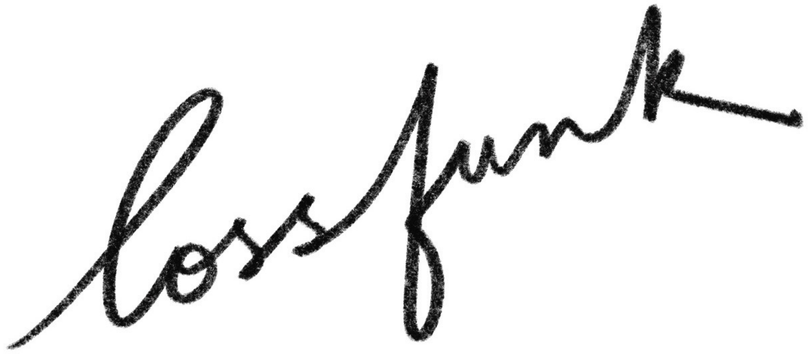
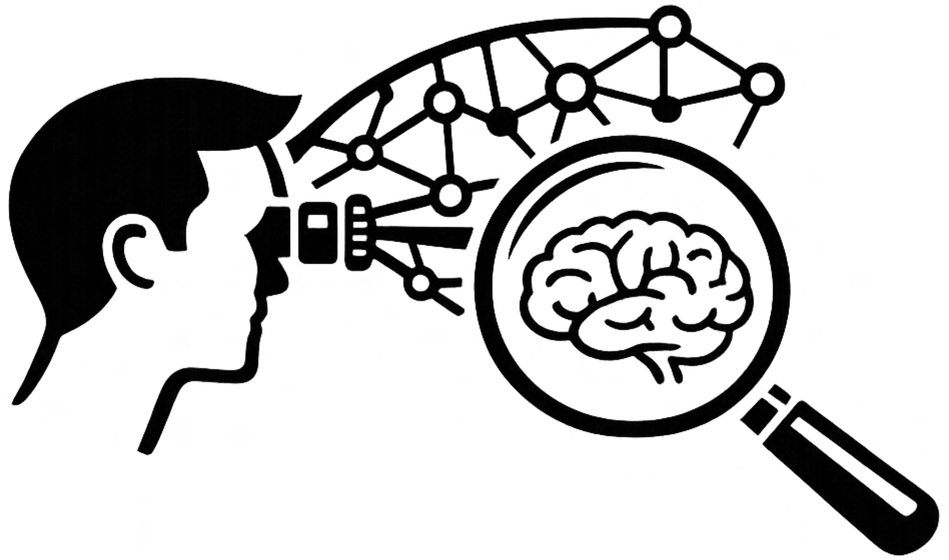




