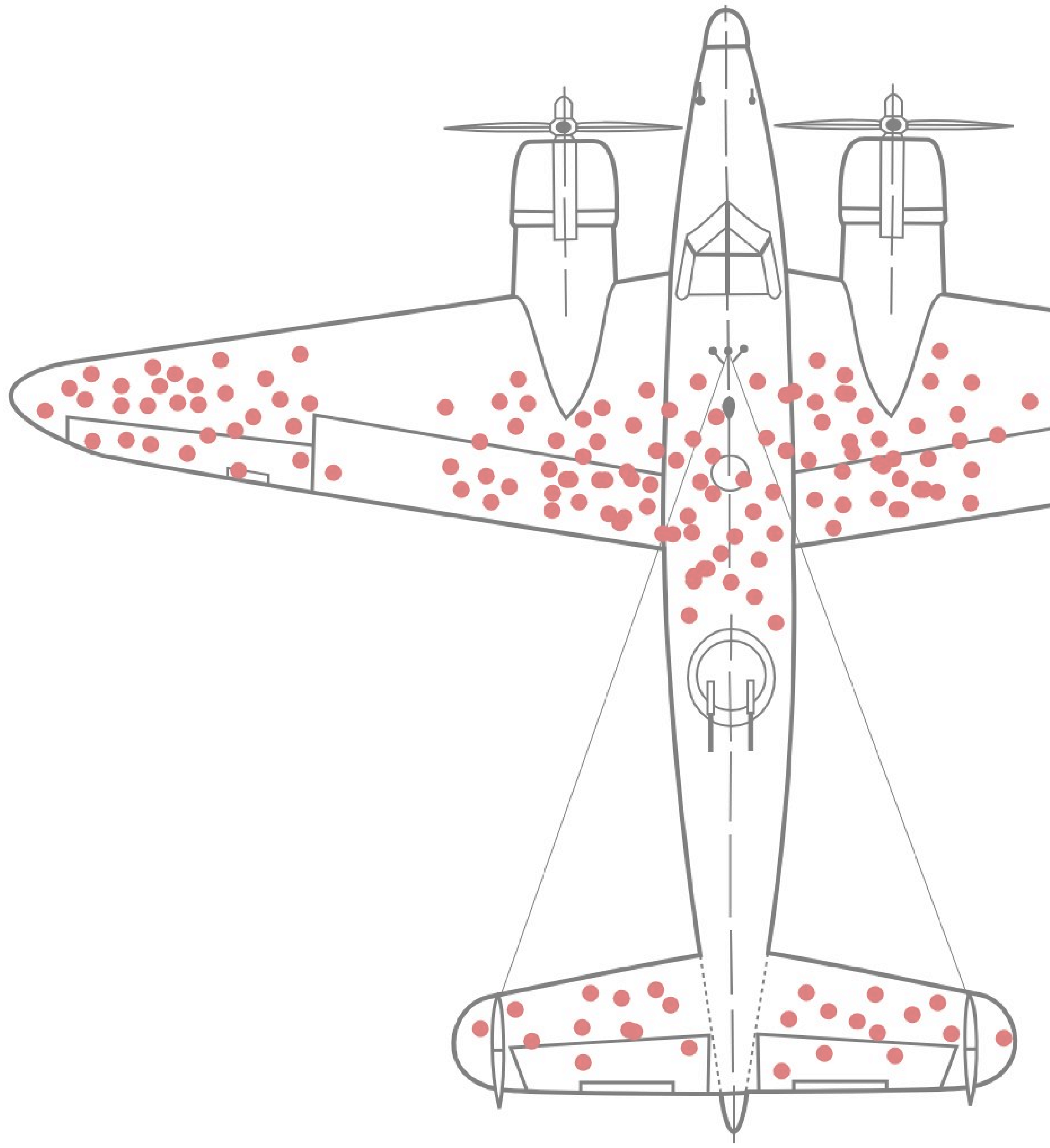
DNNs can be scaled
to **use everywhere**.

DNNs excel at **single tasks.**

DNNs features **transfer across settings.**

DNNs can be scaled to **use everywhere.**

We're here (?) Now what?



DNNs excel at **single tasks.**

DNNs features **transfer across settings.**

DNNs can be scaled to **use everywhere.**

We're here (?) Now what?

We're here (?) Now what?

In machine learning,
evaluations drive our understanding of progress.

(Guess and check will never end)

DNNs excel at **single tasks.**

DNNs features **transfer across settings.**

DNNs can be scaled to **use everywhere.**

We're here (?) Now what?

In machine learning, evaluations drive our understanding of progress.

How should we prioritize evaluations?

1. Evaluating for pragmatics
2. Evaluating for generalization

In machine learning,
evaluations drive our understanding of progress.

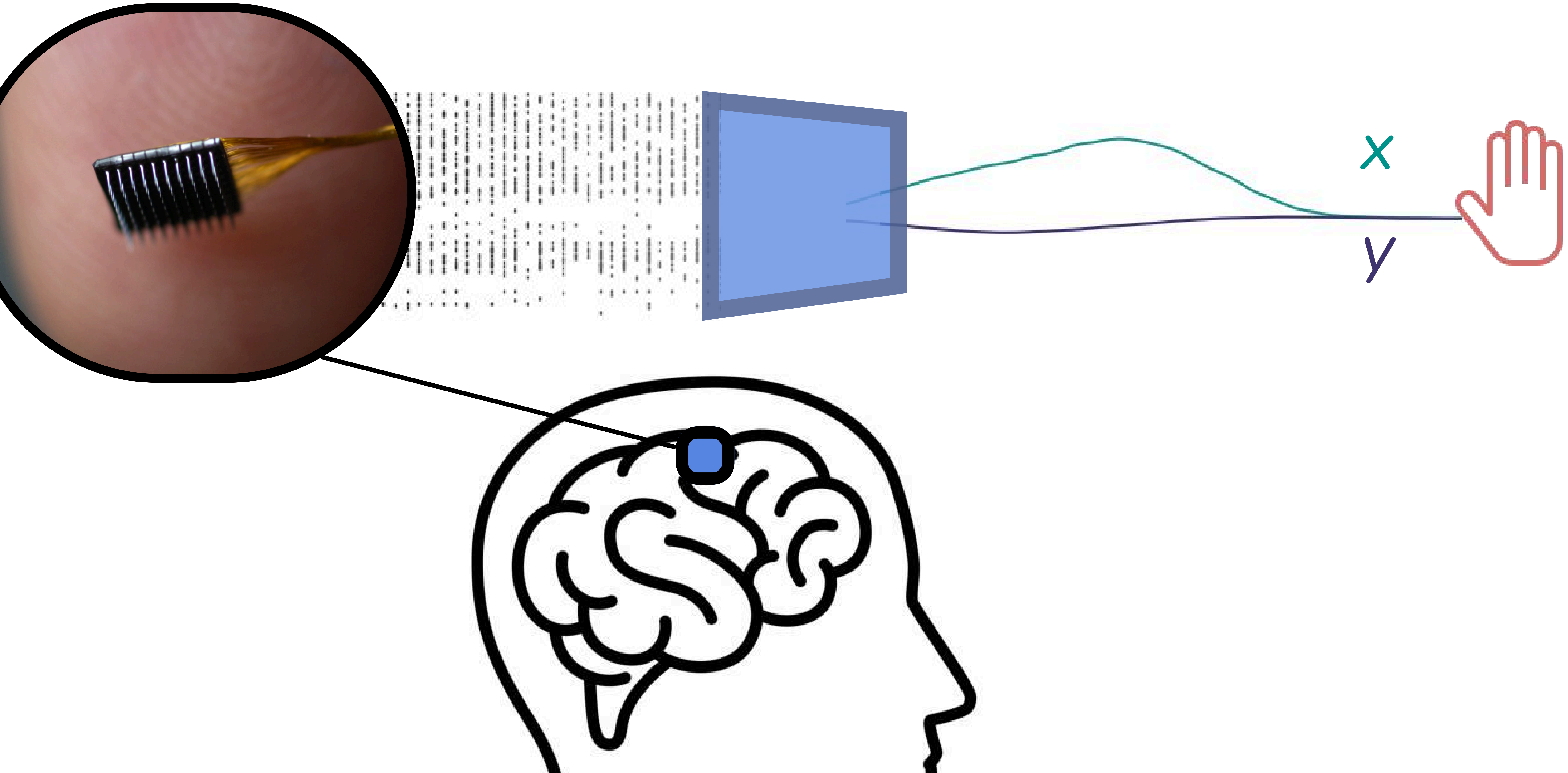
How should we prioritize evaluations?

What is the **practical** benefit of **unified** pretraining?

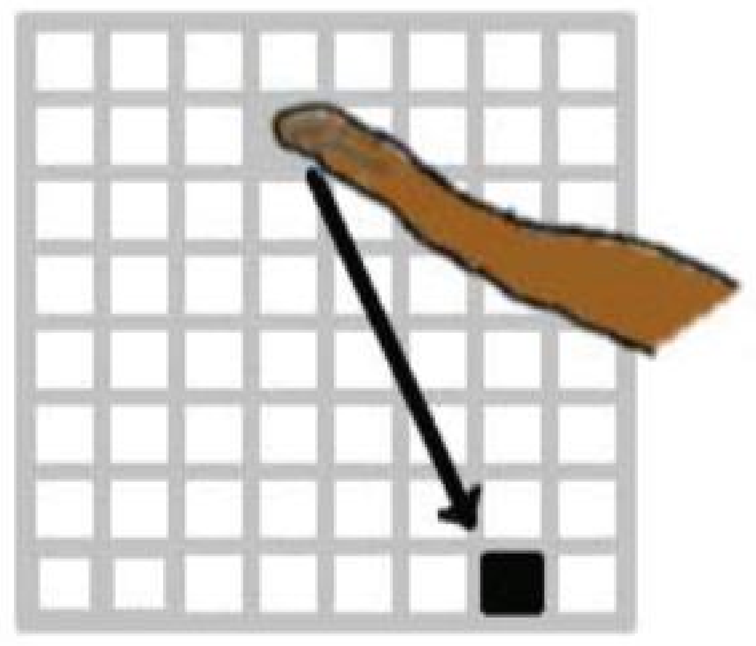
1. **Evaluating for pragmatics**

What is the **practical** benefit of **unified** pretraining?

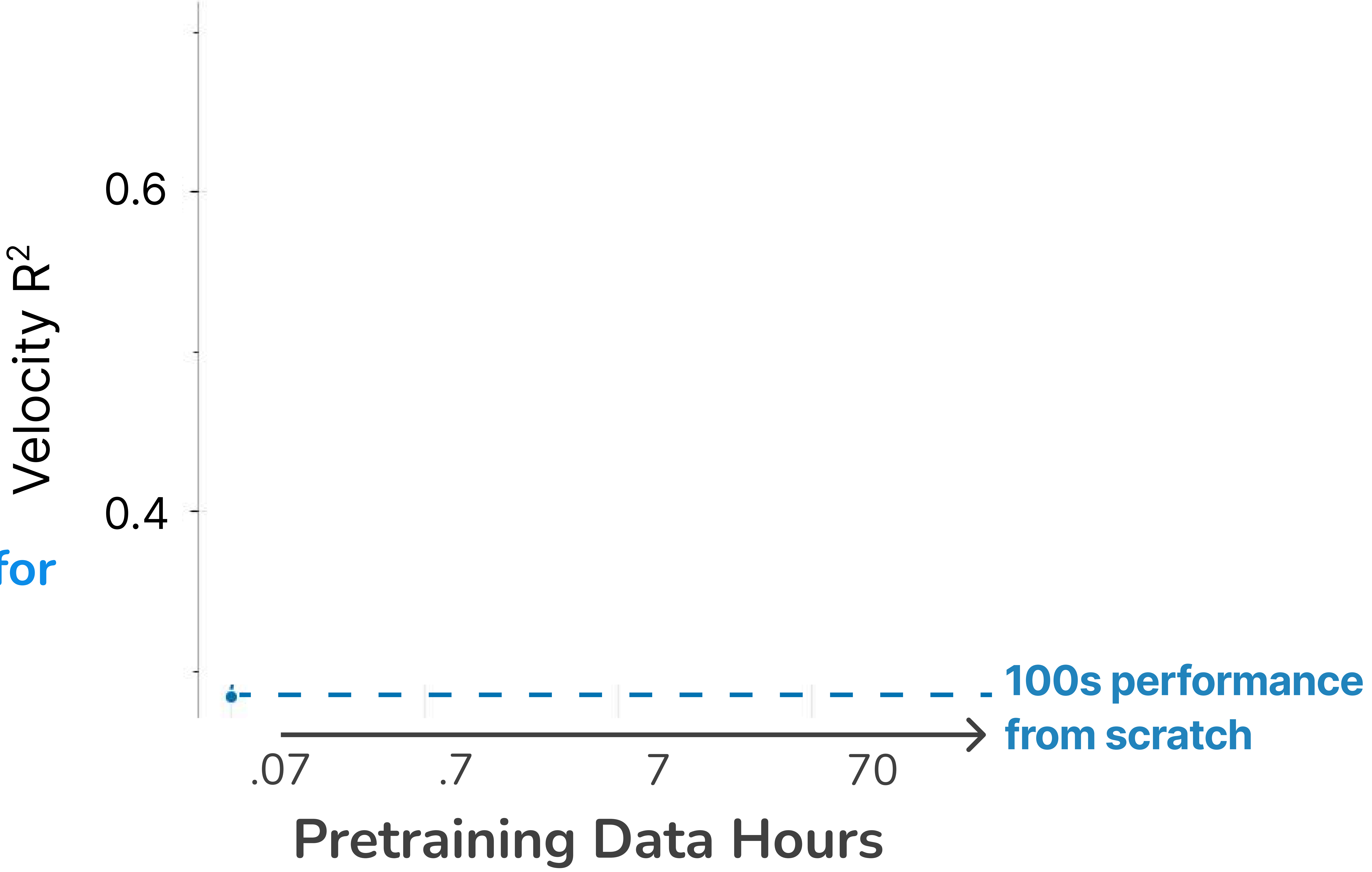
1. Evaluating for pragmatics



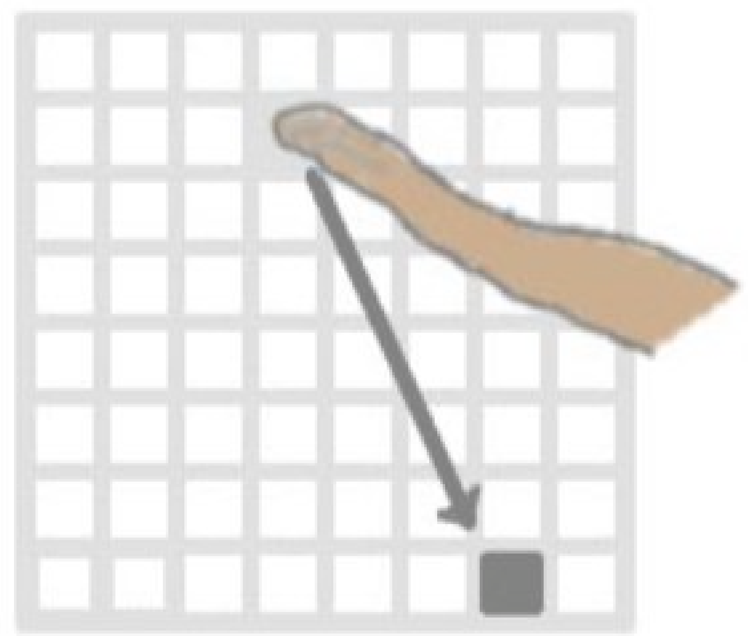
What is the **practical** benefit of **unified** pretraining?

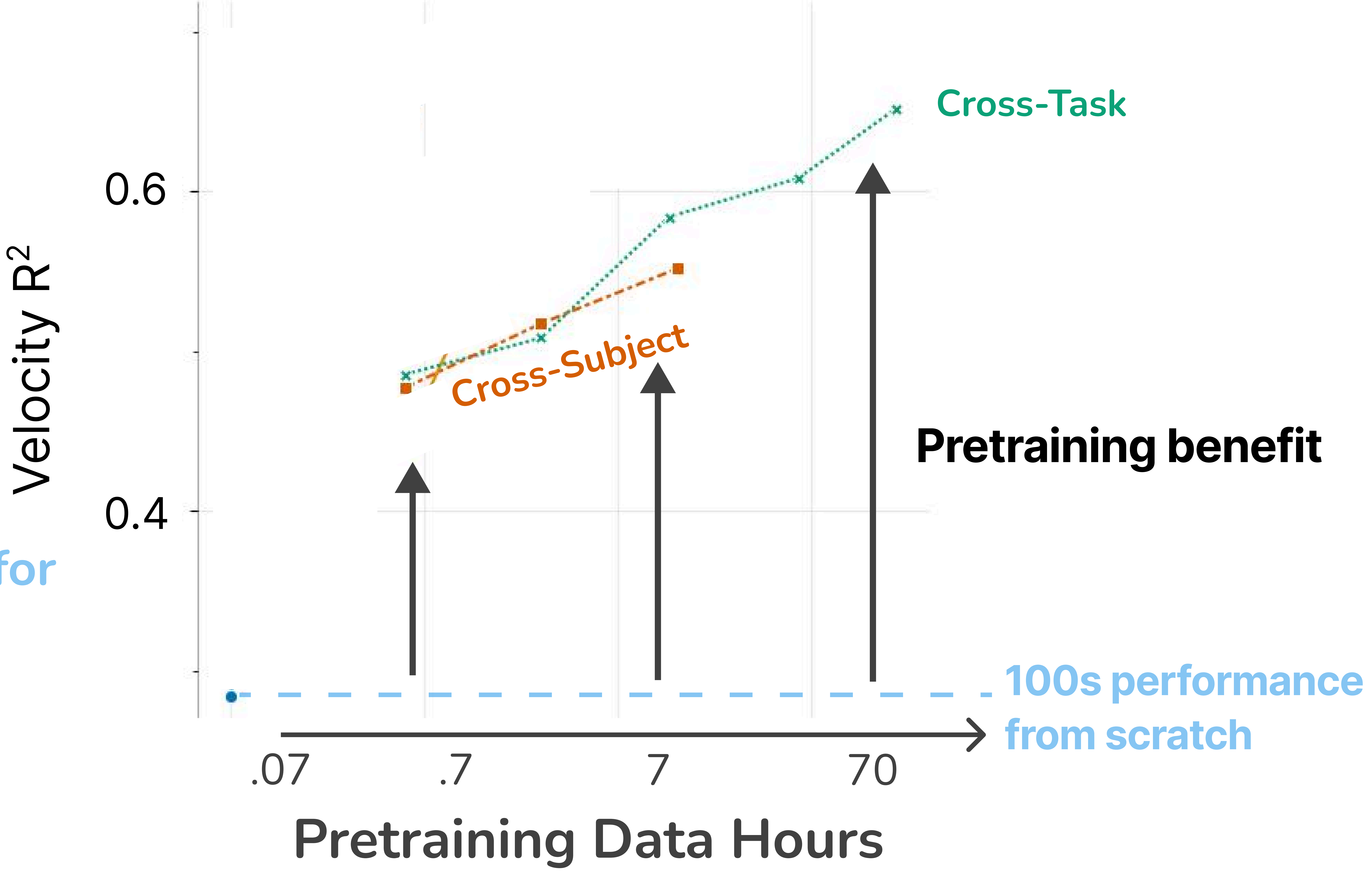


100 seconds for fine-tuning

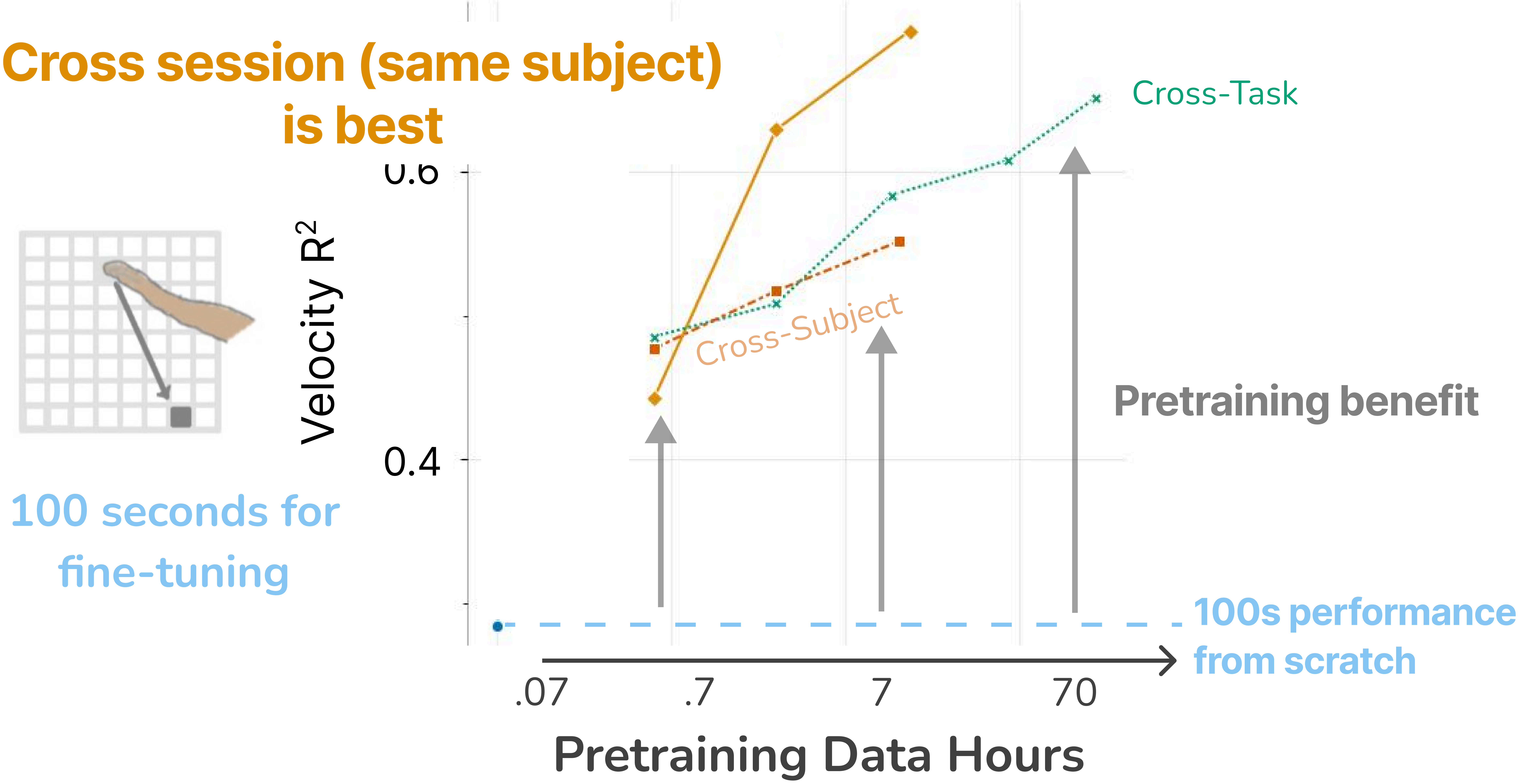


What is the **practical** benefit of **unified** pretraining?