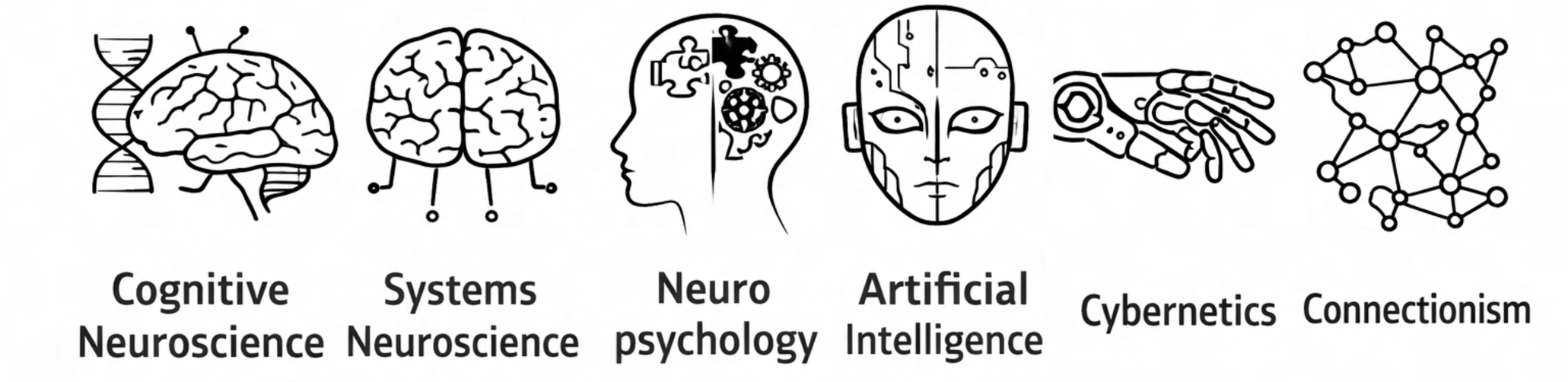
Neural Nets as Scientific Instruments

Towards unraveling human shape bias and representational drift & what's next…

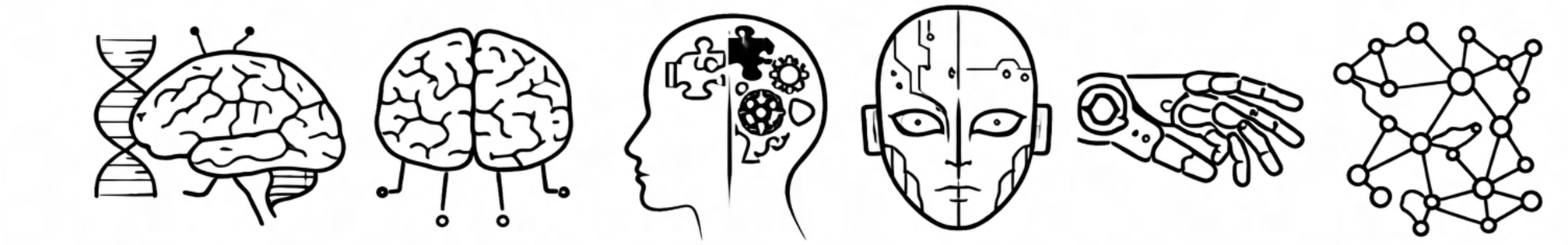
Sushrut Thorat *(sushrutthorat.com)*

***Calibre Many Worlds Seminar @ ICTS*, 26 October, 2026**

Studying the machinery underlying Human Behavior



Studying the machinery underlying Human Behavior



A primary goal:

Process-level descriptions of cognition

Desiderata

Promising theories, and associated models need to:

Be expressive

Expose *relevant* levers

Rosenbleuth & Weiner, 1945; Milner et al. 1968; McClelland et al. 1995

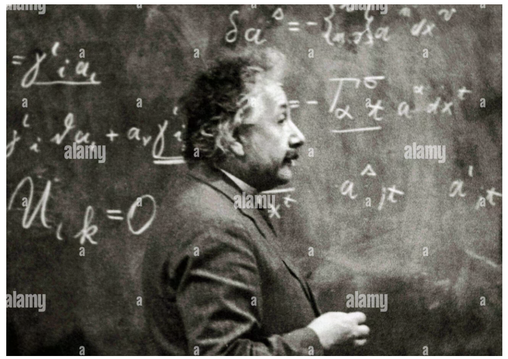
Desiderata

Promising theories, and associated models need to:

Be expressive

Perform a plethora of cognitive tasks



Expose *relevant* levers

Rosenbleuth & Weiner, 1945; Milner et al. 1968; McClelland et al. 1995

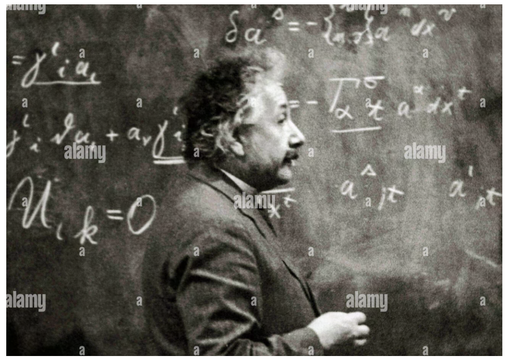
Desiderata

Promising theories, and associated models need to:

Be expressive

Perform a plethora of cognitive tasks



Expose *relevant* levers

Allow interpretable control over behaviour

Interventional

Formative

Rosenbleuth & Weiner, 1945; Milner et al. 1968; McClelland et al. 1995

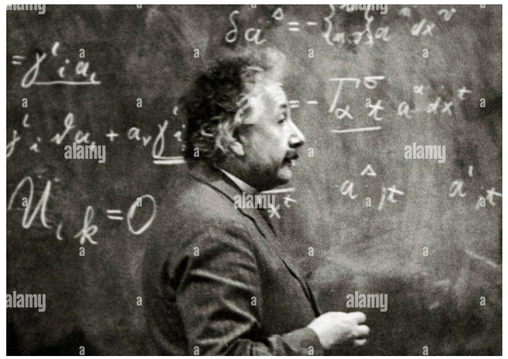
Desiderata

Promising theories, and associated models need to:

Be expressive

Perform a plethora of cognitive tasks



Expose *relevant* levers

Allow interpretable control over behaviour

Interventional

Formative



Hippocampus removed; H.M.

Rosenbleuth & Weiner, 1945; Milner et al. 1968; McClelland et al. 1995

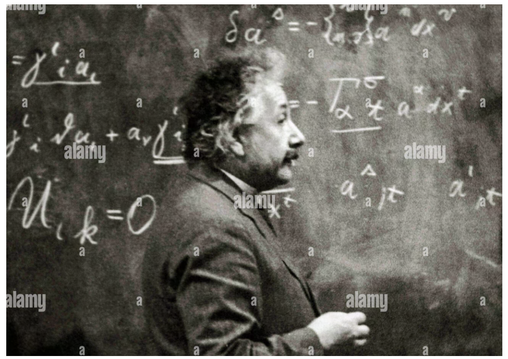
Desiderata

Promising theories, and associated models need to:

Be expressive

Perform a plethora of cognitive tasks



Expose *relevant* levers

Allow interpretable control over behaviour

Interventional

New episodic memories ❌

Skill learning ✅

Formative



Hippocampus removed; H.M.