What is the practical benefit of unified pretraining?



What is the **practical** benefit of **unified** pretraining?

THE NORM

Cross-Session

Cross-Subject **Cross-Task**

Single-session

What is the **practical** benefit of **unified** pretraining?

THE NORM

Cross-Session

Cross-Subject

Cross-Task

Single-session

PRAGMATIC EVAL

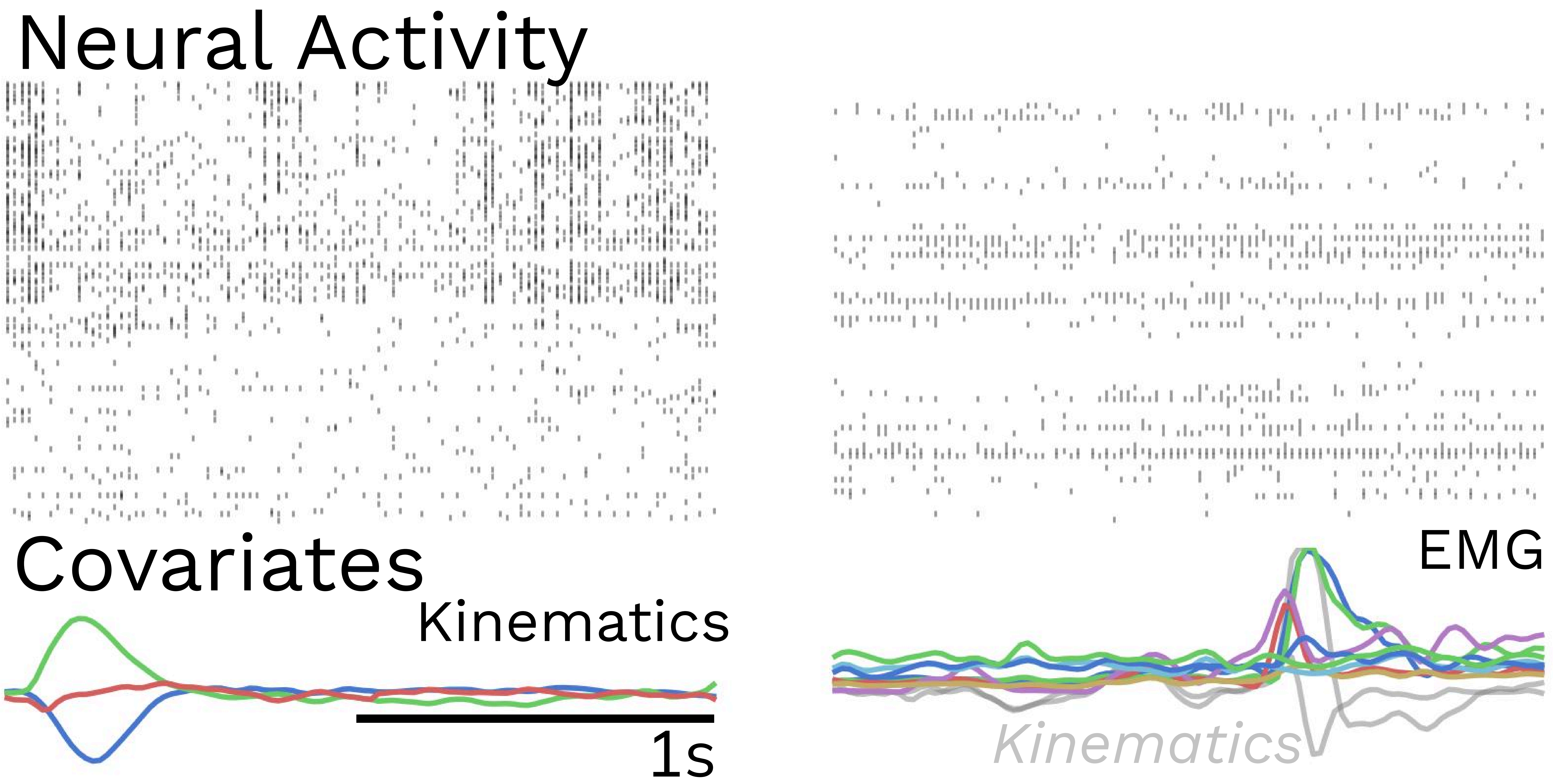
(For chronic BCI)

Cross-Subject

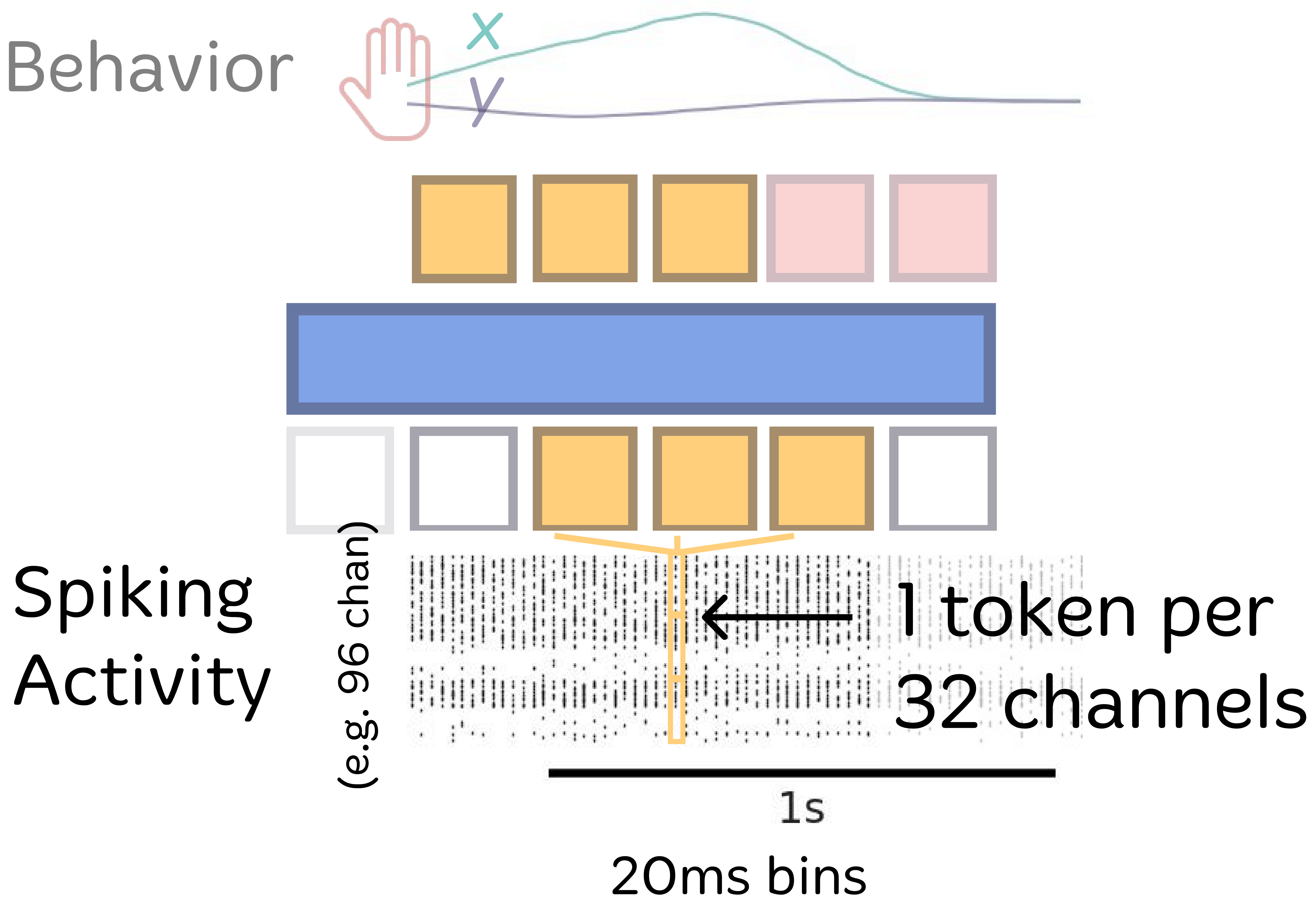
Cross-Task

Multi-session

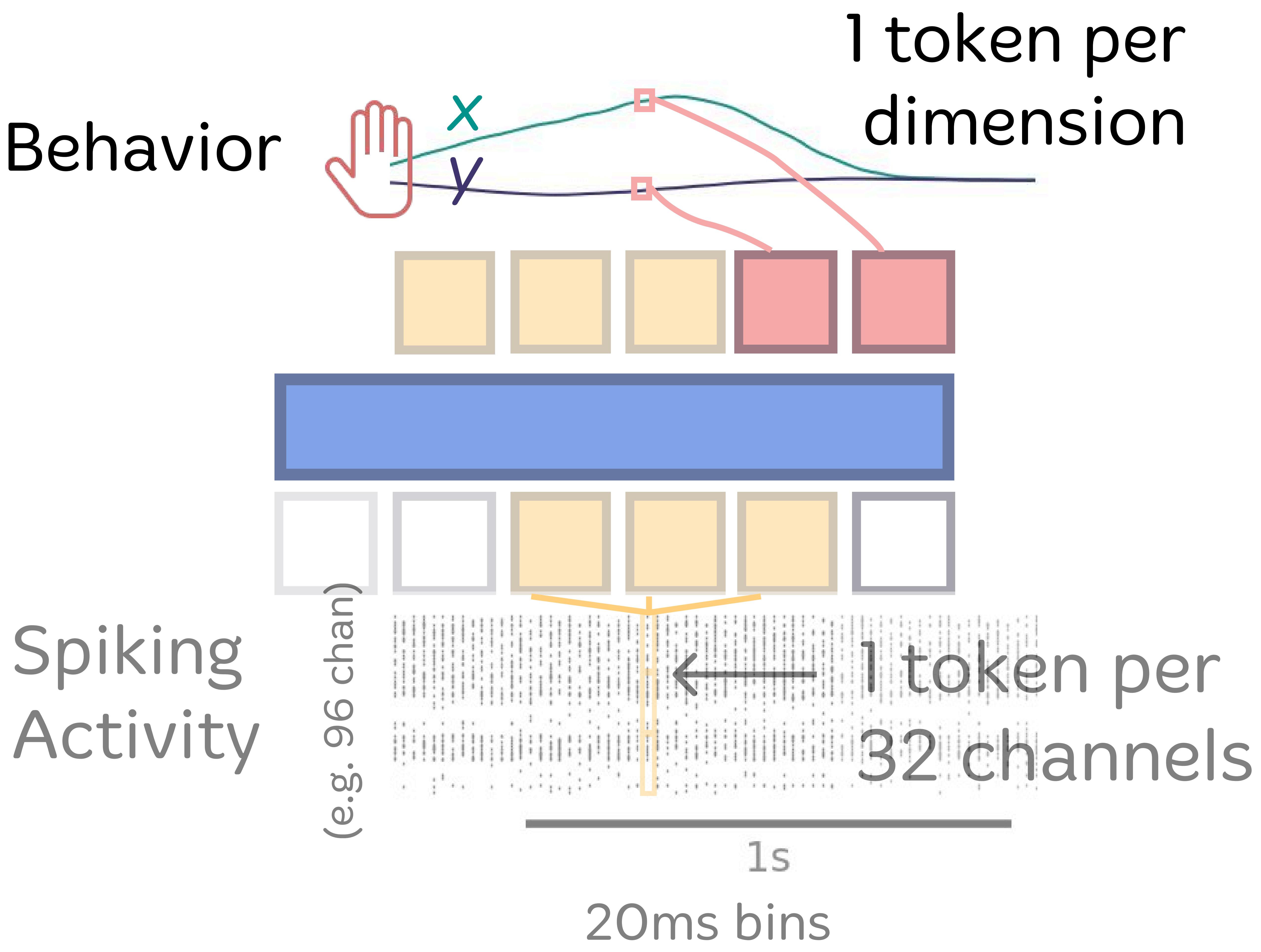
Neural Data Transformer 3 (NDT3):
Scaling motor BCI pretraining to 2000 hours



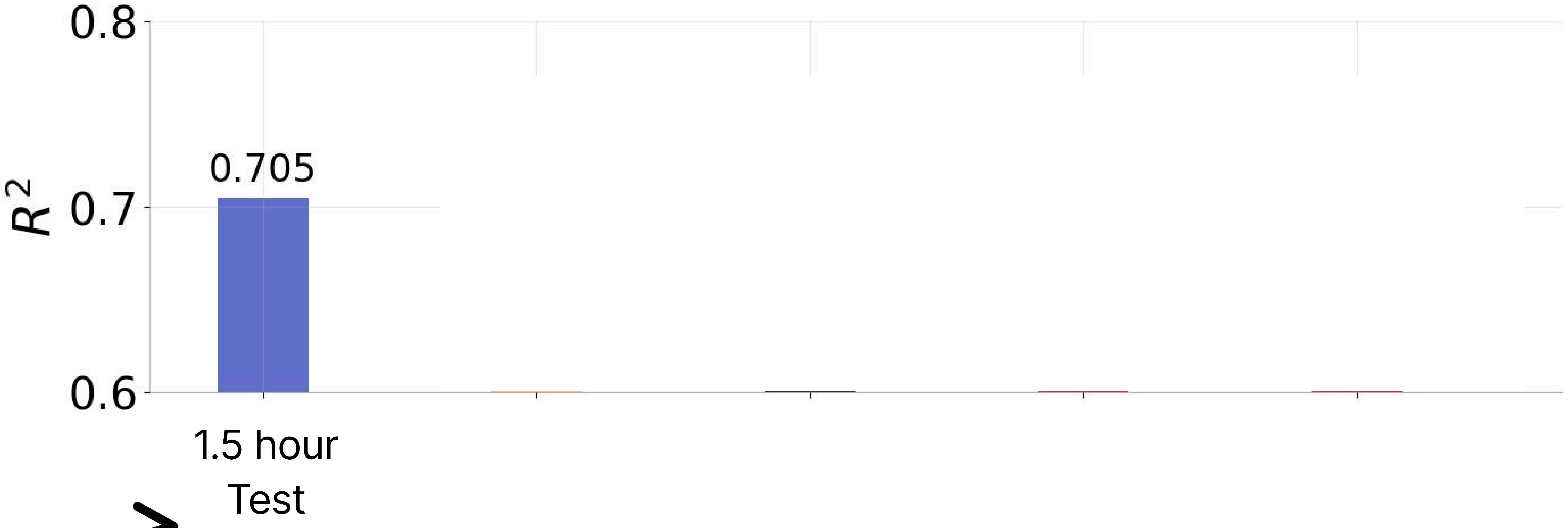
Neural Data Transformer 3 (NDT3):
Scaling motor BCI pretraining to 2000 hours



Neural Data Transformer 3 (NDT3):
Scaling motor BCI pretraining to 2000 hours

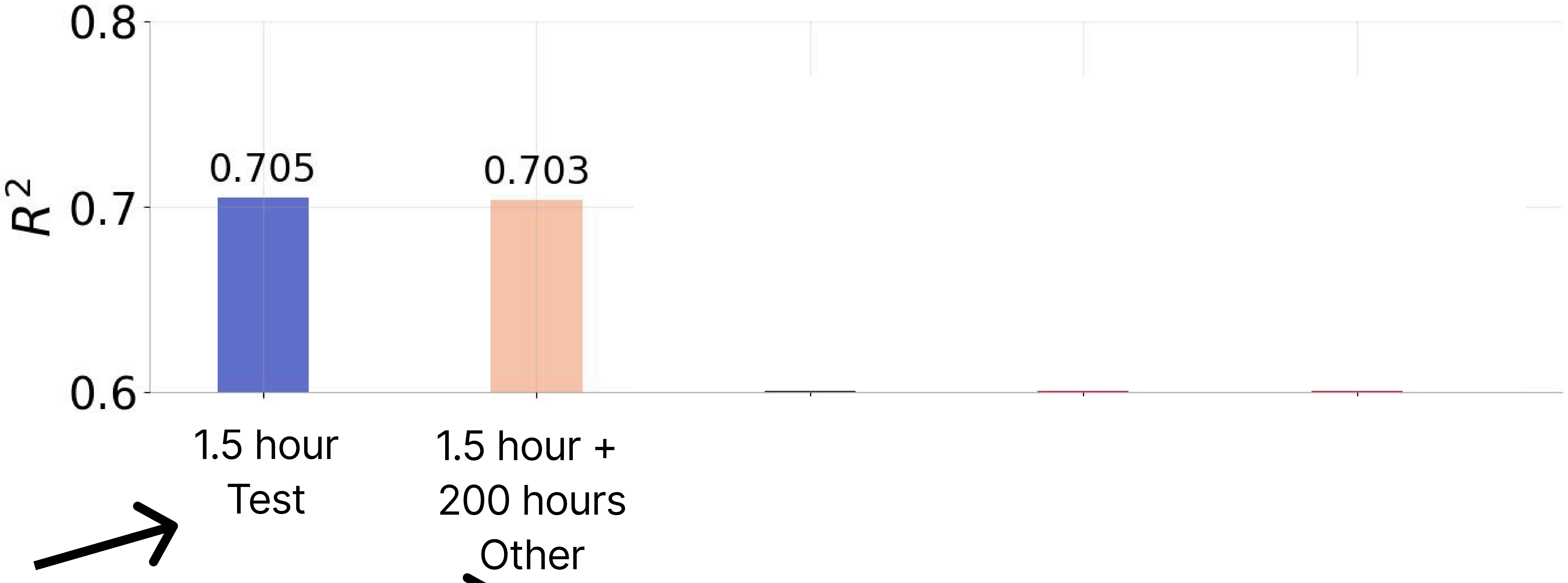


Scaled pretraining shows weak pragmatic (cross-subject) transfer



Target monkey

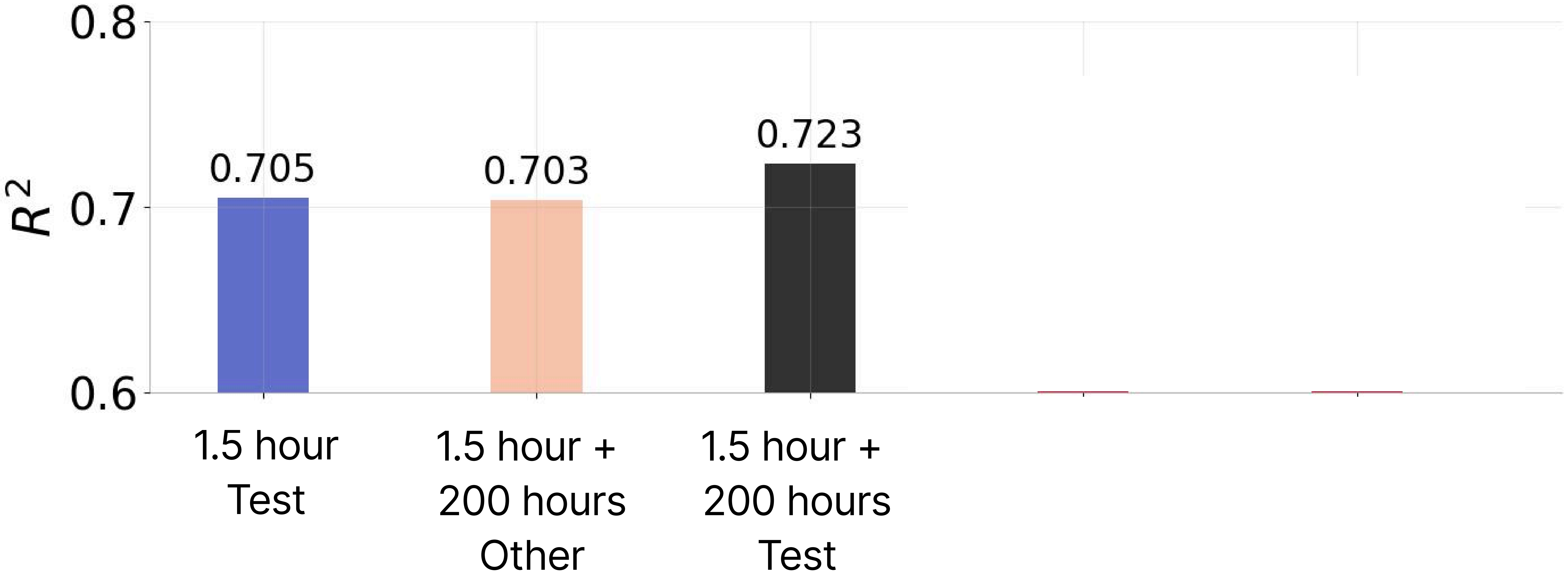
Scaled pretraining shows weak pragmatic (cross-subject) transfer