Rosenbleuth & Weiner, 1945; Milner et al. 1968; McClelland et al. 1995

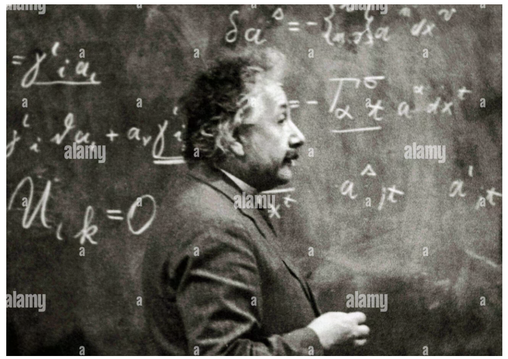
Desiderata

Promising theories, and associated models need to:

Be expressive

Perform a plethora of cognitive tasks



Expose *relevant* levers

Allow interpretable control over behaviour

Interventional

New episodic memories ❌

Skill learning ✅

Formative

Complimentary learning systems (fast vs slow) ← *Stability-plasticity dilemma*



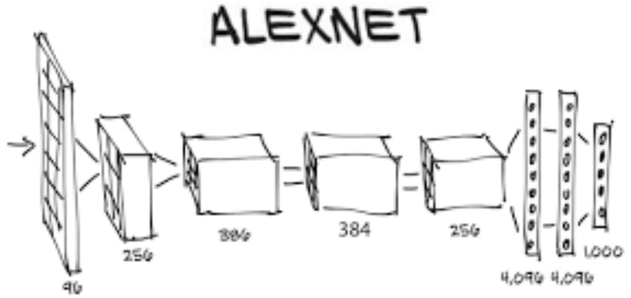
Hippocampus removed; H.M.

emergence analytical accounts

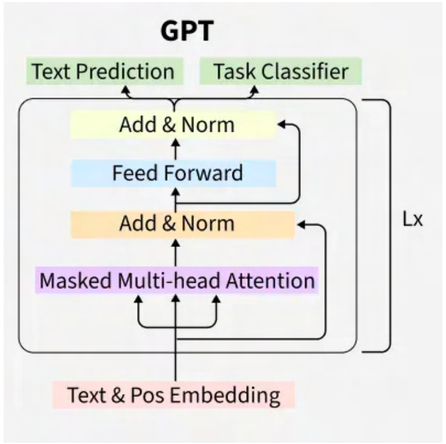
Rosenbleuth & Weiner, 1945; Milner et al. 1968; McClelland et al. 1995

Enter: Artificial Neural Networks

Expressive and performant







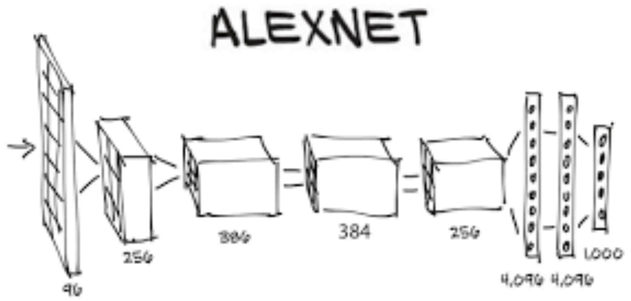


Universal approximation

Leshno et al. 1993; Doerig et al. 2023; Cichy & Kaiser, 2019;

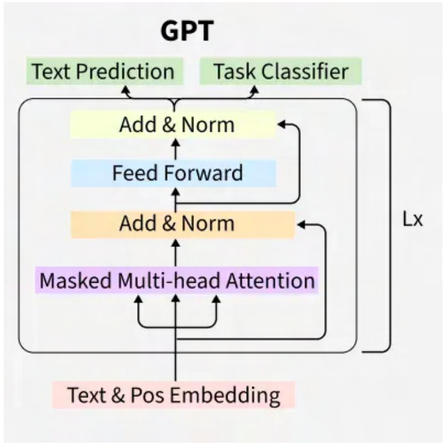
Enter: Artificial Neural Networks

Expressive and performant



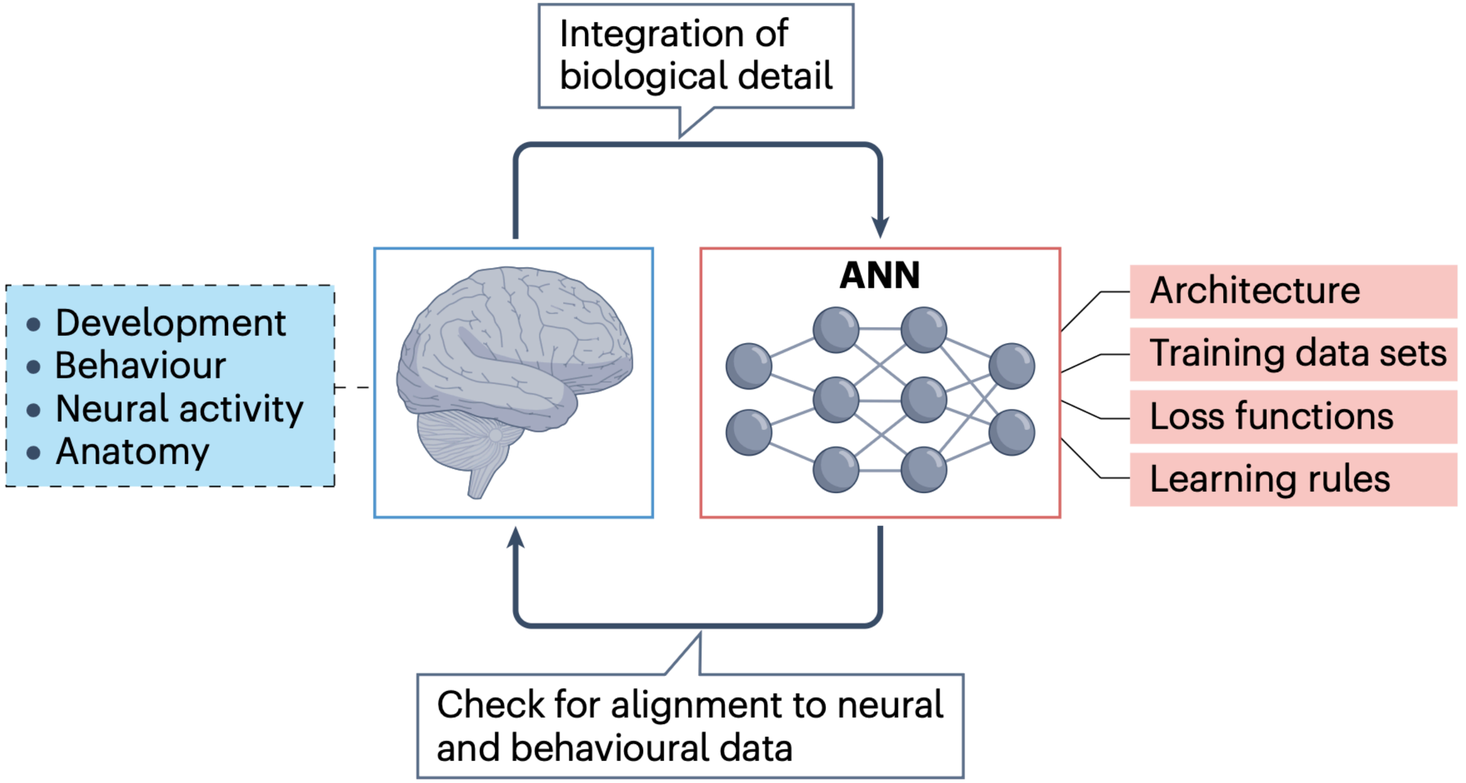






Universal approximation

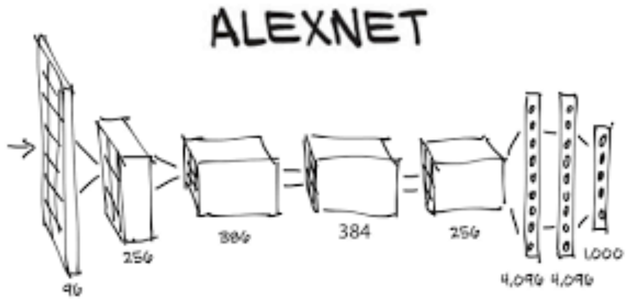
Relevant levers (e.g., Neuroconnectionism)



Leshno et al. 1993; Doerig et al. 2023; Cichy & Kaiser, 2019;