Target monkey

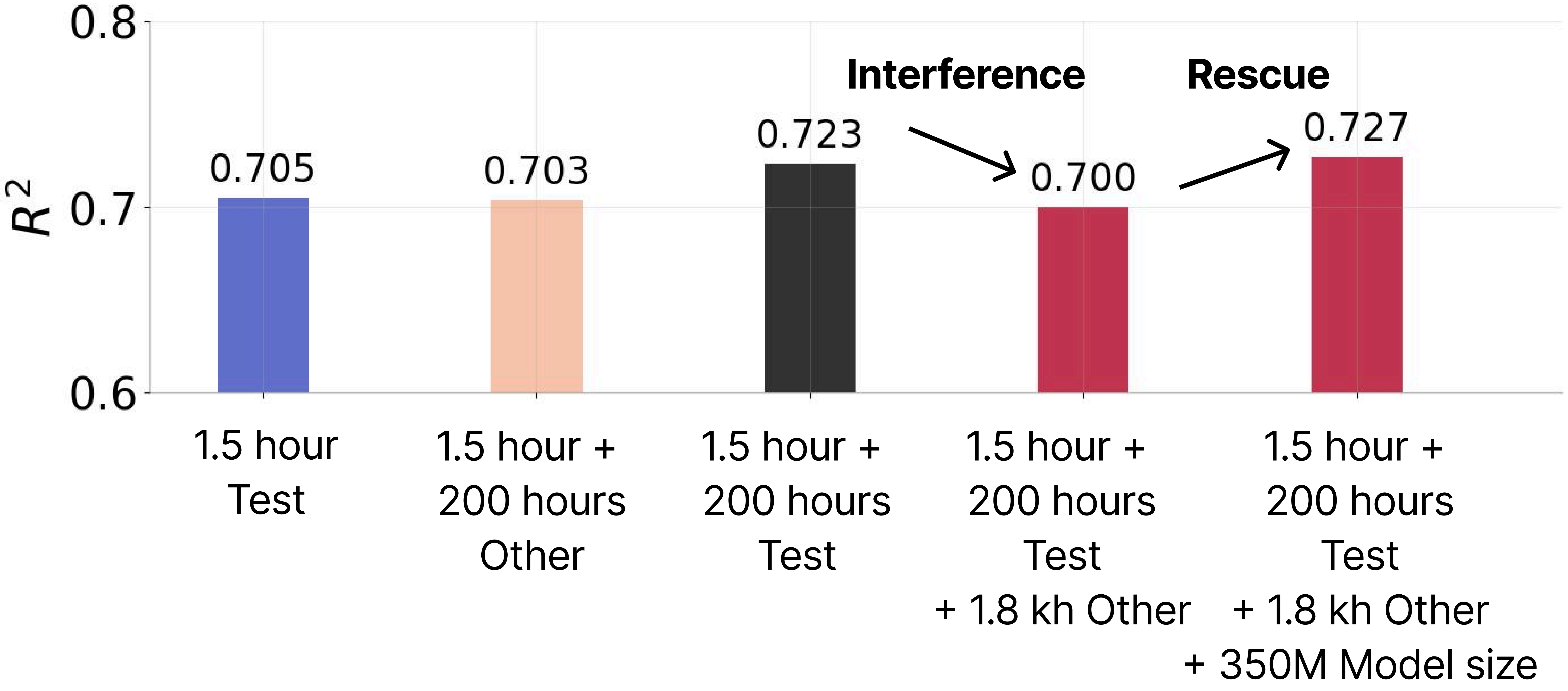
Other monkeys

Scaled pretraining shows weak pragmatic (cross-subject) transfer

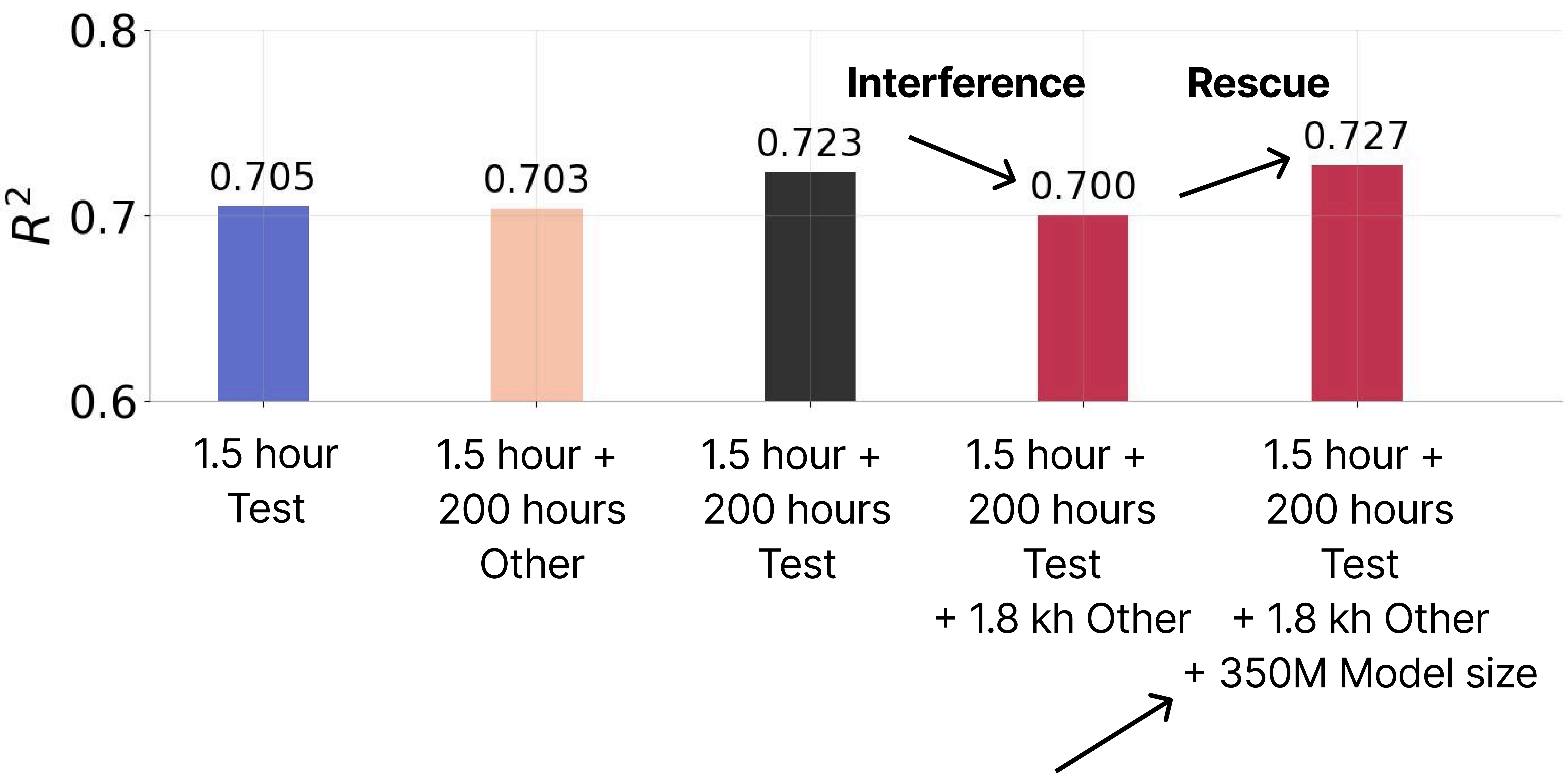


Cross-subject transfer fails for the long tail?

Scaled pretraining shows weak pragmatic (cross-subject) transfer

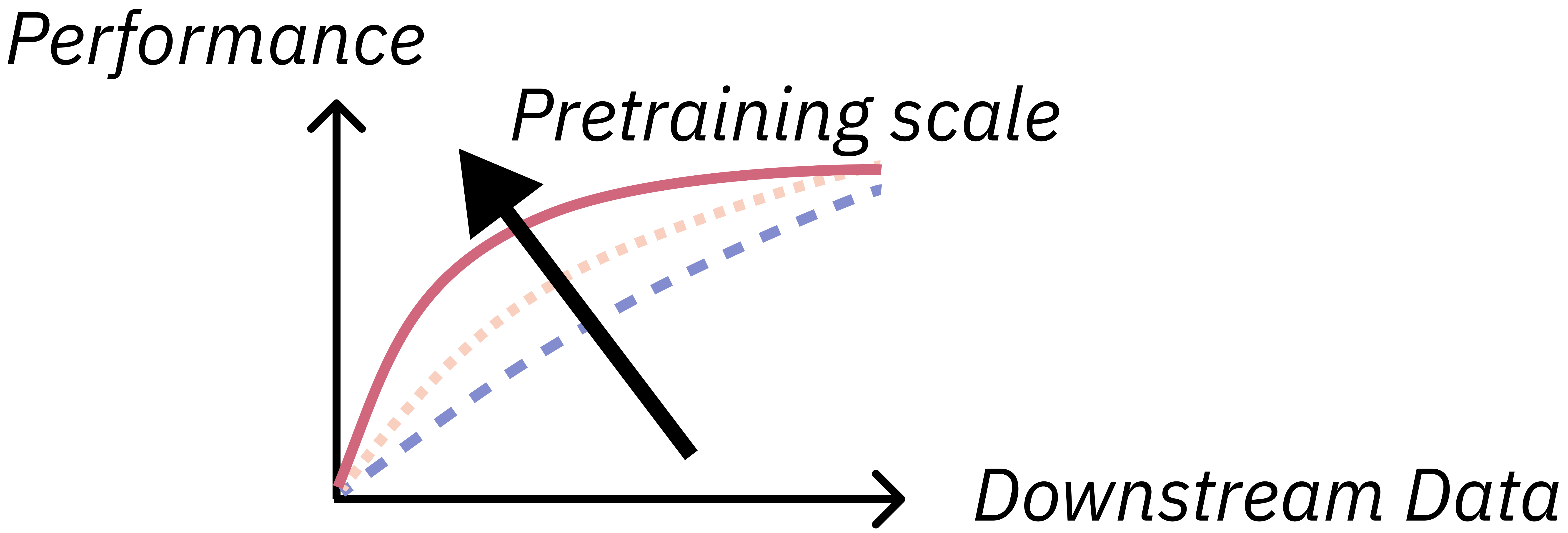


Scaled pretraining shows weak pragmatic (cross-subject) transfer

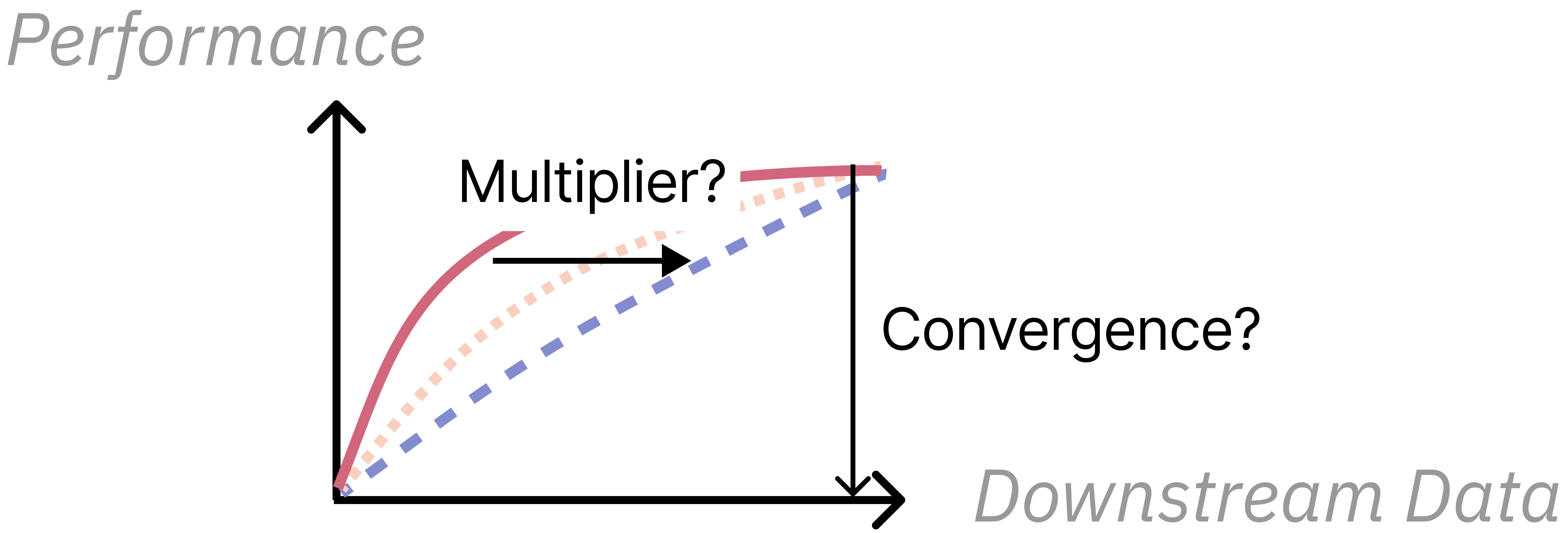


A cost to our ambition, for no gain?