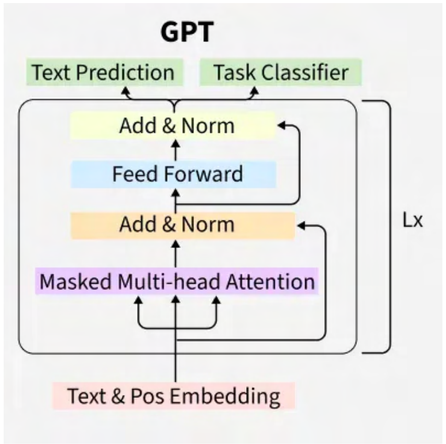
Enter: Artificial Neural Networks

Expressive and performant



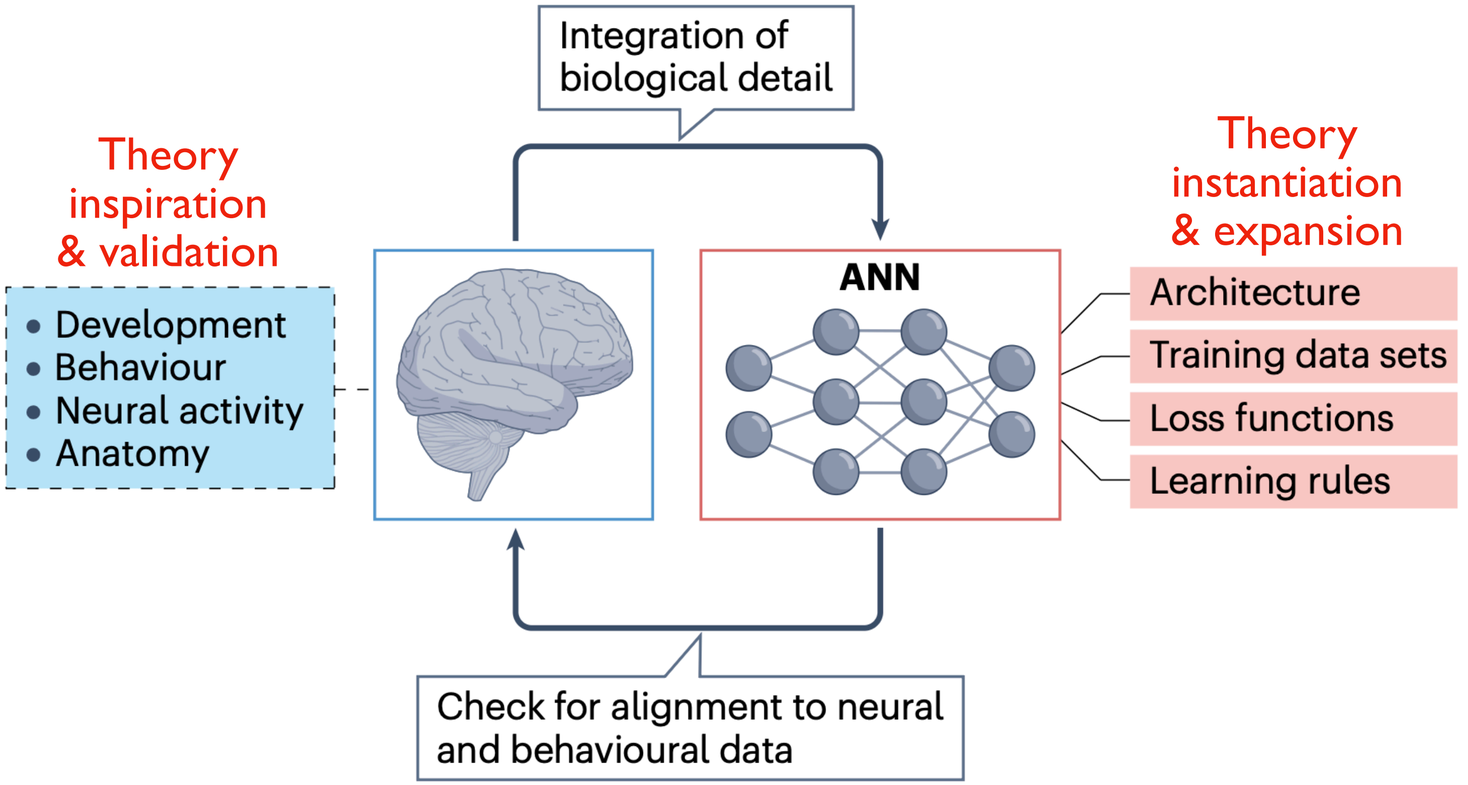






Universal approximation

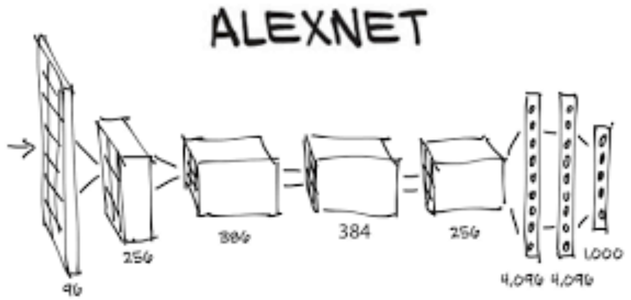
Relevant levers (e.g., Neuroconnectionism)



Leshno et al. 1993; Doerig et al. 2023; Cichy & Kaiser, 2019;