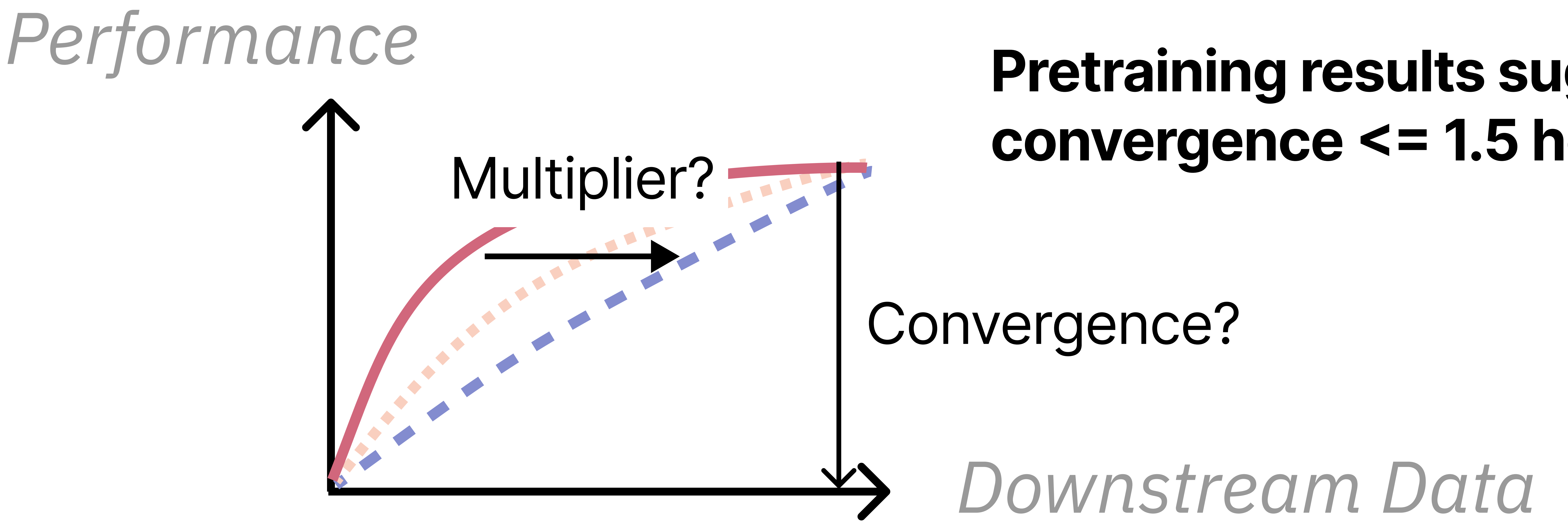
Quantifying downstream scaling gains



Quantifying downstream scaling gains

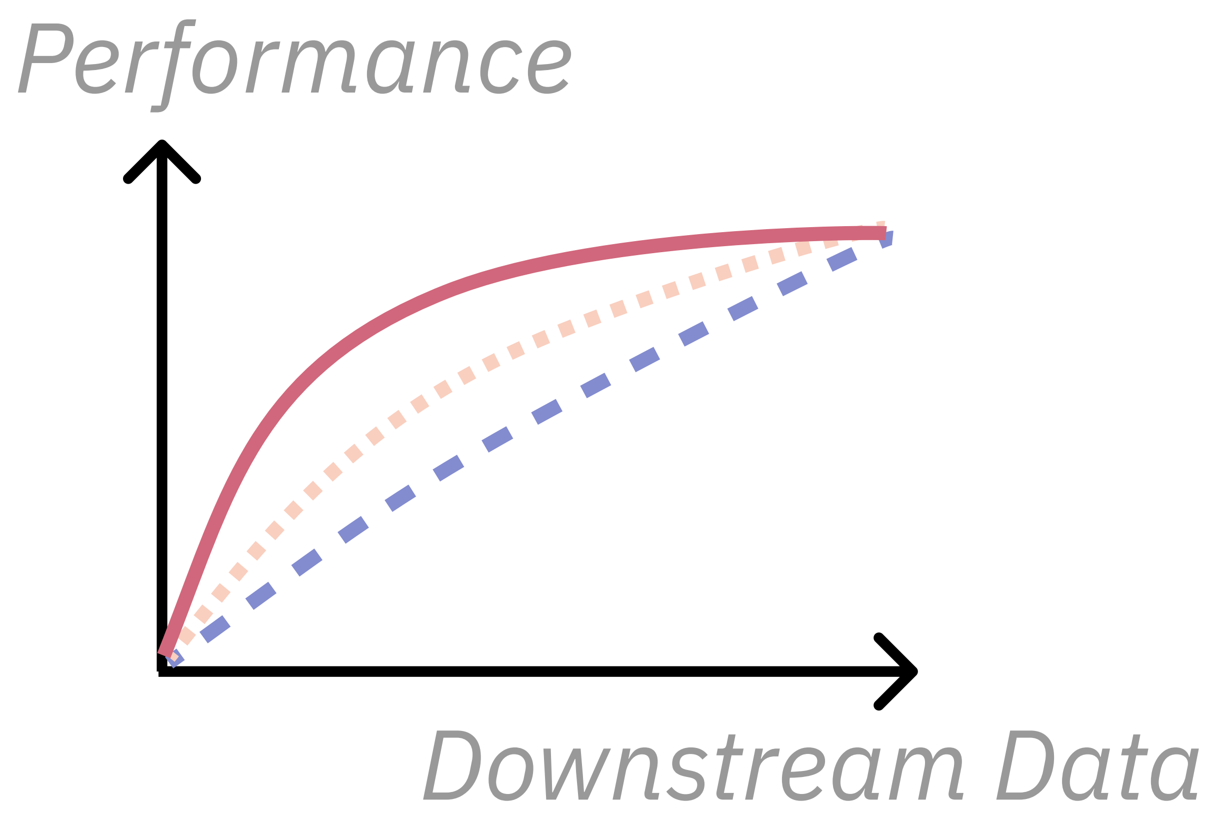


Quantifying downstream scaling gains



Pretraining results suggest convergence <= 1.5 hours

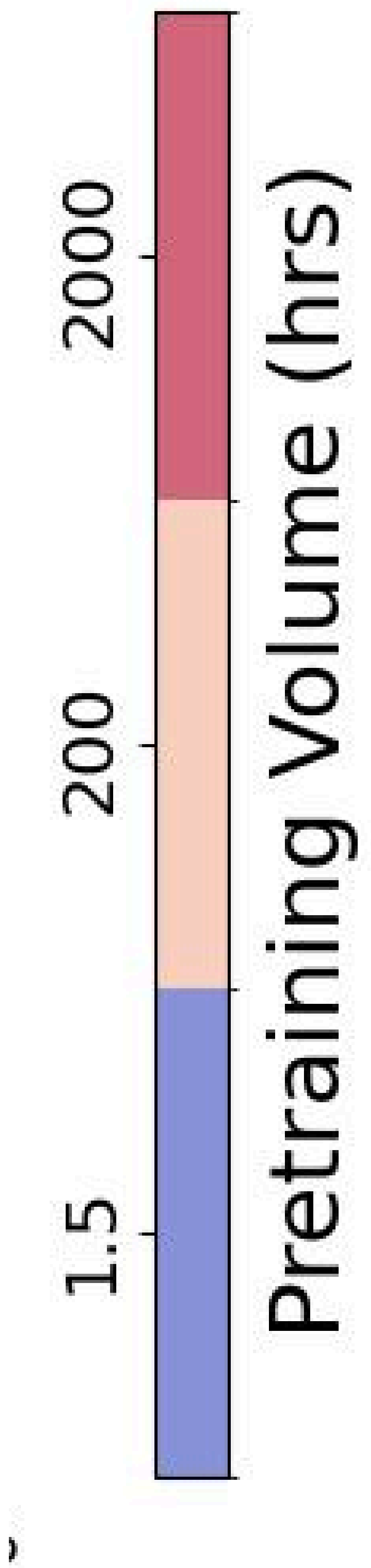
Quantifying downstream scaling gains

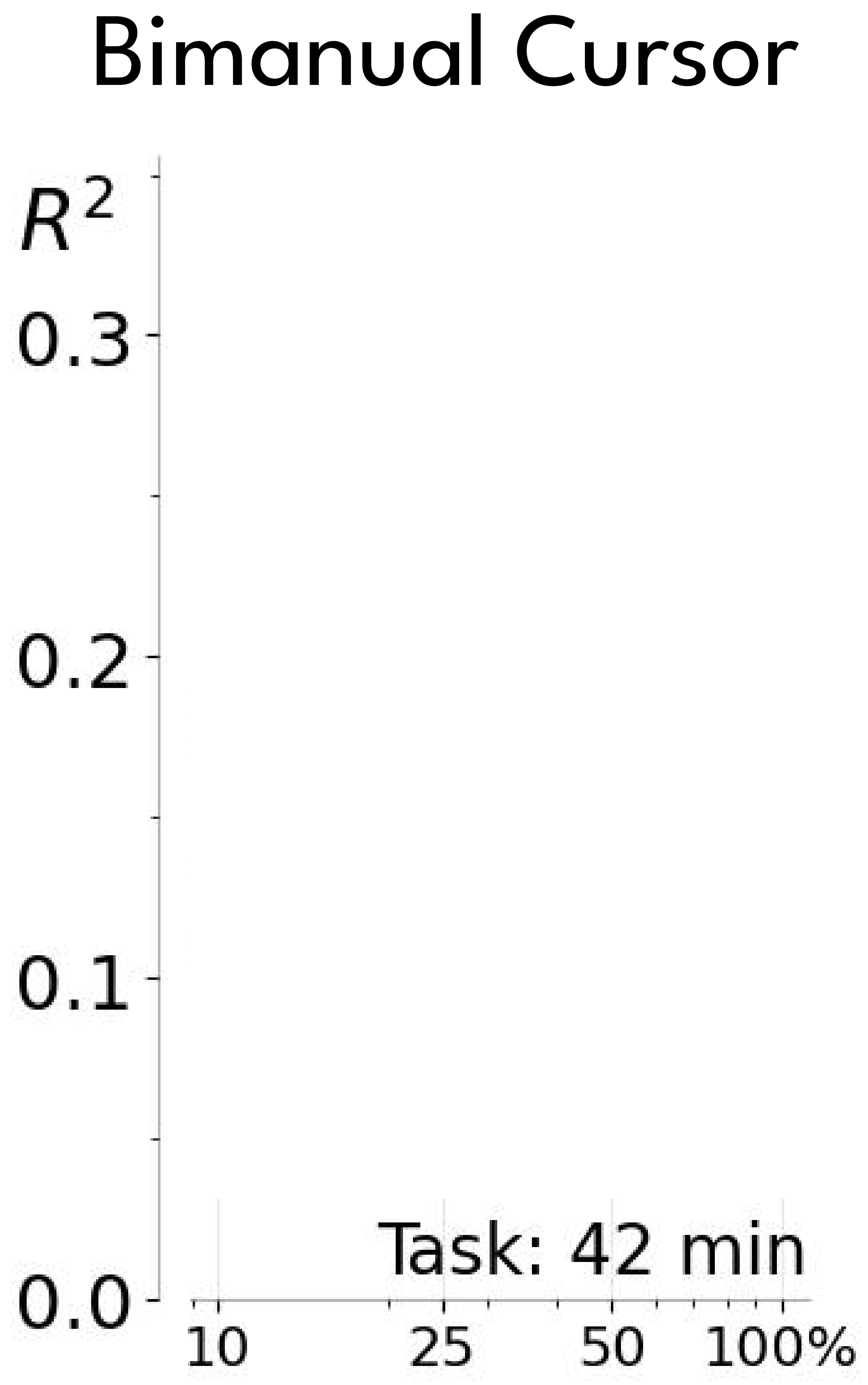


Scratch

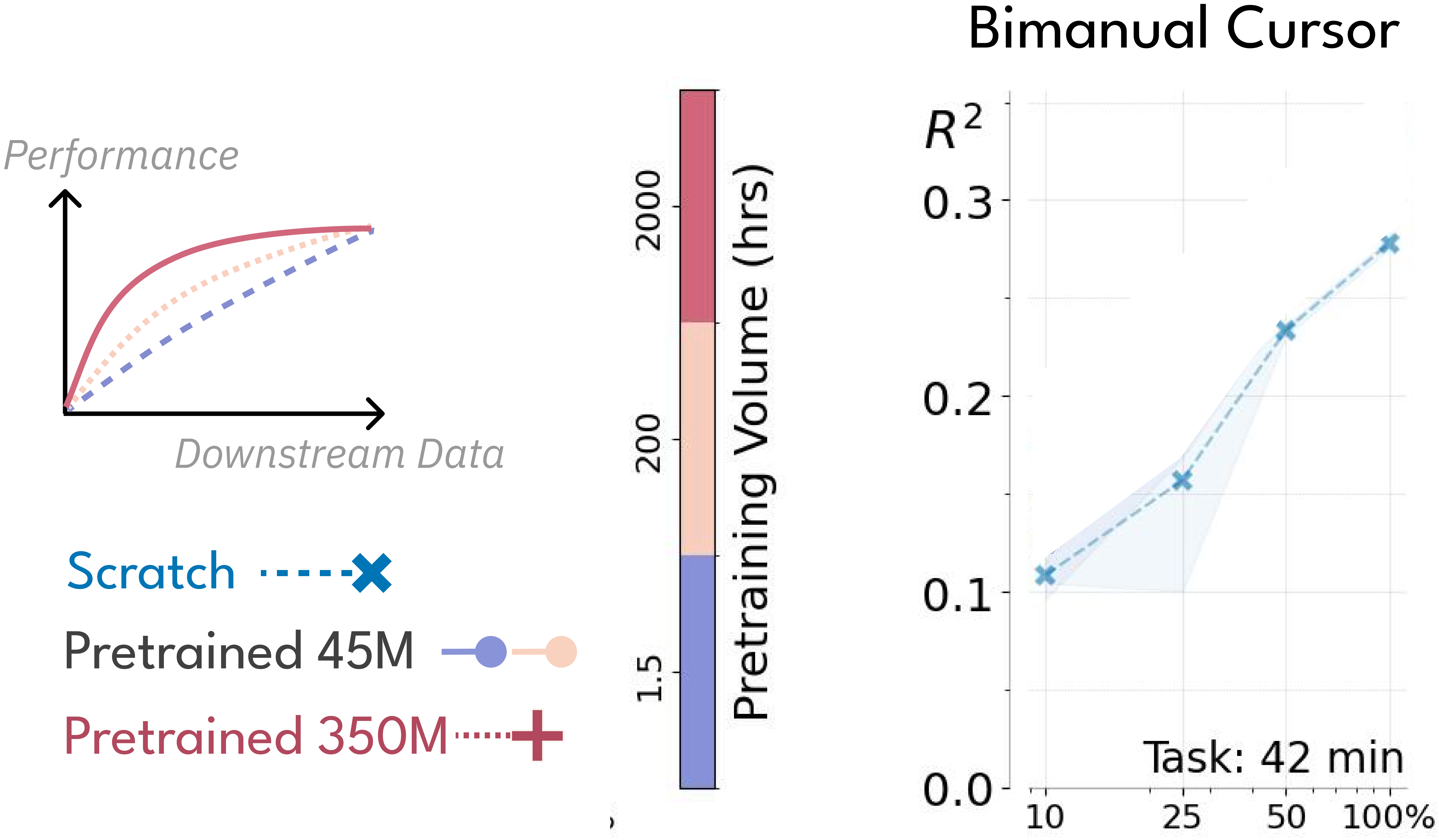
Pretrained 45M

Pretrained 350M

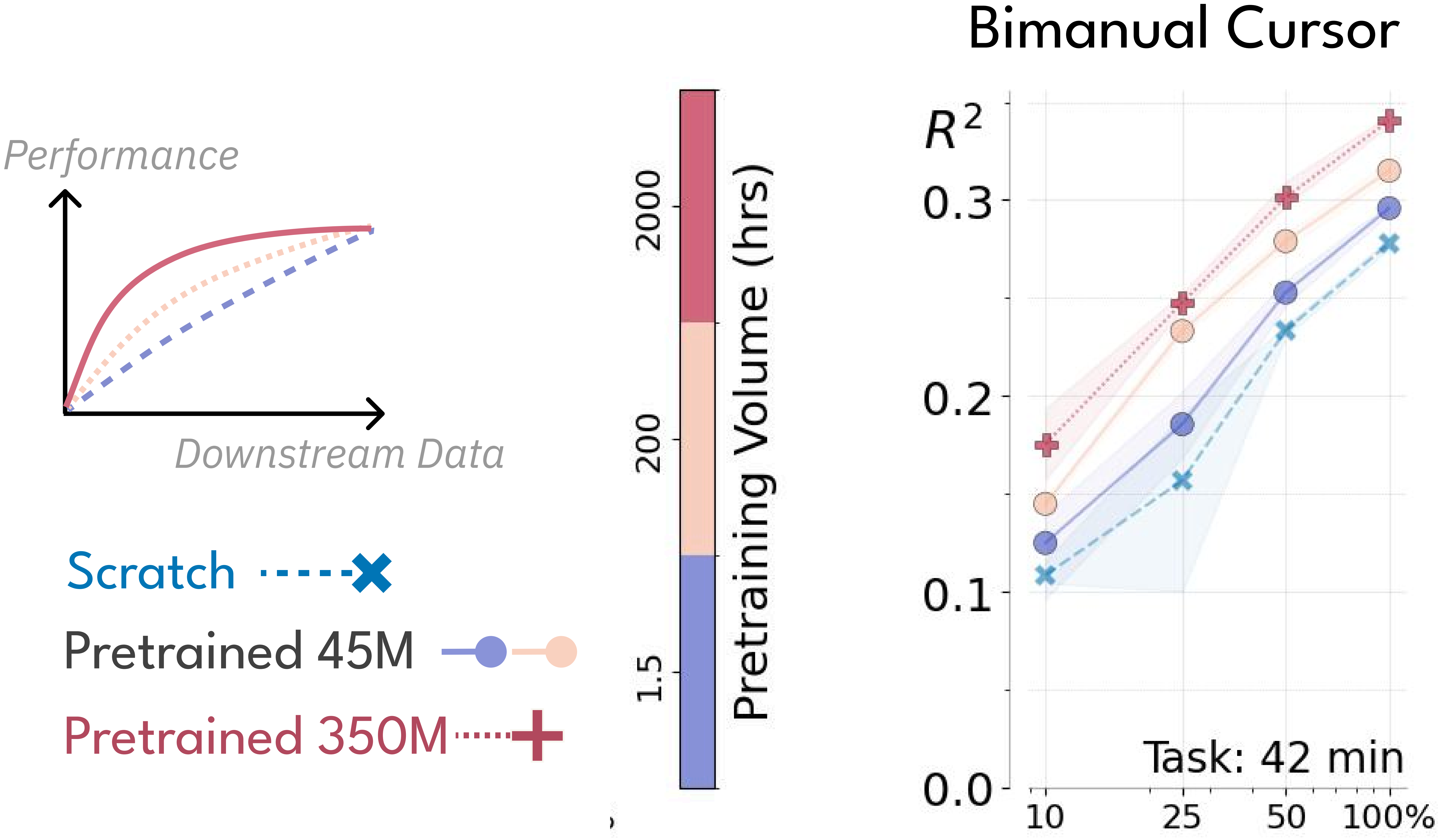




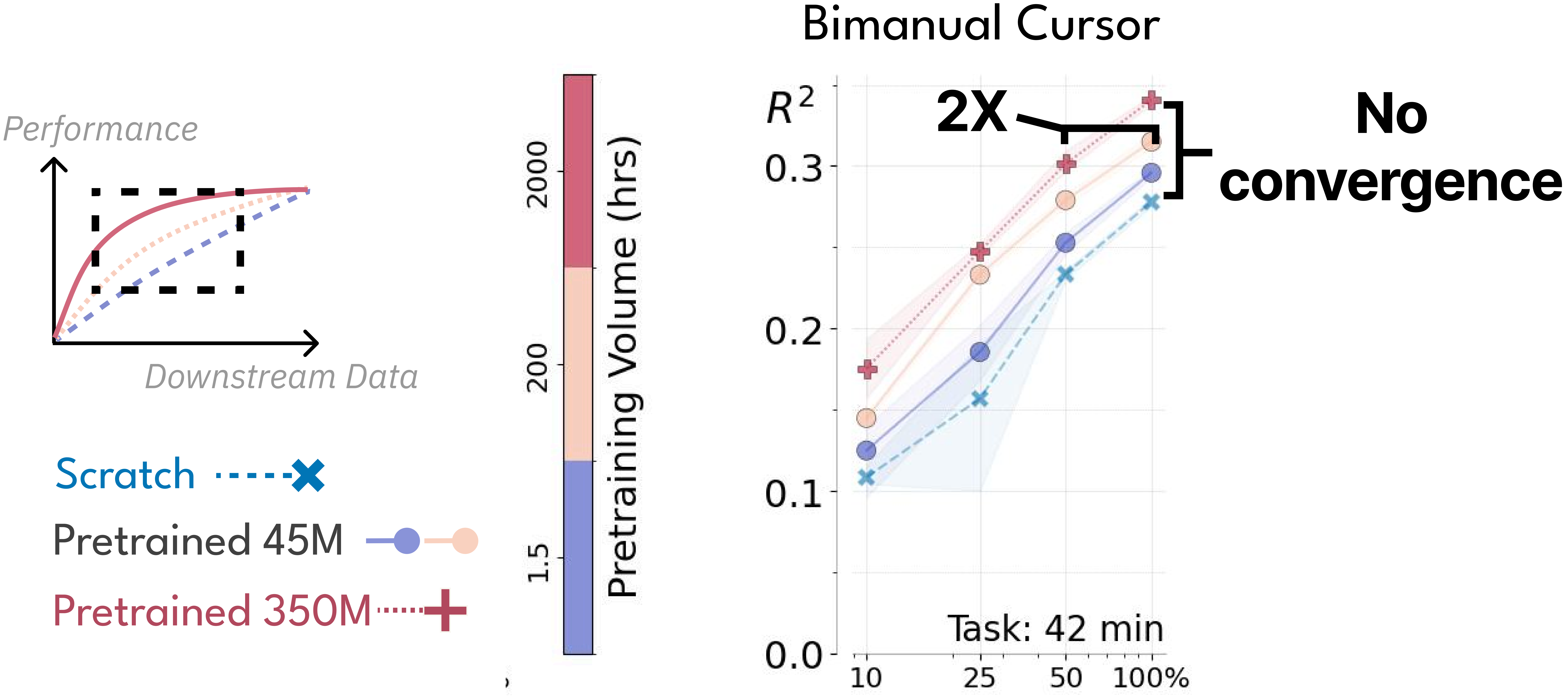
Quantifying downstream scaling gains



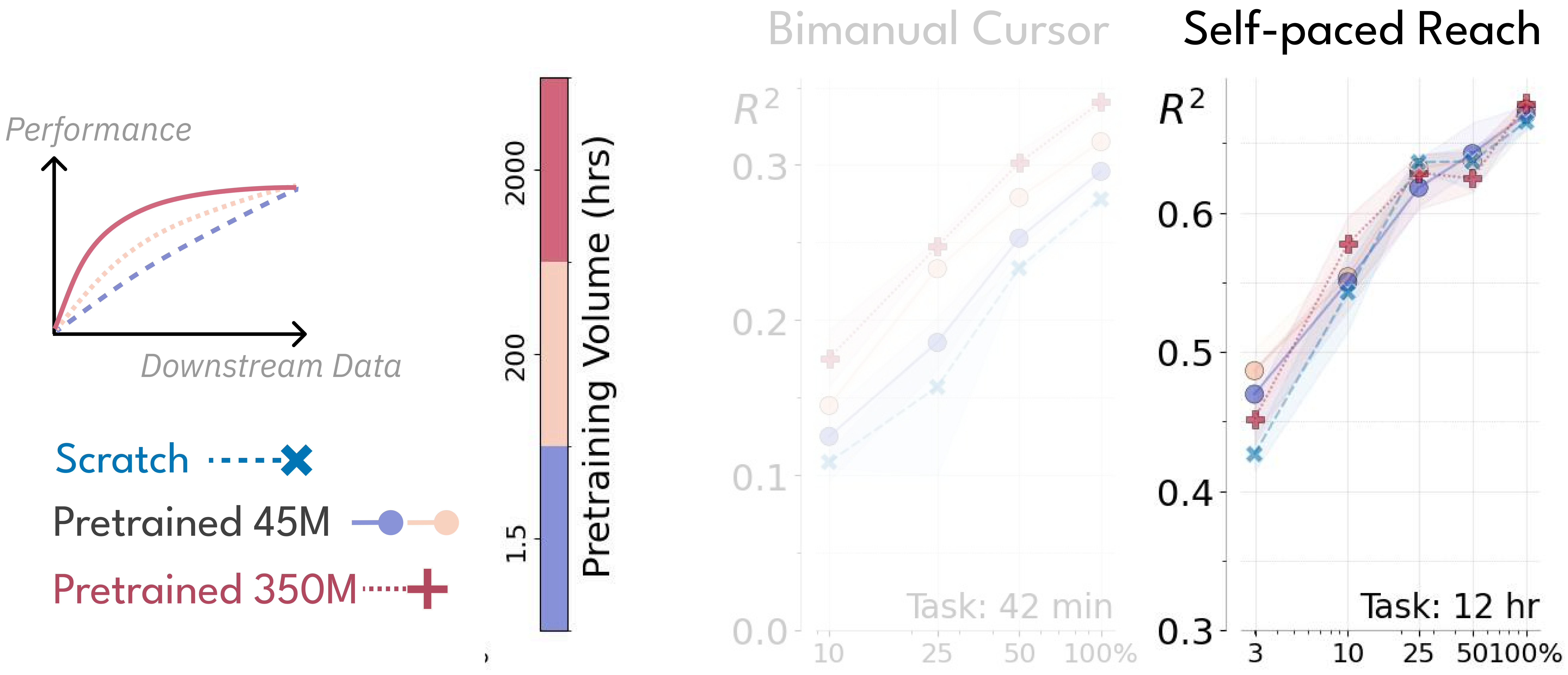
Quantifying downstream scaling gains



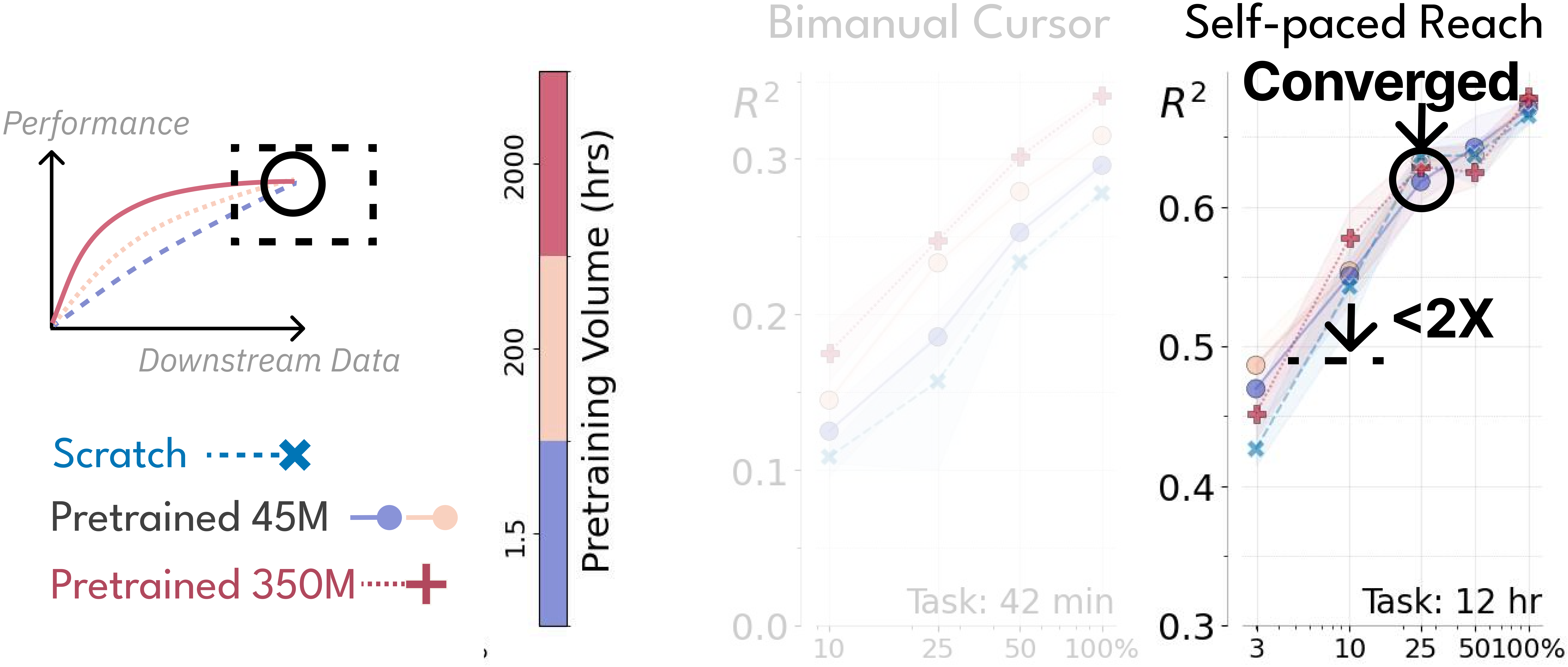
Quantifying downstream scaling gains



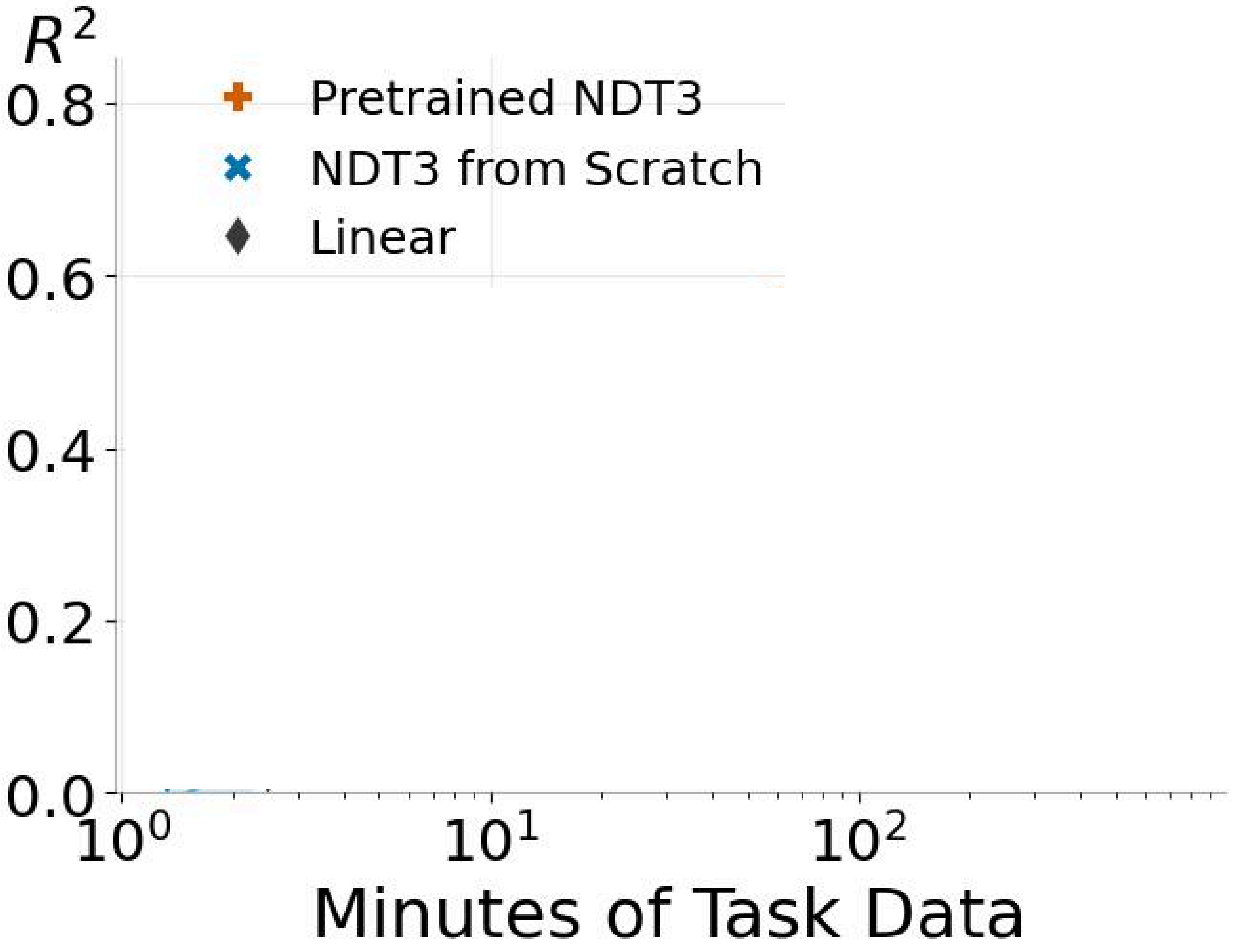
Quantifying downstream scaling gains across tasks



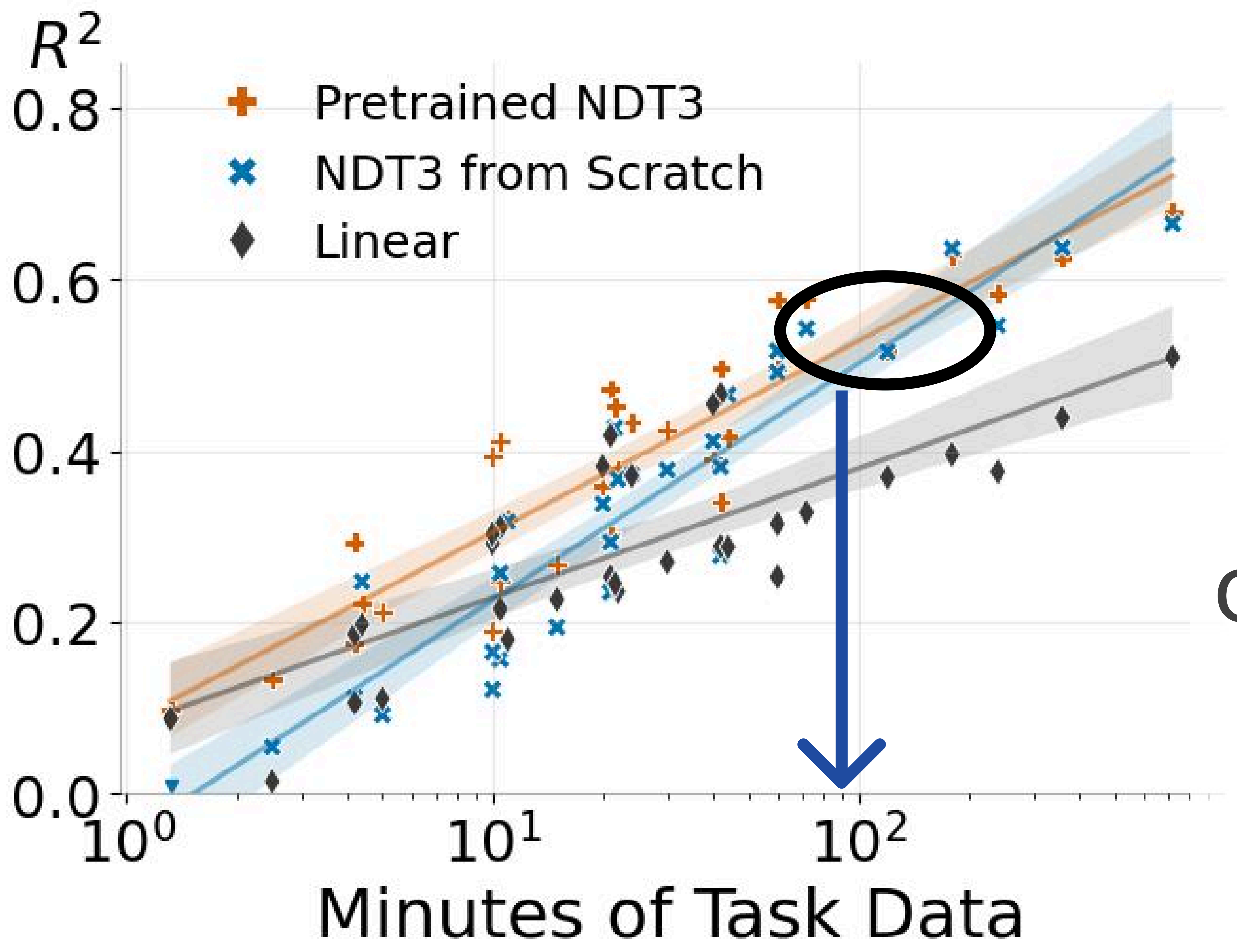
Quantifying downstream scaling gains across tasks



Quantifying downstream scaling gains on 8 datasets

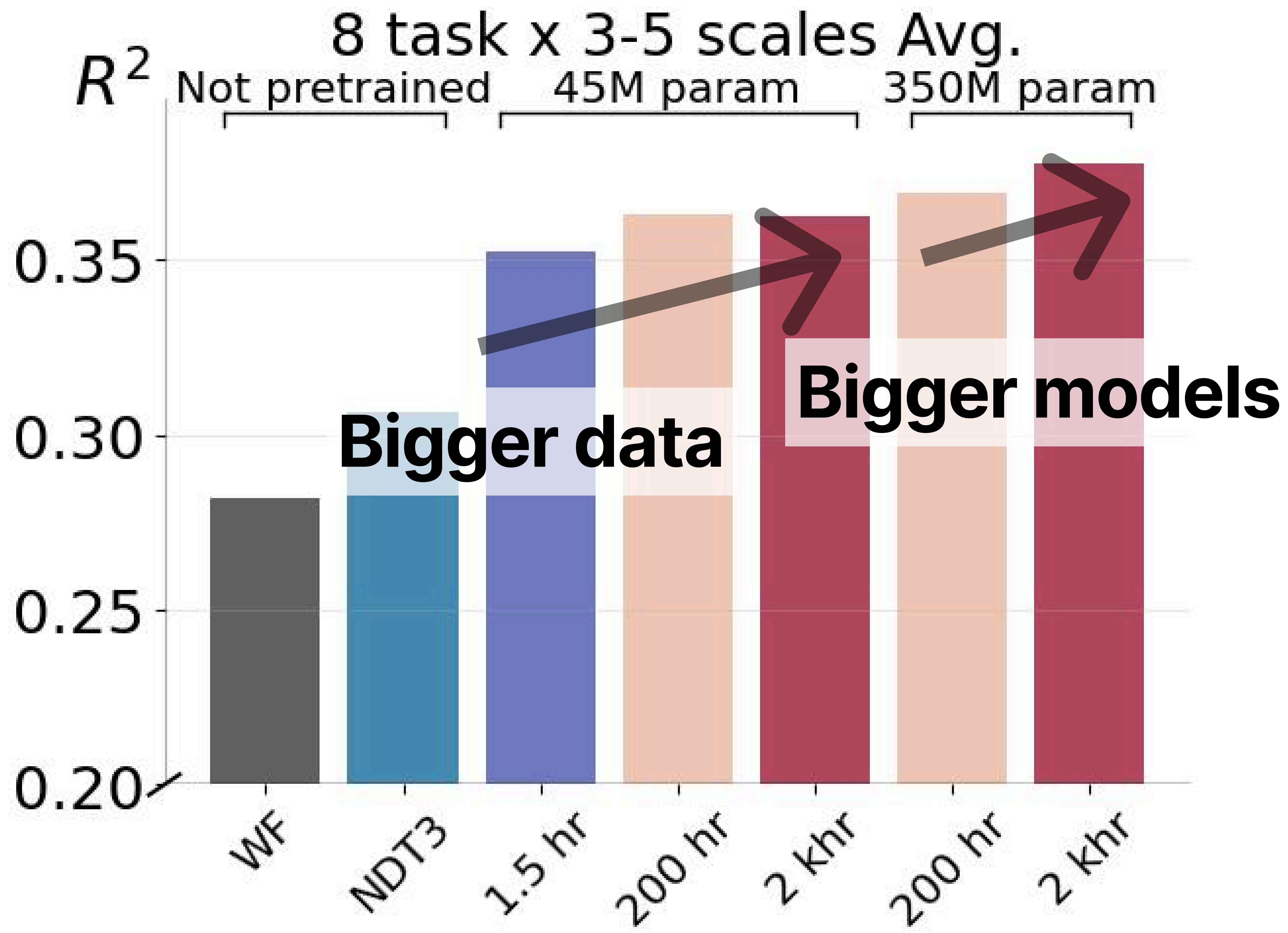


Quantifying downstream scaling gains on 8 datasets

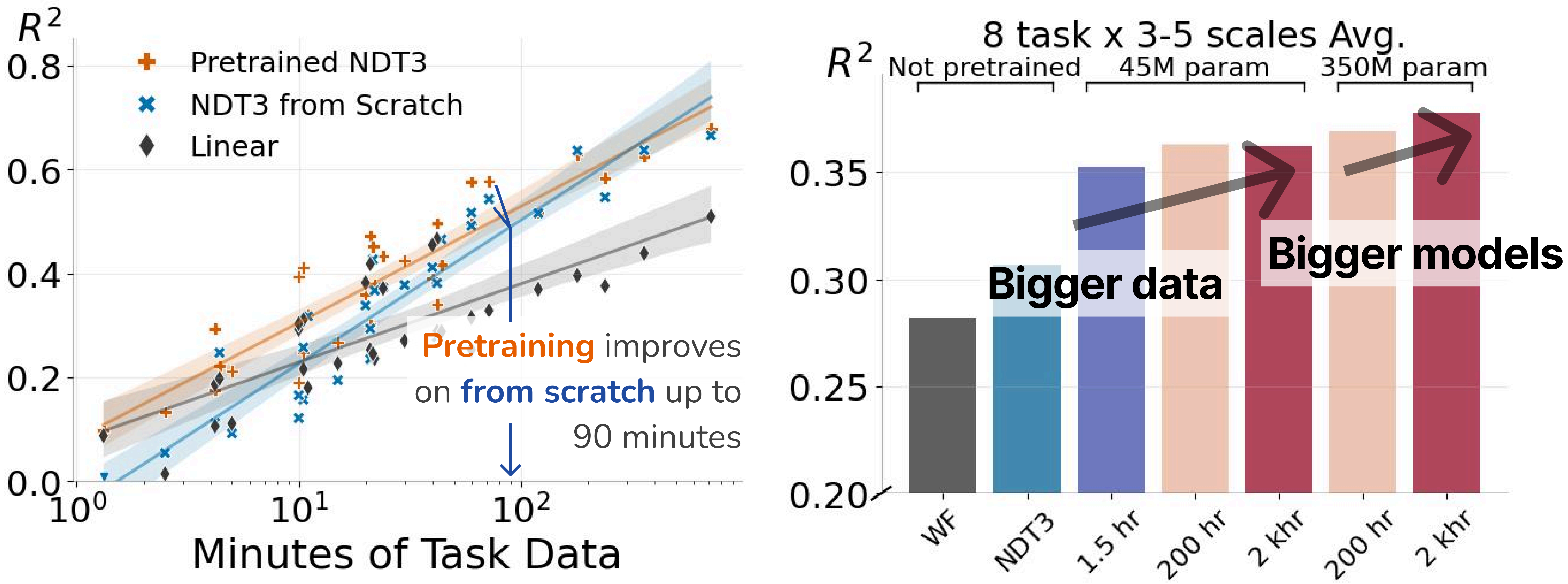


Pretraining improves on **from scratch** up to 90 minutes

Quantifying downstream scaling gains on 8 datasets



NDT has modest practical value for BCI, and scaling likely won't change this.



1. Evaluating* for pragmatics (is hard)

2. Evaluating for generalization

Neural foundation models need evaluations to progress.

1. Evaluating for pragmatics

2. Evaluating for generalization

Pretrain → **Tune A** → **Evaluate B**

- Ecological covariate shifts
- Angular attractors
- Cursor control

Neural foundation models need evaluations to progress.