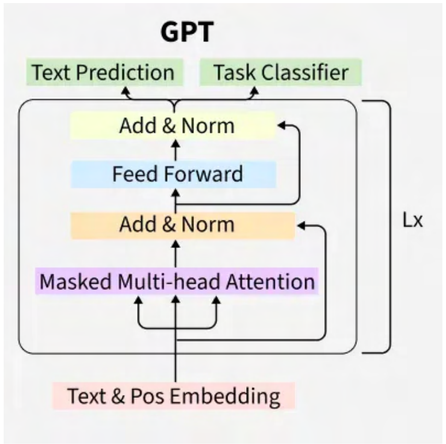
Enter: Artificial Neural Networks

Expressive and performant



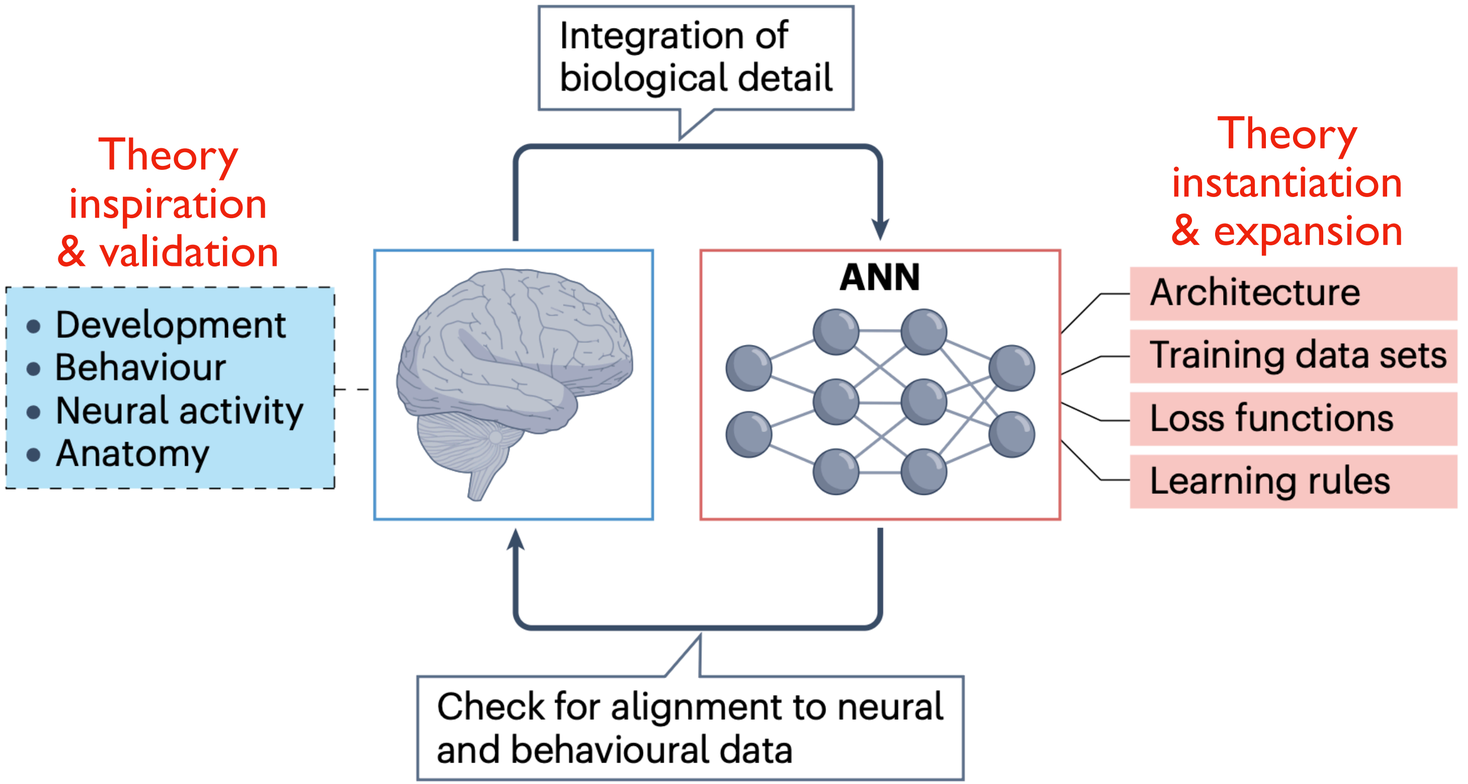






Universal approximation

Relevant levers (e.g., Neuroconnectionism)

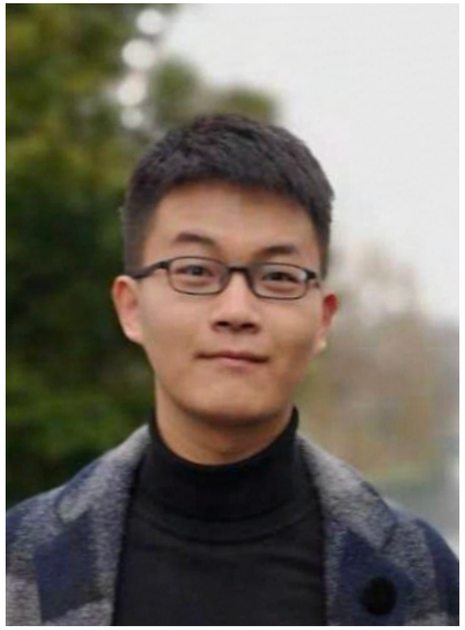


Let's look at two examples…

Leshno et al. 1993; Doerig et al. 2023; Cichy & Kaiser, 2019;

1. Shape-biased image classification

"Adopting a human developmental visual diet yields robust and shape-based AI vision."
Nature Machine Intelligence 8, 735–748 (2026).