Pretrain → Tune A → Evaluate B

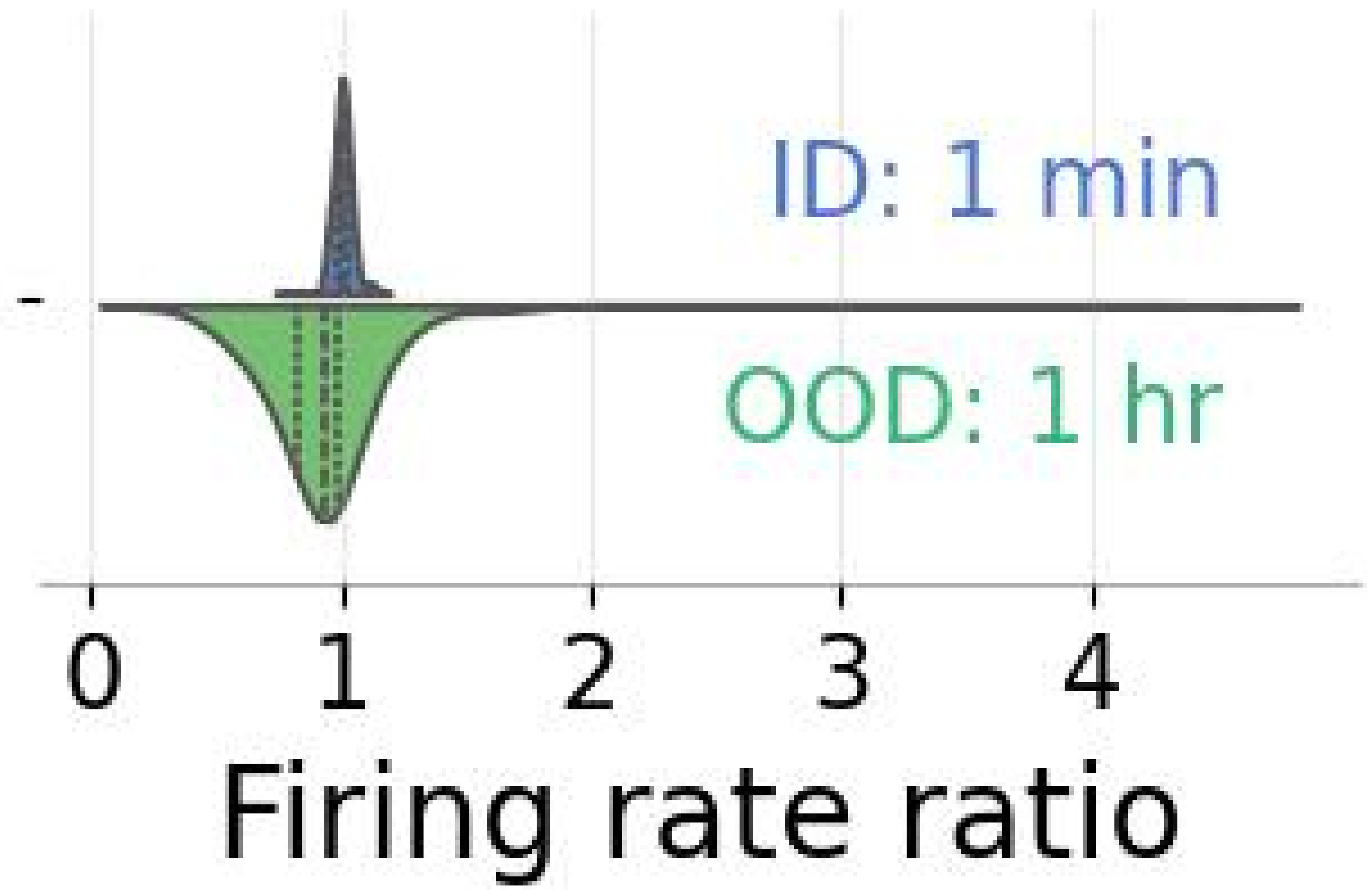
Ecological covariate shifts

Shift: **Time**

Pretrain → Tune A → Evaluate B

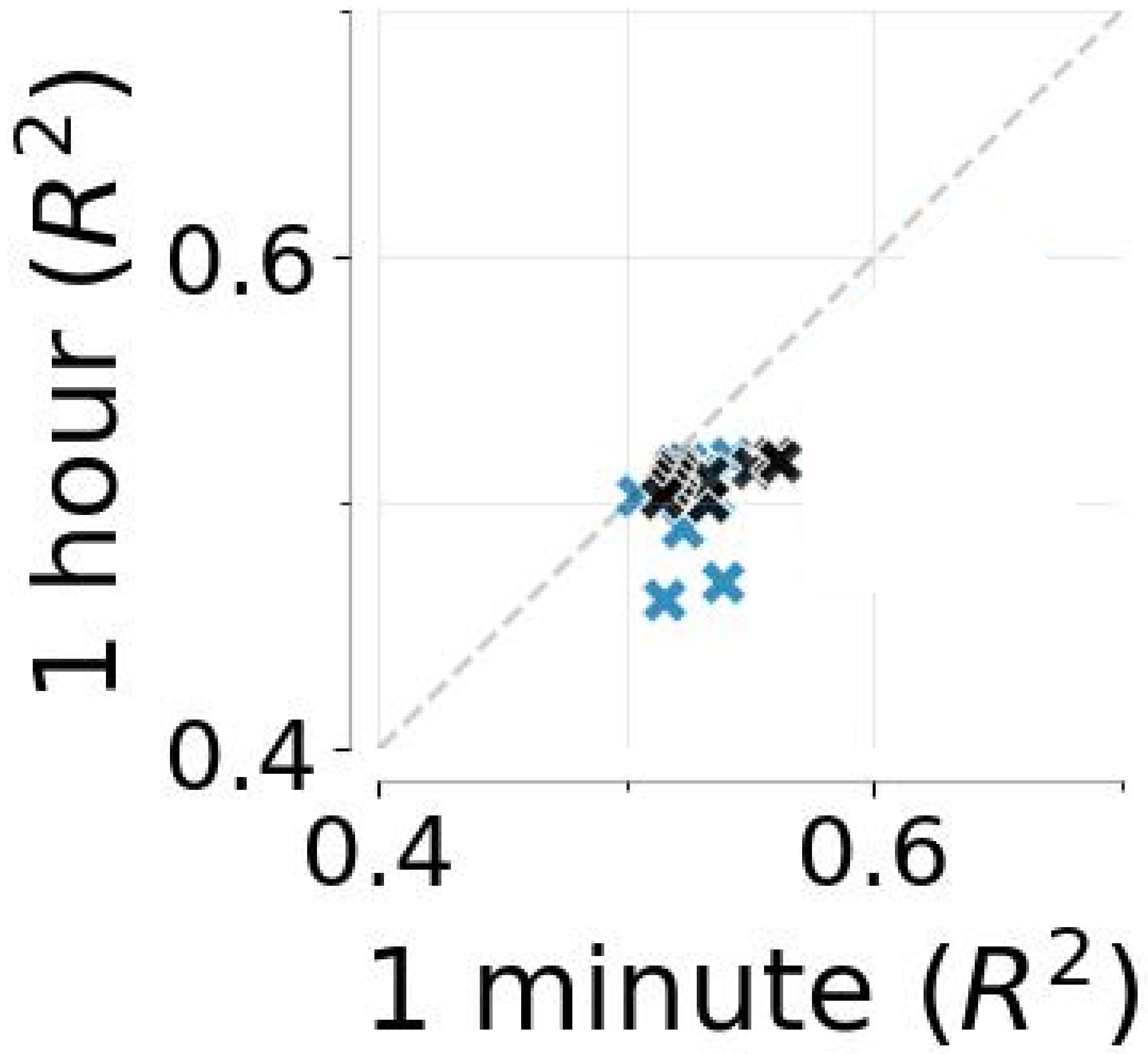
Ecological covariate shifts

Shift: **Time**



Pretrain → Tune A → Evaluate B

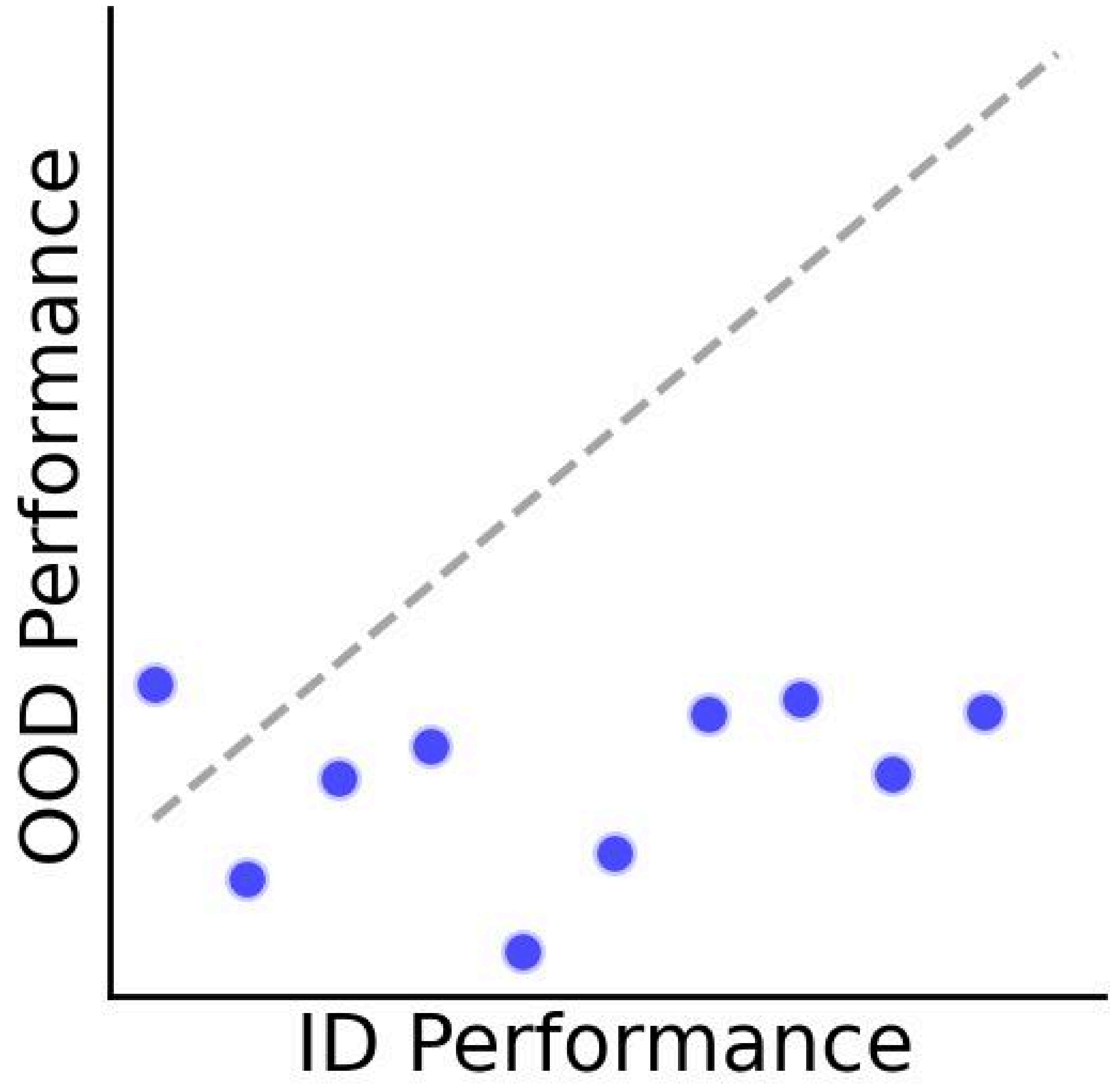
Ecological covariate shifts



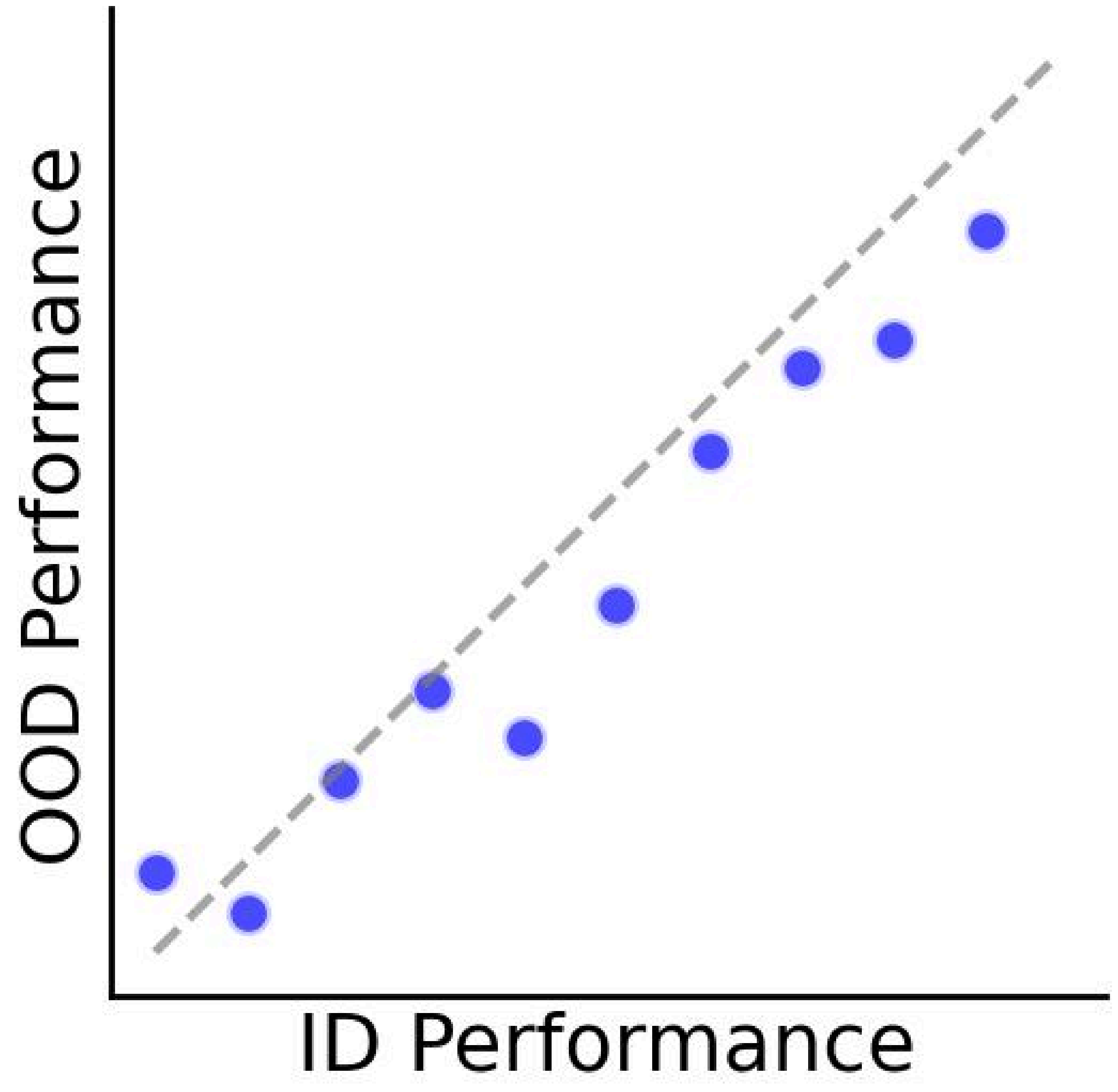
✖ Wiener Filter ✖ Scratch ● 200hr 45M ✚ 2khr 350M

Pretrain → Tune A → Evaluate B

Ecological covariate shifts



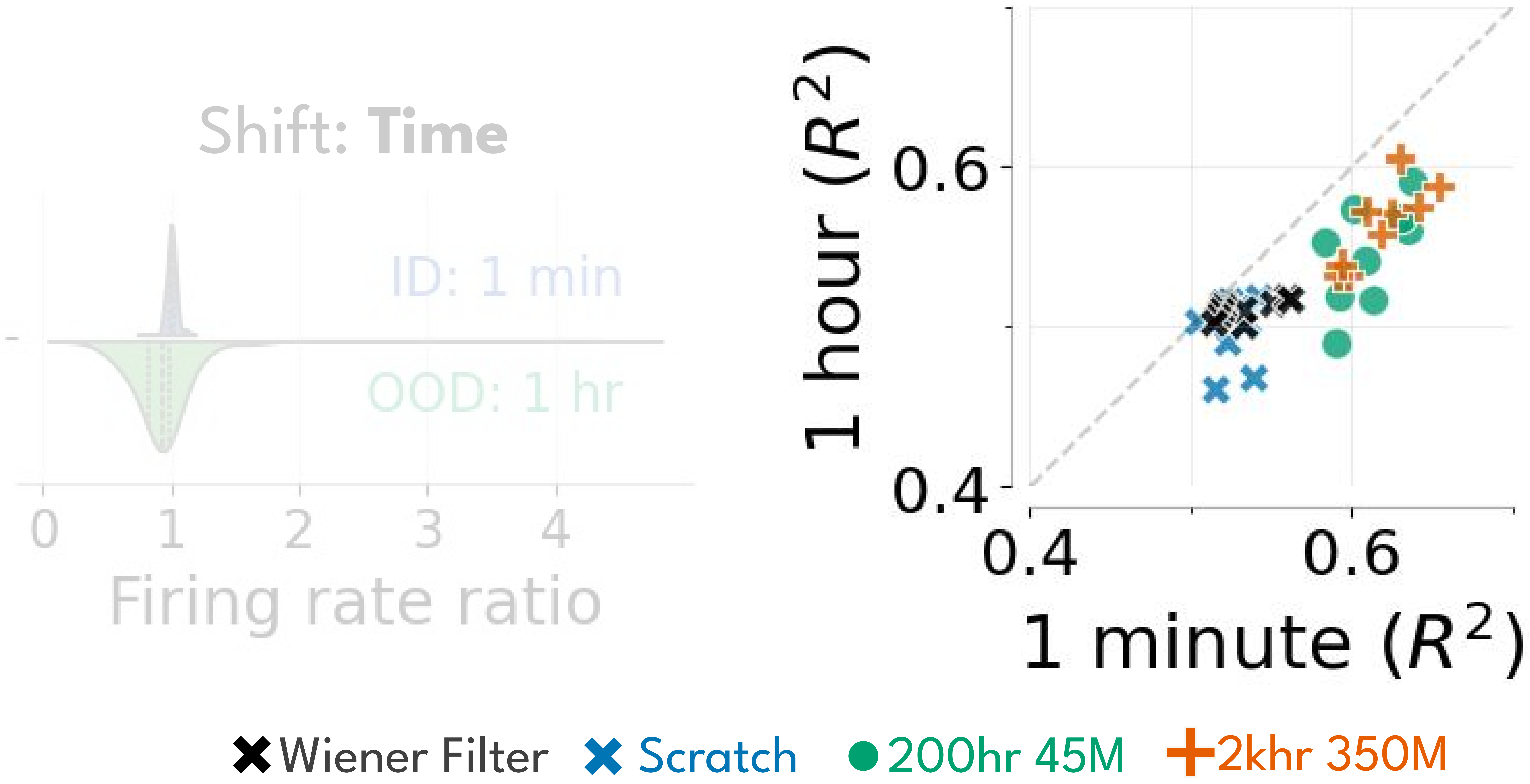
Hypothesis 1: Fragile features



Hypothesis 2: Robust gains

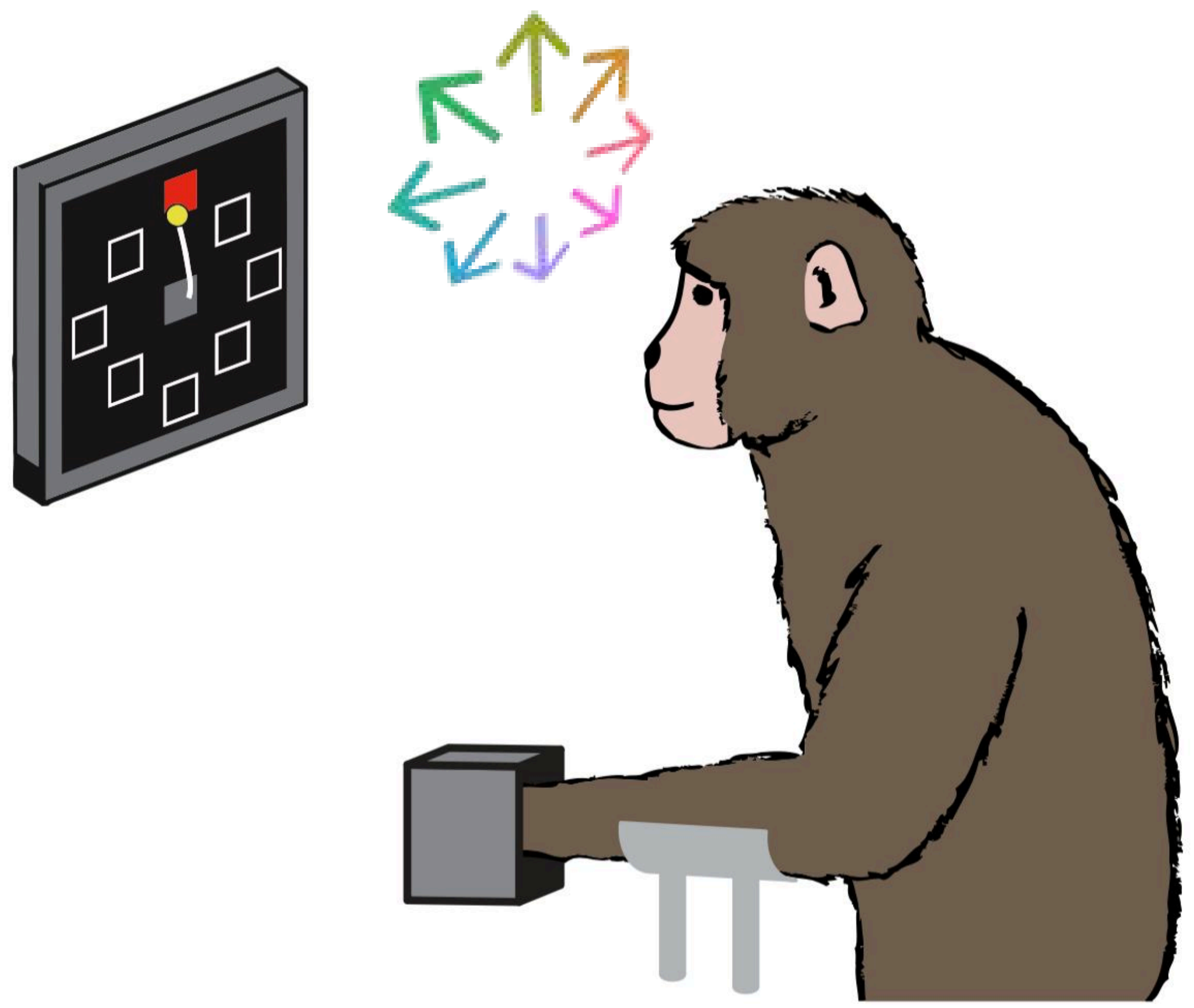
Pretrain → Tune A → Evaluate B

Ecological covariate shifts



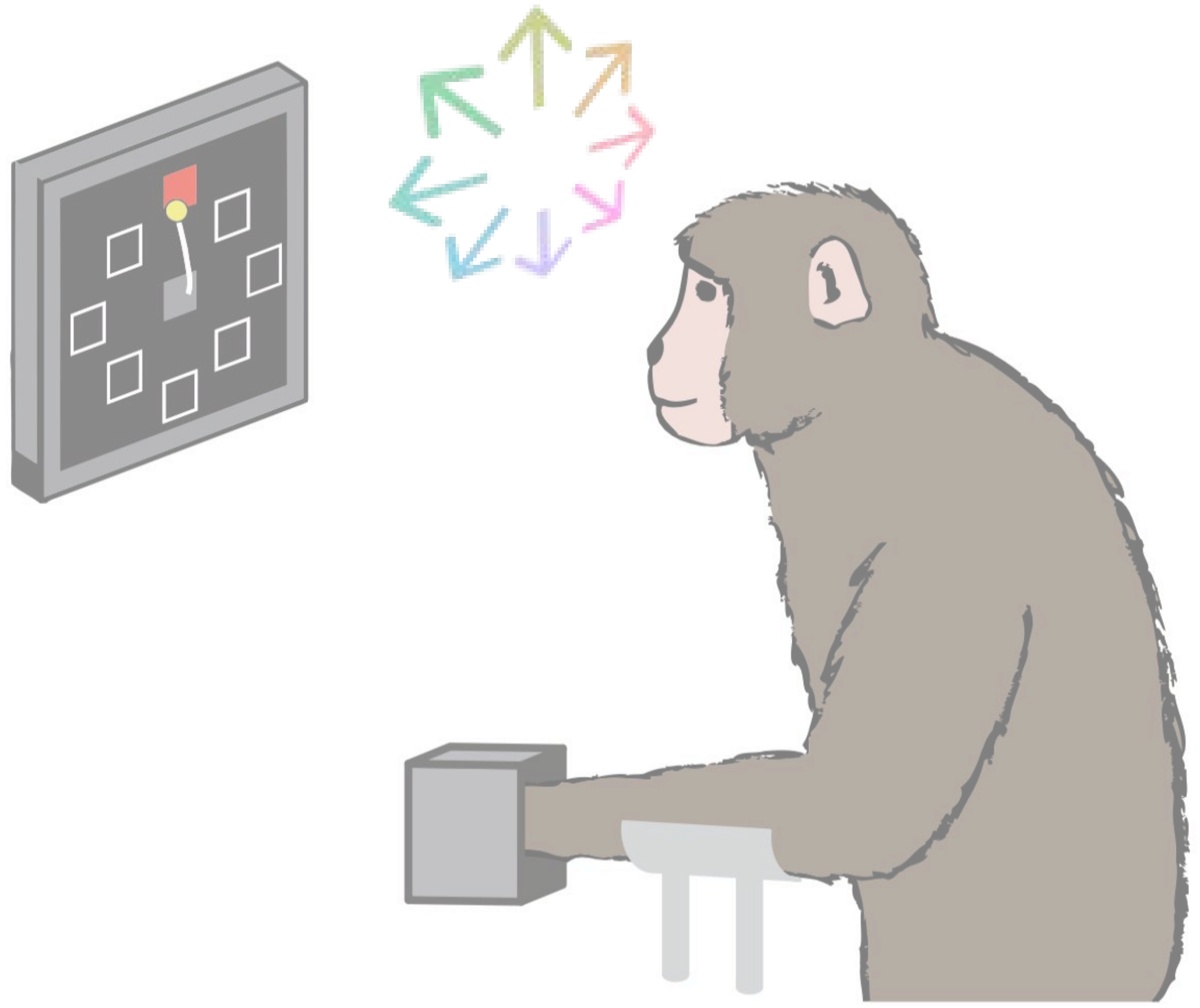
Pretrain → Tune A → Evaluate B

Angular attractors: Probing for "qualitative priors"