KIETZMANN LAB

European Research Council
Established by the European Commission

UNIVERSITÄT OSNABRÜCK

Zejin Lu Radek Cichy Tim Kietzmann

Configural processing behaviour in humans

Configural difference

200ms

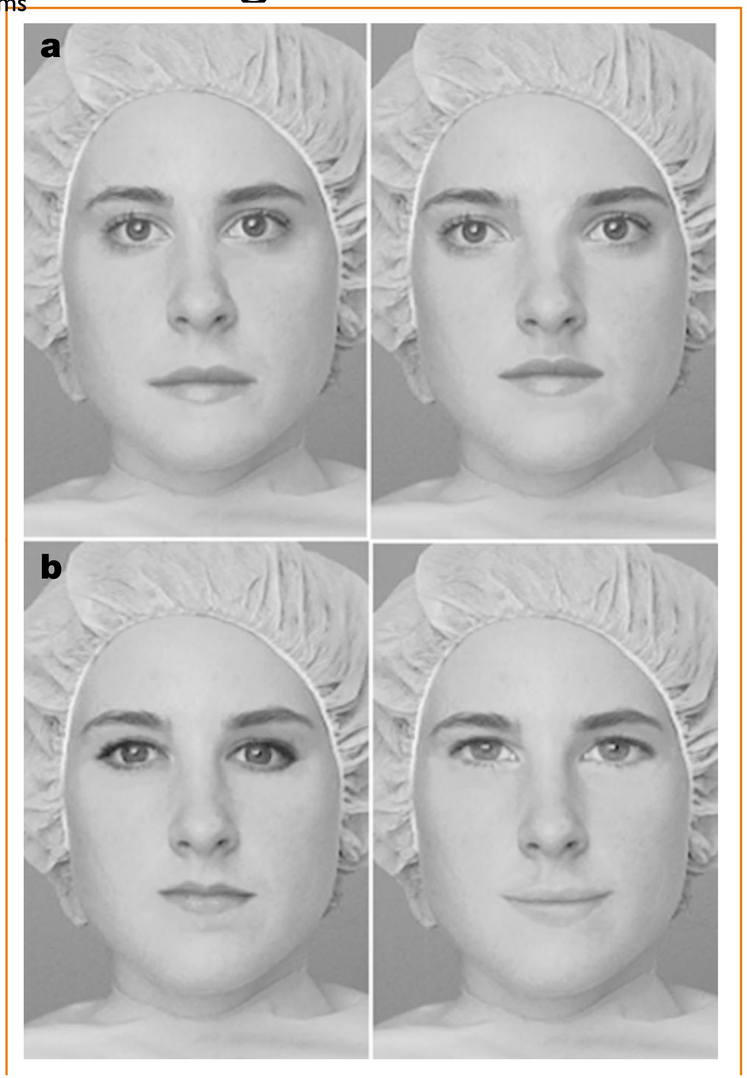


Le Grand et al. Nature 2001; Mondloch et al. Dev. Sci. 2013; Le Grand et al. Psych Sci. 2004; Vogelsang et al. 2024

Configural processing behaviour in humans

Configural difference

200ms



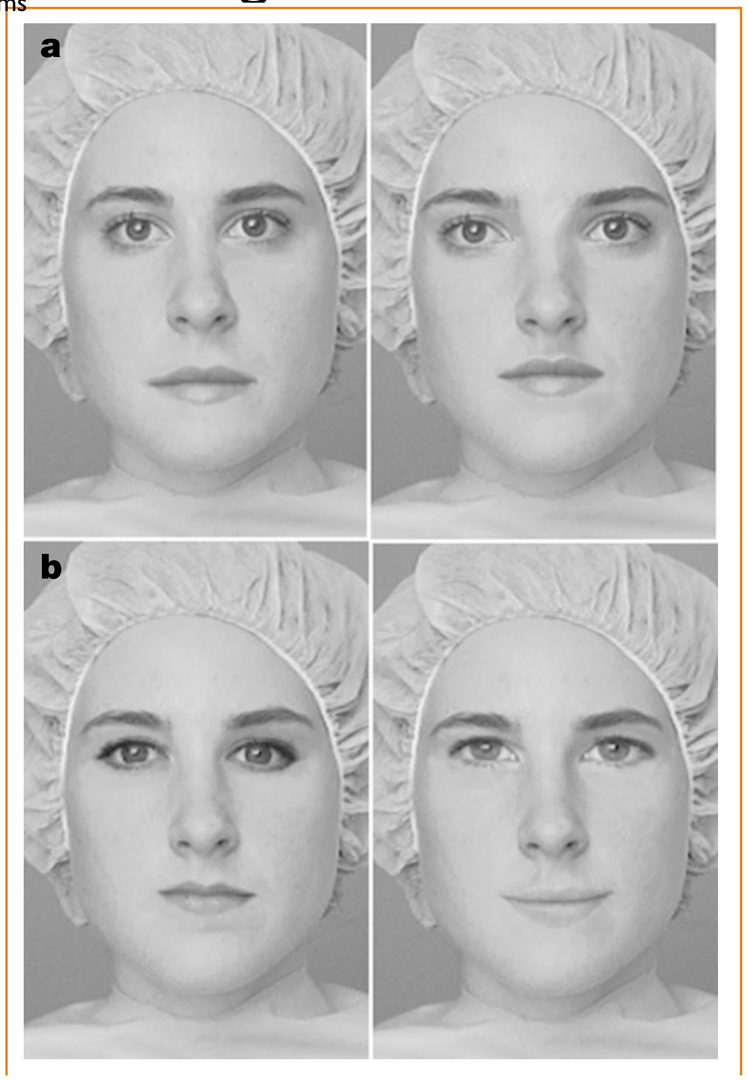
Featural difference

Le Grand et al. Nature 2001; Mondloch et al. Dev. Sci. 2013; Le Grand et al. Psych Sci. 2004; Vogelsang et al. 2024

Configural processing behaviour in humans

Configural difference

200ms



Featural difference

Table 1 Detection of facial configural and featural differences

Group	N	*Mean age (range)*	*Configural*		*Featural*	
			Upright	*Inverted*	*Upright*	*Inverted*
Adults	26	19 (18–22)	80 (1.9)	63 (2.2)	80 (1.4)	81 (1.8)