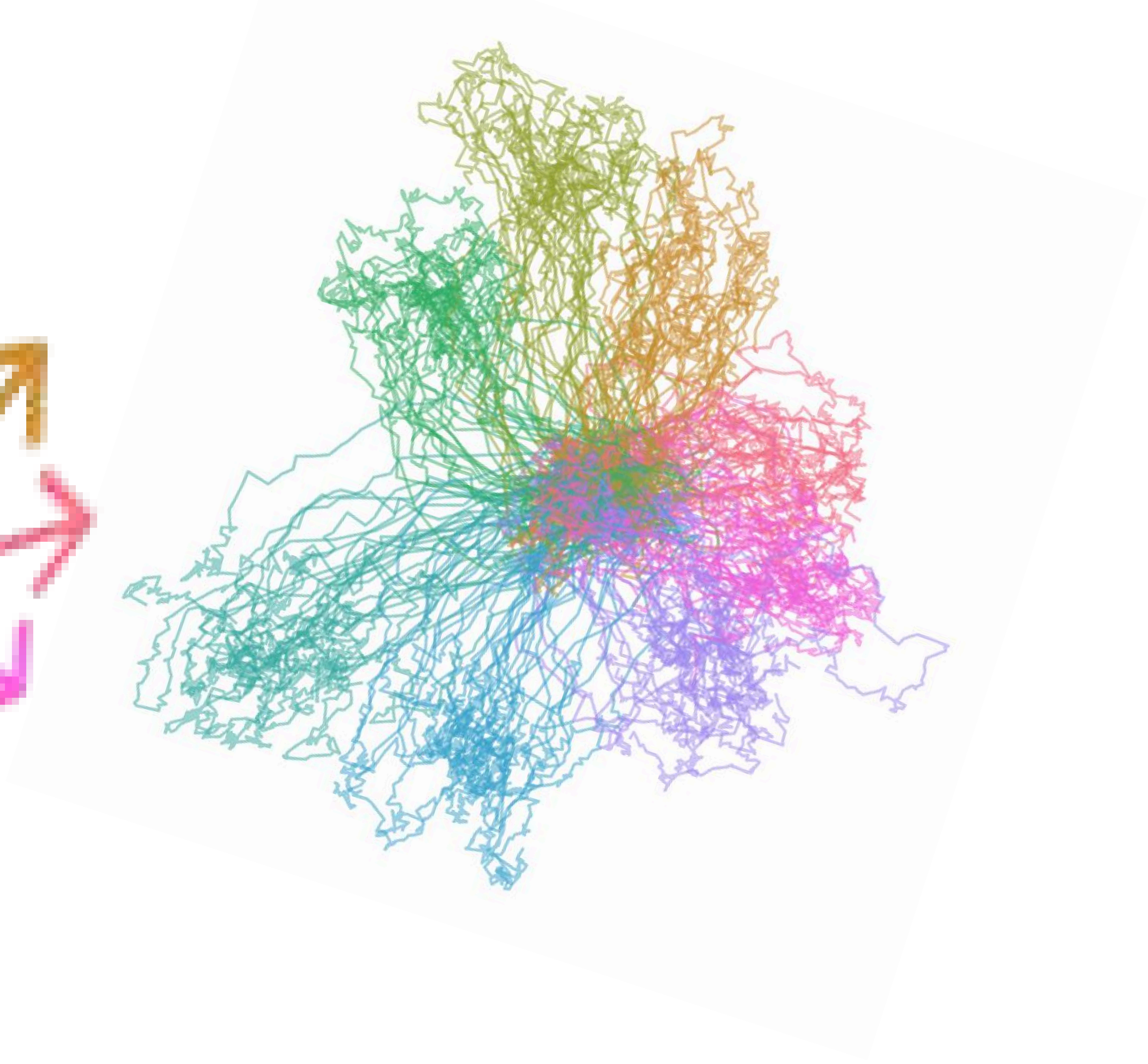


Pretrain → Tune A → Evaluate B

Angular attractors

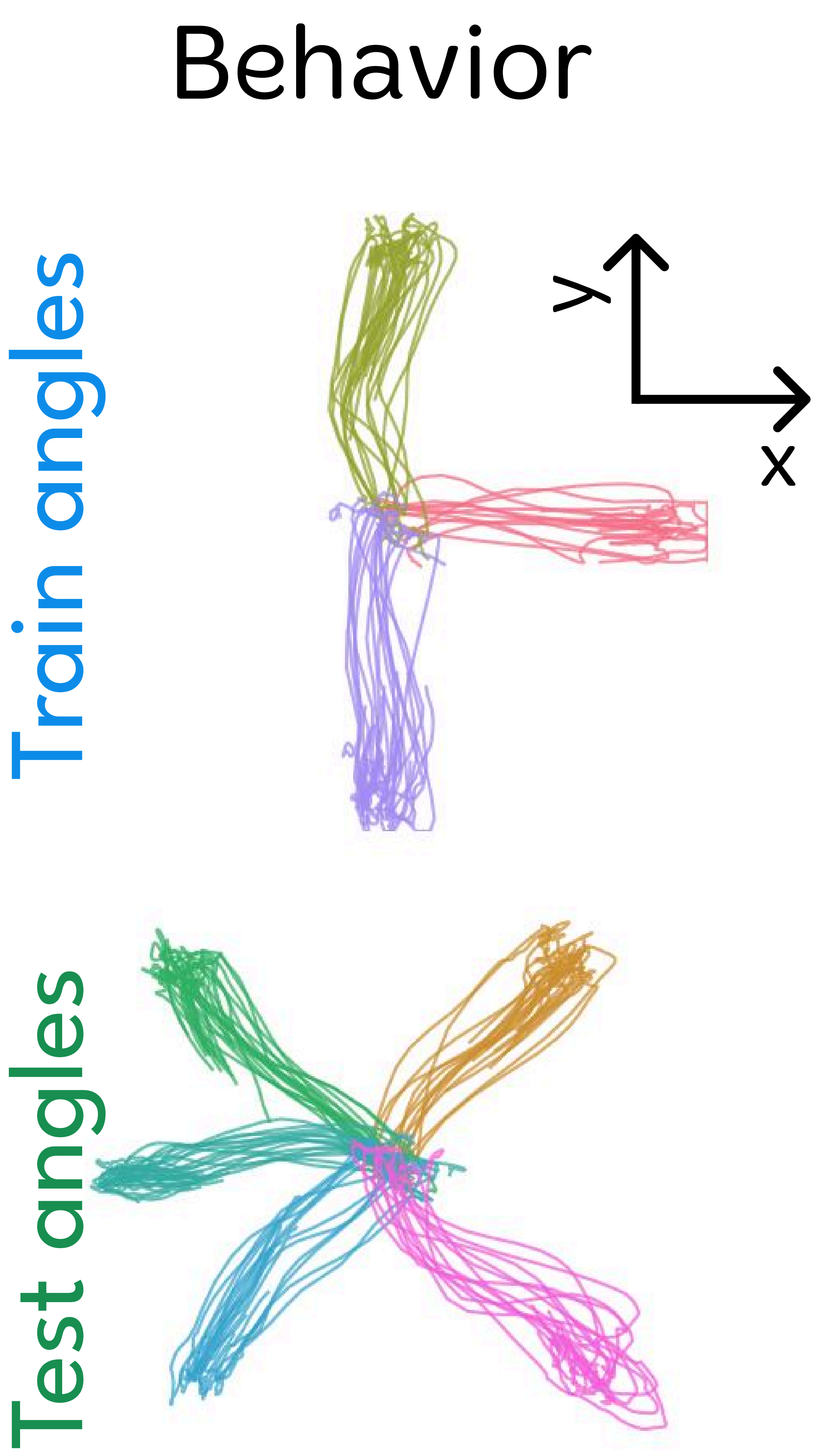


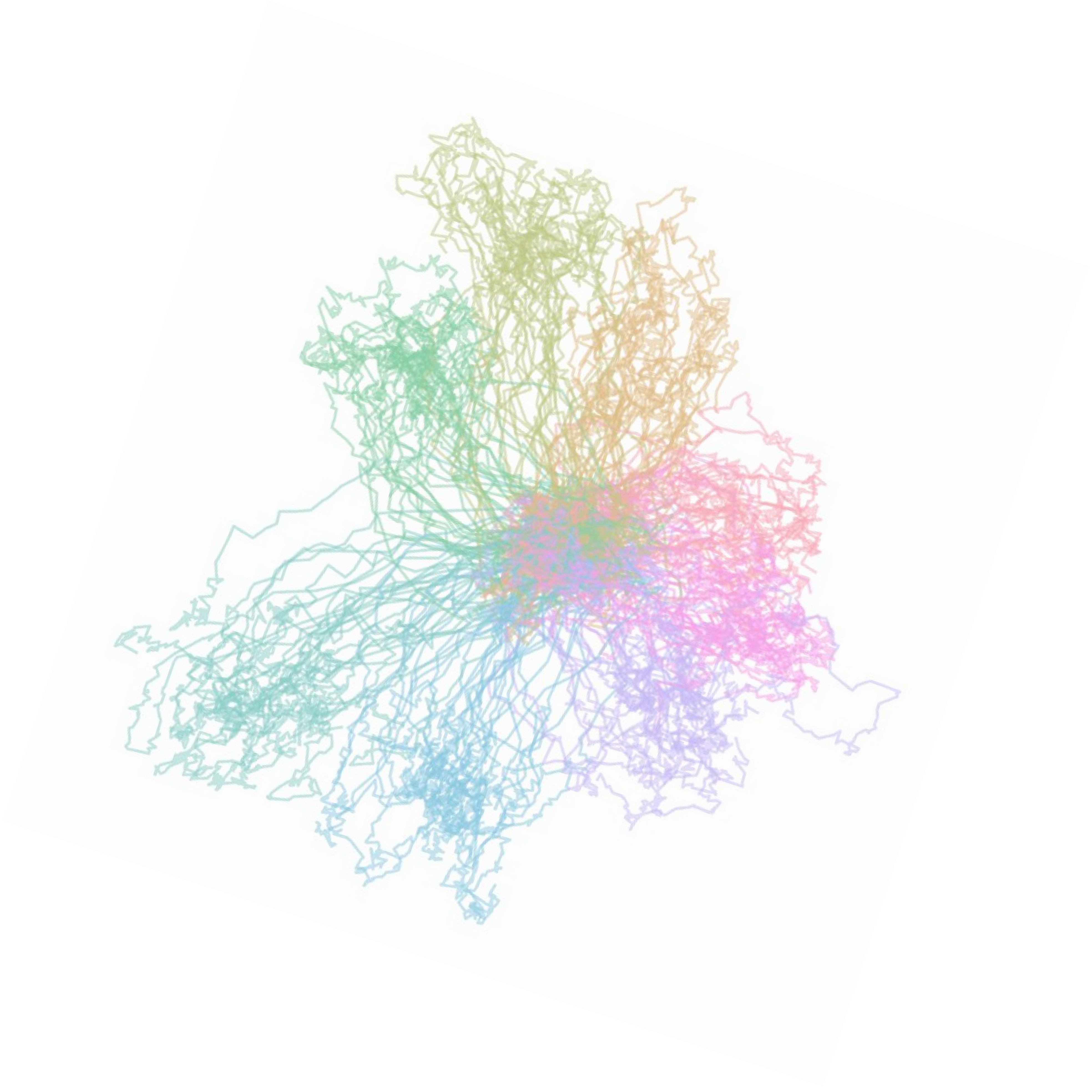
PCA + LDA Trajectories

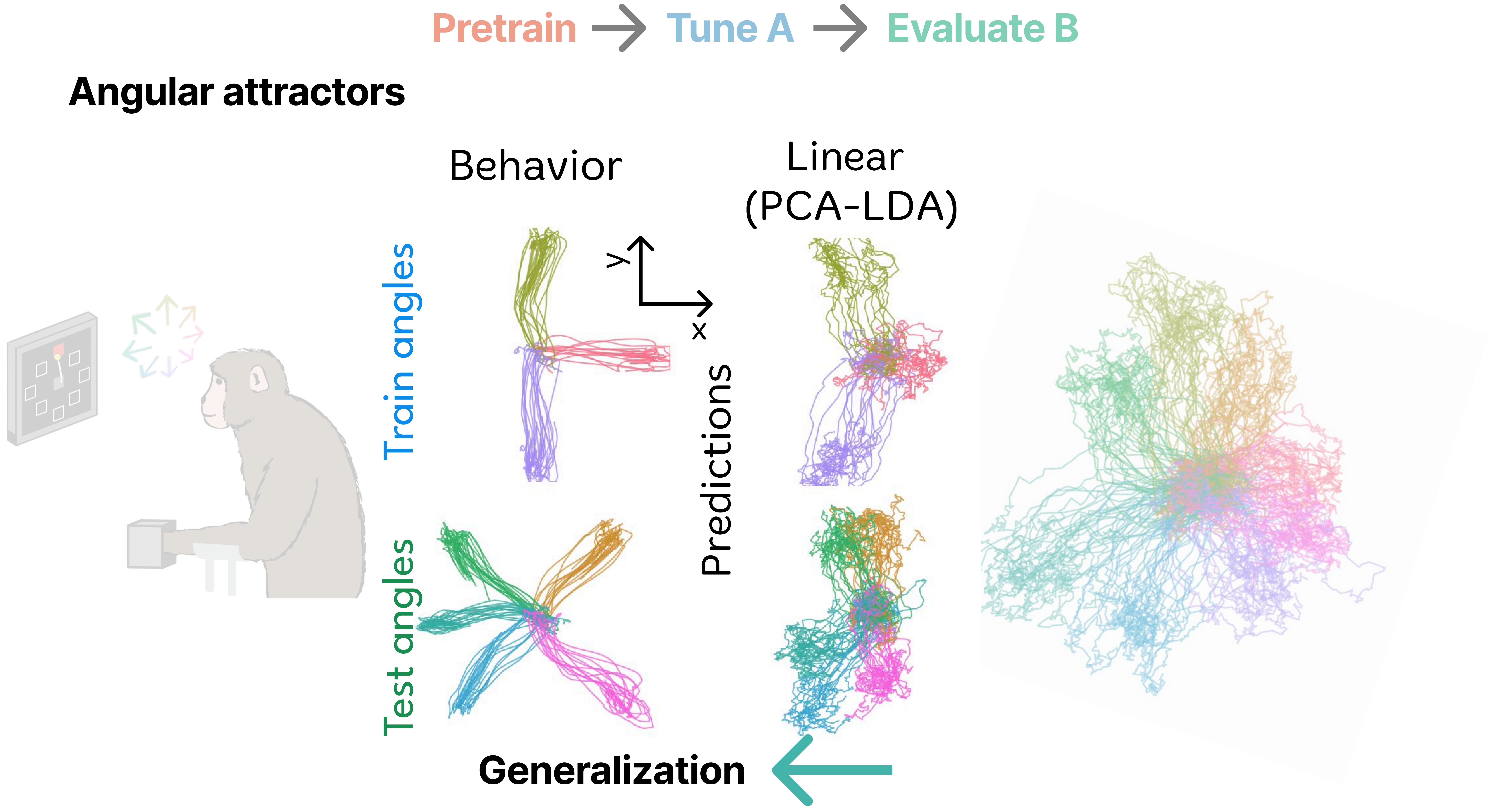


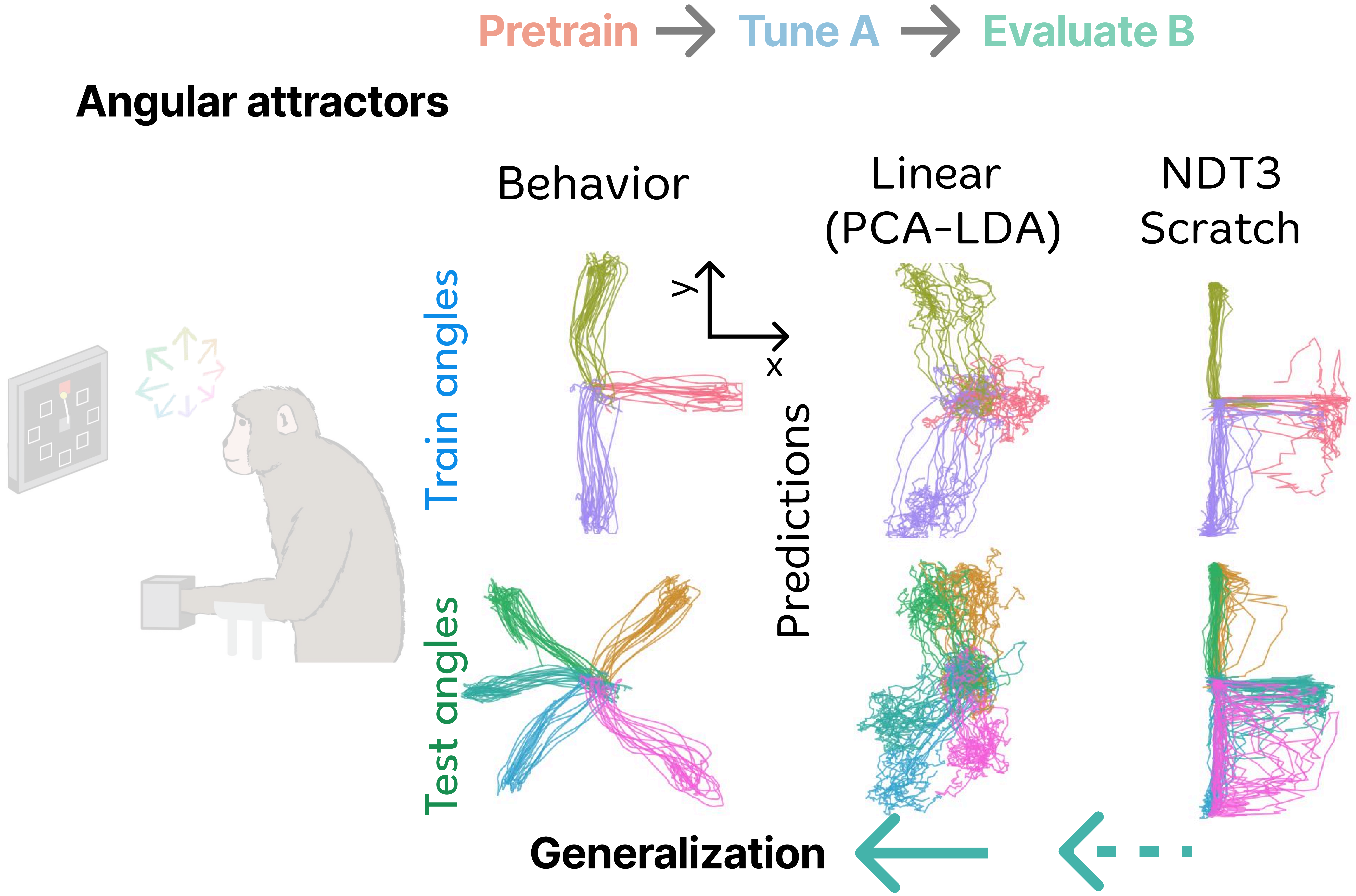
Pretrain → Tune A → Evaluate B

Angular attractors



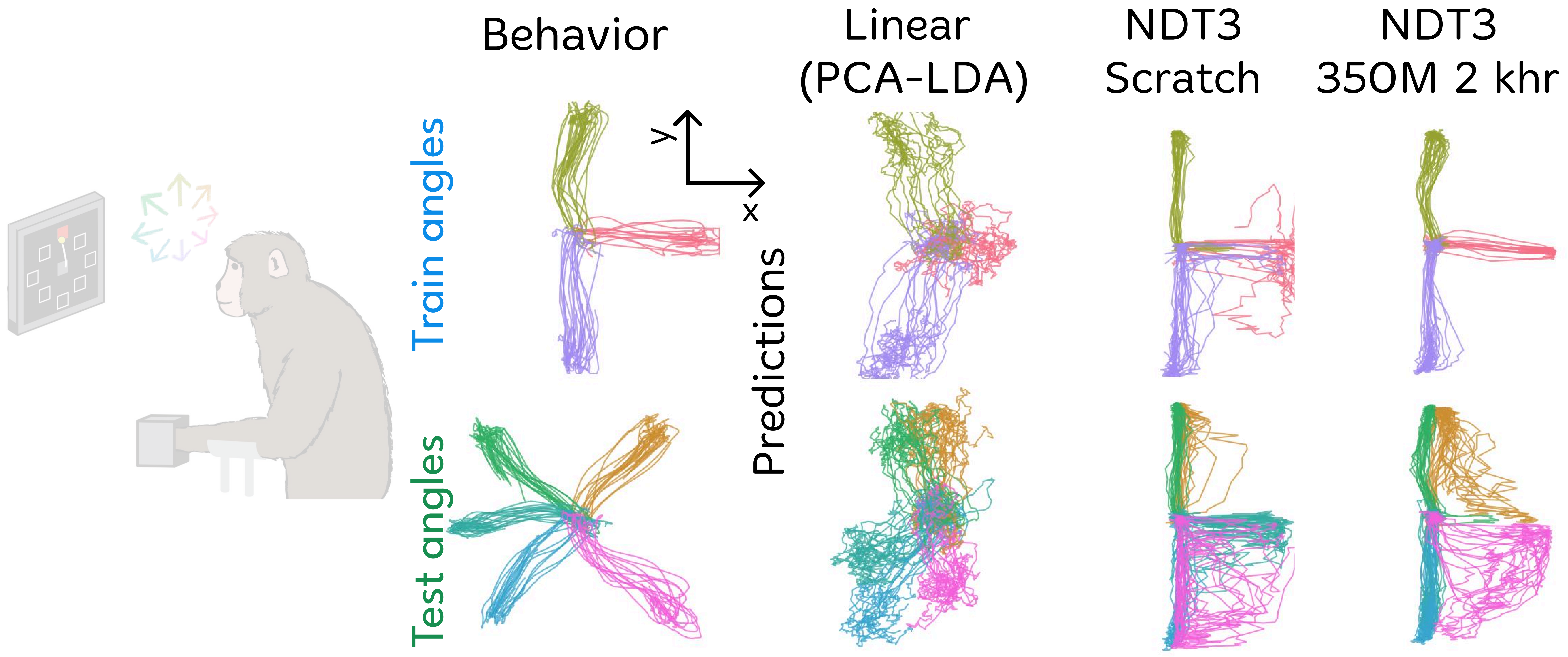



Pretrain → Tune A → Evaluate B
Angular attractors
Behavior
Linear
(PCA-LDA)
Train angles
Test angles
y
x
Predictions
Generalization


Pretrain → Tune A → Evaluate B
Angular attractors
Behavior
Linear (PCA-LDA)
NDT3 Scratch
Train angles
Test angles
y
x
Predictions
Generalization

Pretrain → Tune A → Evaluate B

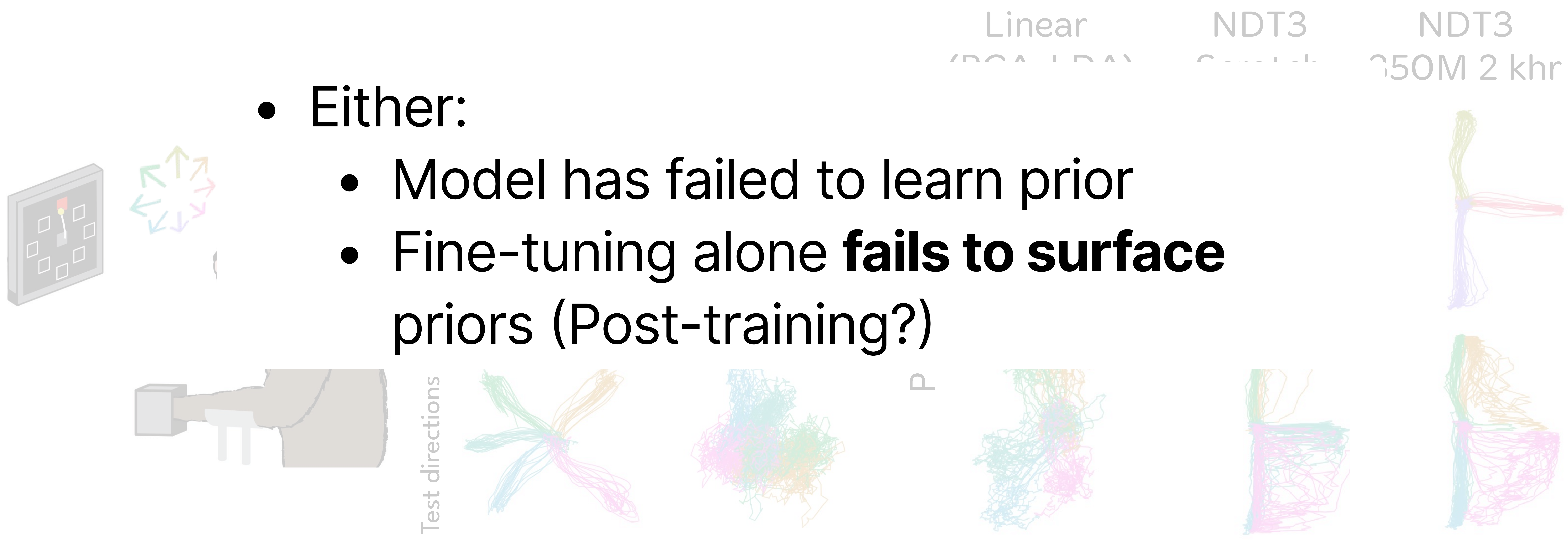
Angular attractors



Pretrain → Tune A → Evaluate B

Angular attractors

- Either:
  - Model has failed to learn prior
  - Fine-tuning alone **fails to surface** priors (Post-training?)



Pretrain → Tune A → Evaluate B

Cursor control

Tune A

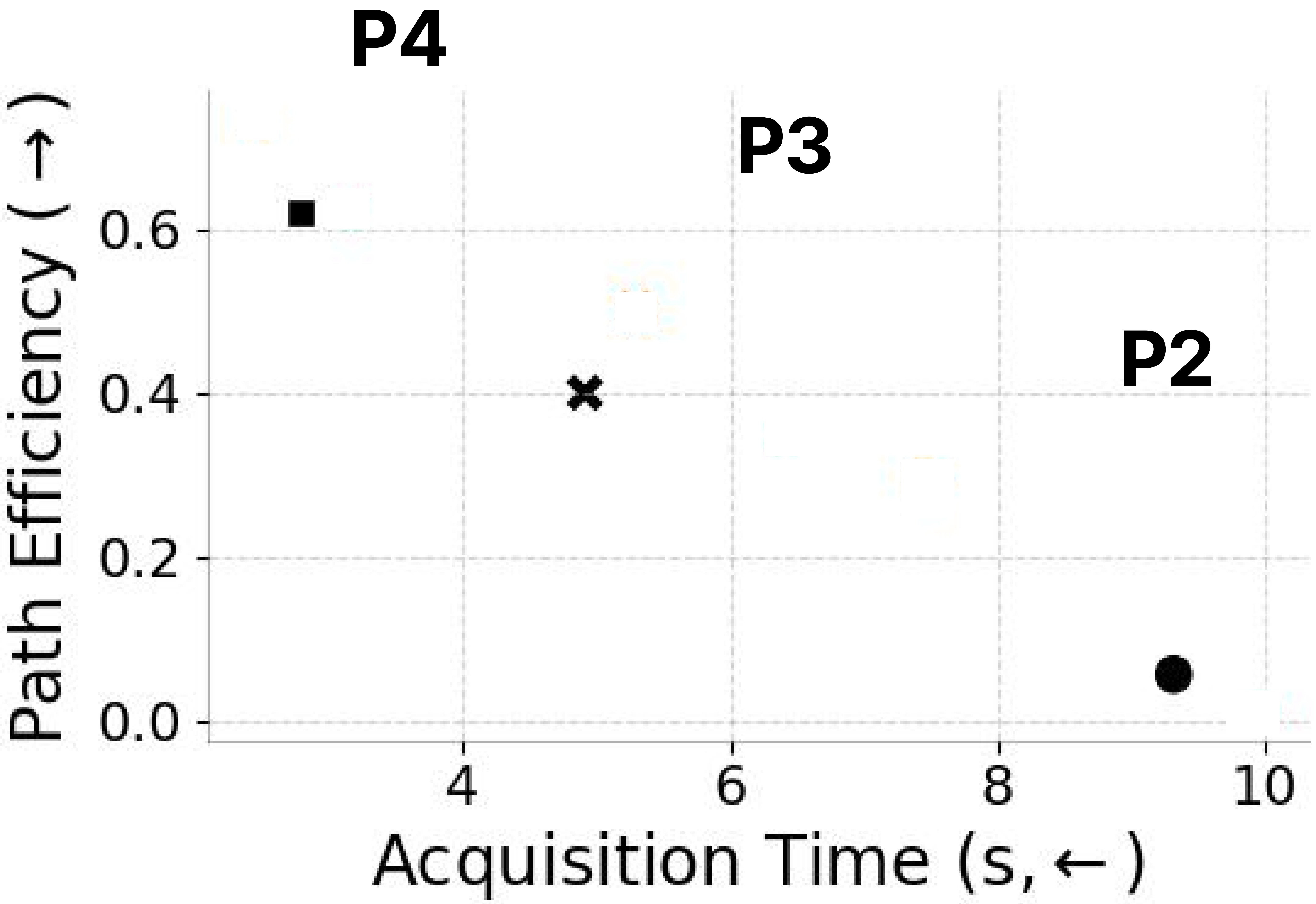
Open loop calibration
(~1 minute)

Evaluate B (Control)

Pretrain → Tune A → Evaluate B

Cursor control

Linear (iOLE)



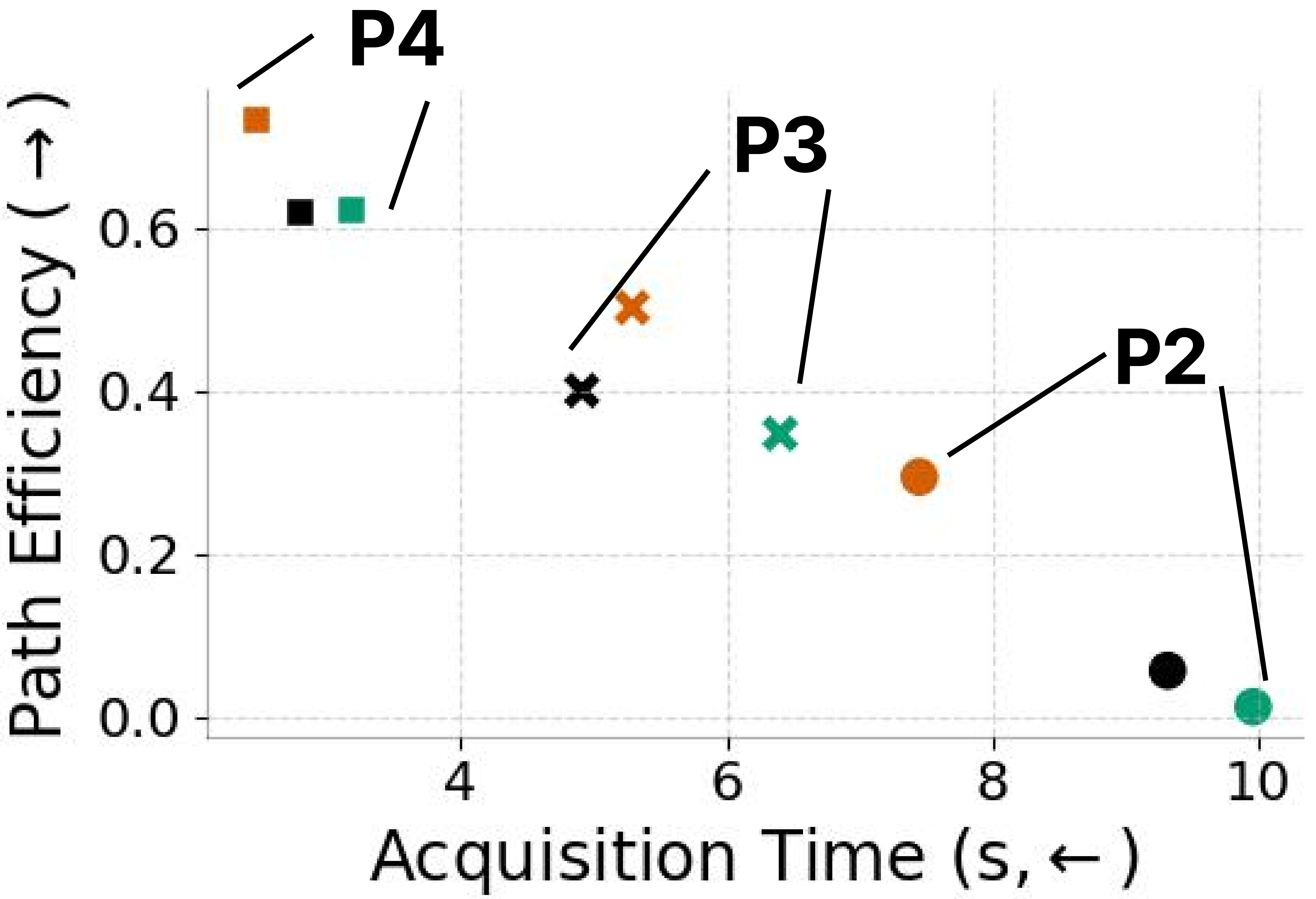
Pretrain → Tune A → Evaluate B

Cursor control

Linear (iOLE)

NDT Base (200hr 45M)

NDT Big (2khr 350M)



Evaluation will drive progress in neural foundation models.

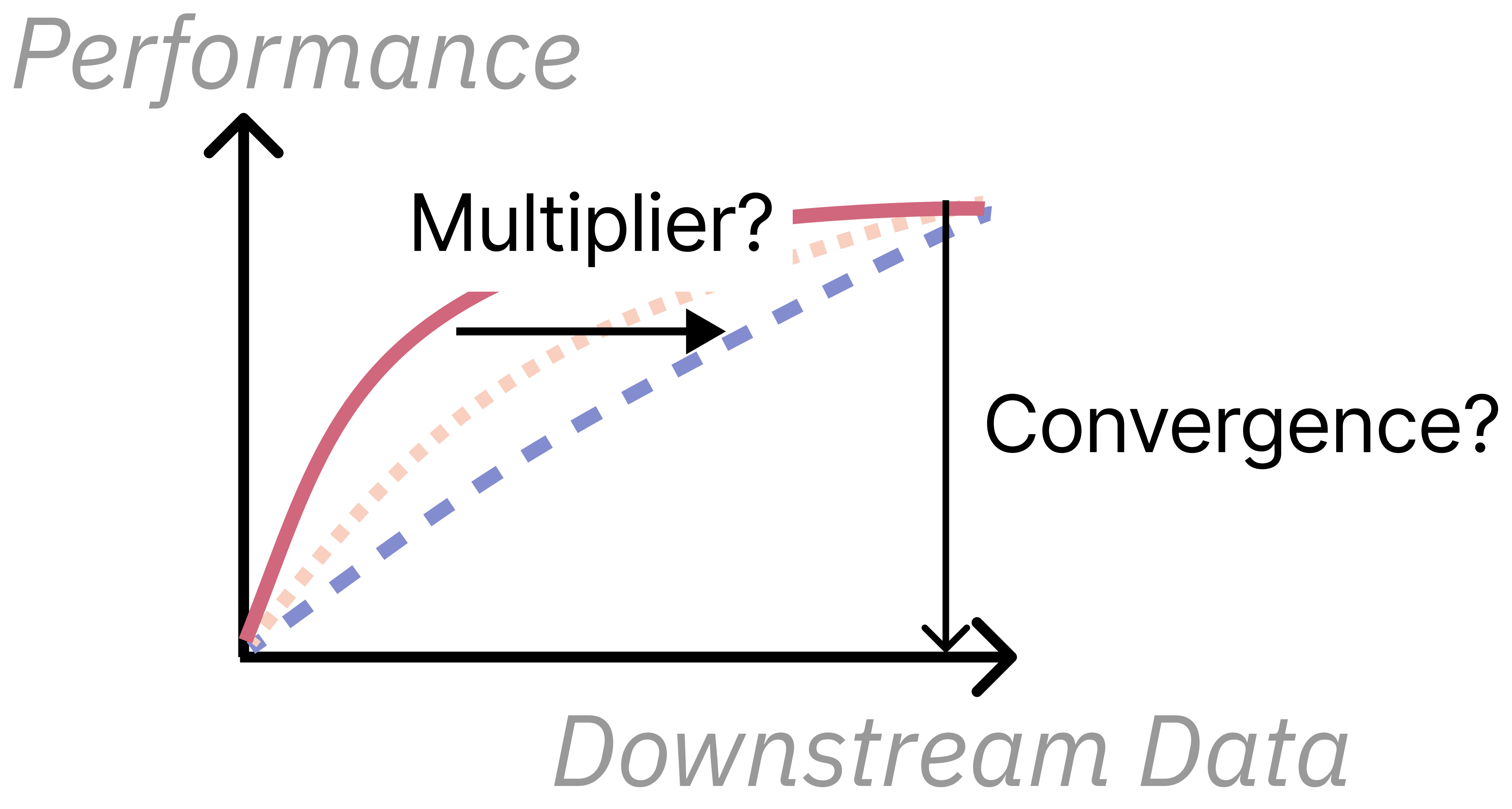
Pragmatics

Generalization

Evaluation will drive progress in neural foundation models.

Pragmatics

- Pretraining gain should be measured over realistic ranges

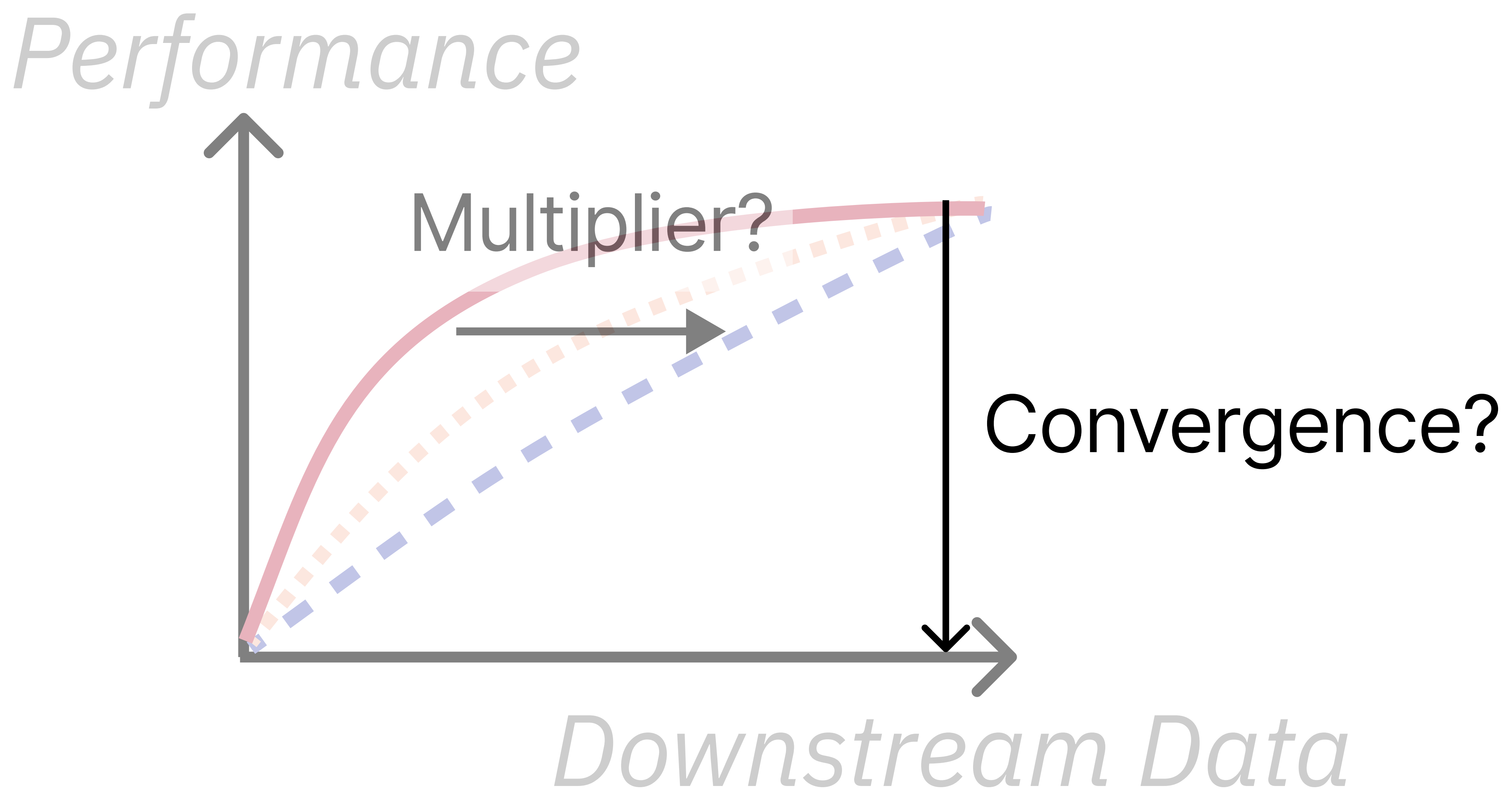


Generalization

Evaluation will drive progress in neural foundation models.

Pragmatics

- Pretraining gain should be measured over realistic ranges



Generalization

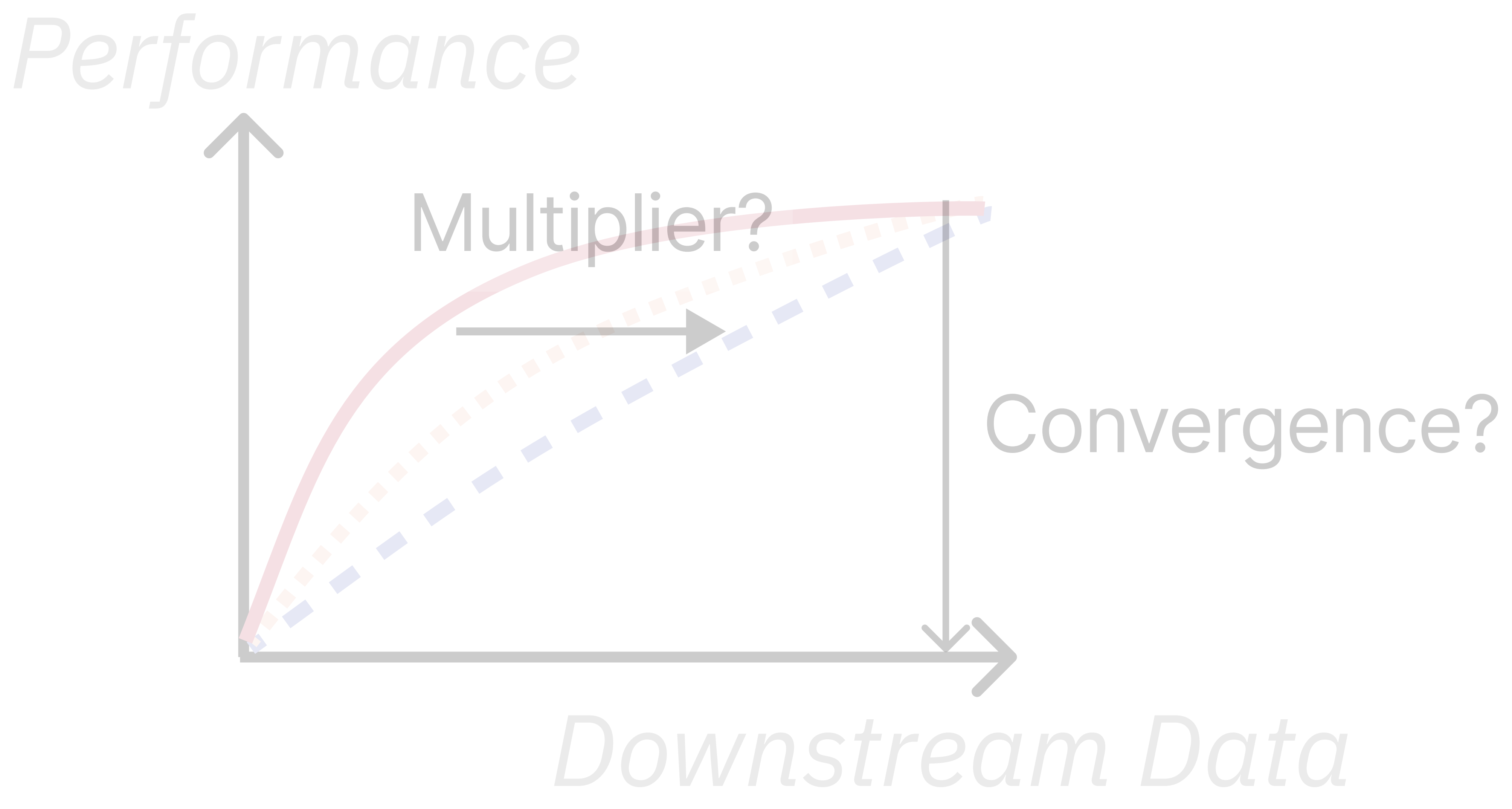
- Downstream often needs generalization
- Targeted probes can surface challenges that scaling will not resolve.

Pretrain → Tune A → Evaluate B

Pragmatics

Generalization

Evaluations define progress in neural foundation models.



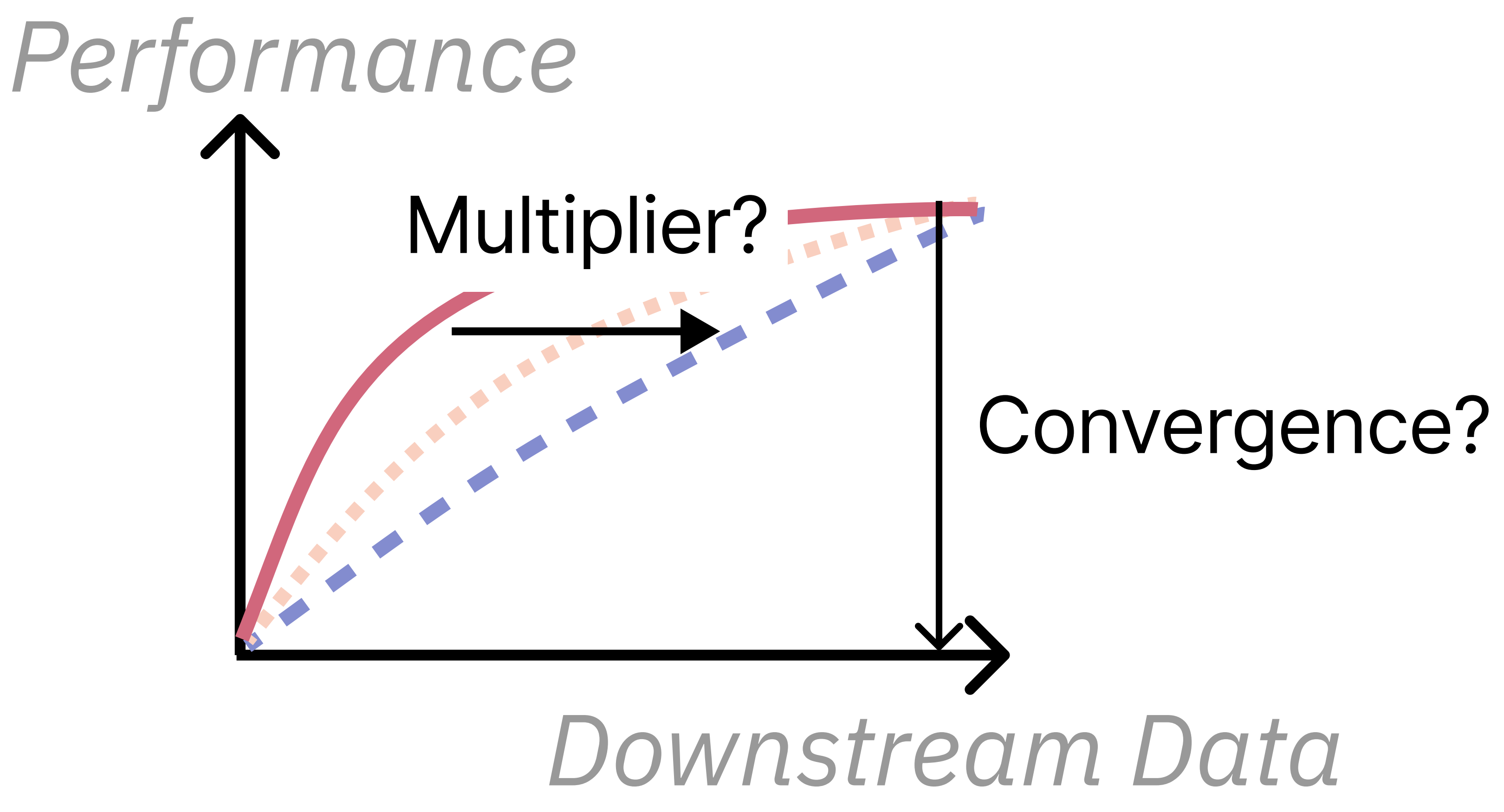
resolve.

Pretrain → Tune A → Evaluate B

Evaluation will drive progress in neural foundation models.

Pragmatics

- Pretraining gain should be measured over realistic ranges



Generalization

- Downstream will need generalization
- Targeted probes can surface challenges that scaling will not resolve.

Pretrain → Tune A → Evaluate B

Support:
DOE CSGF
DE-SC0023112
NIH National Institute of Neurological Disorders and Stroke
R01NS121079, UH3NS107714
PittCRC Center for Research Computing
NERSC

Carnegie Mellon University
RNEL Rehab Neural Engineering Labs
University of Pittsburgh
Cortical Bionics research group
The University of Kansas
The University of Chicago
Northwestern University