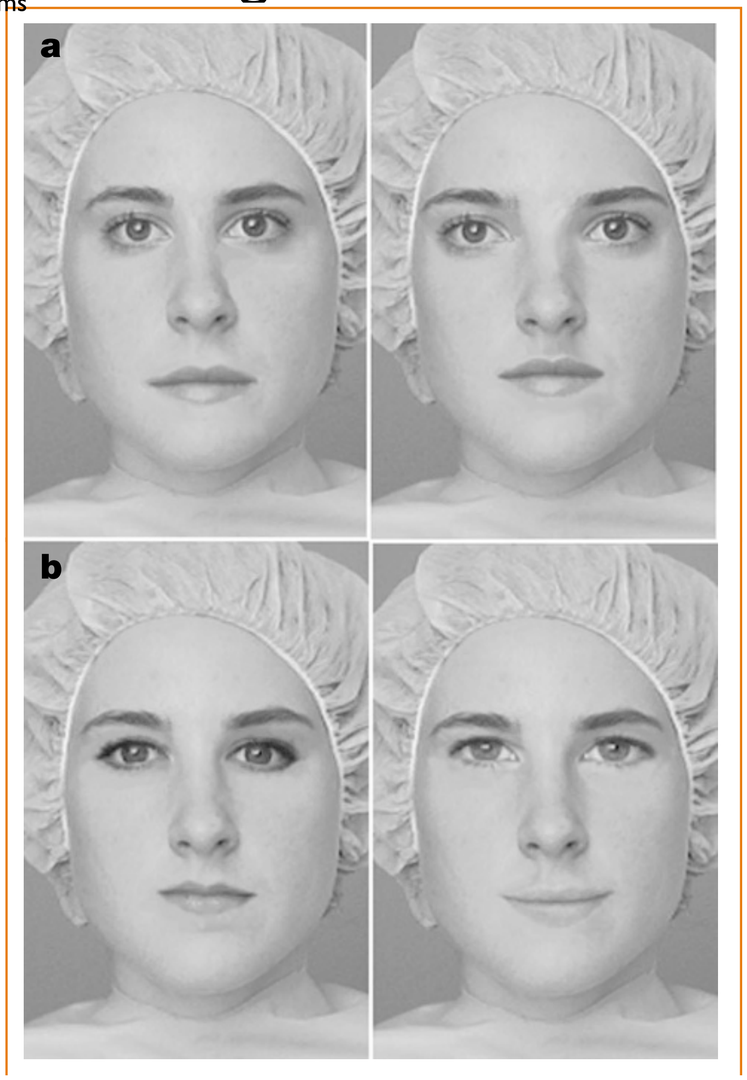
Le Grand et al. Nature 2001; Mondloch et al. Dev. Sci. 2013; Le Grand et al. Psych Sci. 2004; Vogelsang et al. 2024

Configural processing behaviour in humans



Configural difference

200ms



Featural difference

Patients had 62-187 days of visual deprivation post-birth due to dense cataracts

Table 1 Detection of facial configural and featural differences

Group	N	*Mean age (range)*	*Configural*		*Featural*	
			Upright	*Inverted*	*Upright*	*Inverted*
Adults	26	19 (18–22)	80 (1.9)	63 (2.2)	80 (1.4)	81 (1.8)

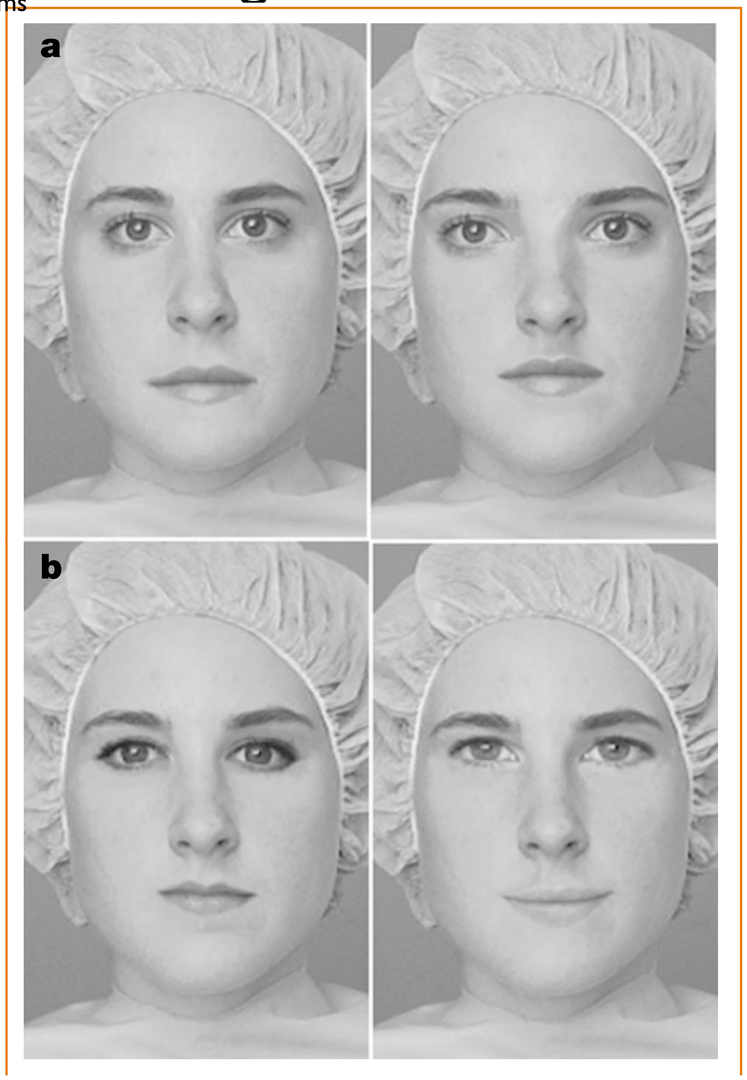
Le Grand et al. Nature 2001; Mondloch et al. Dev. Sci. 2013; Le Grand et al. Psych Sci. 2004; Vogelsang et al. 2024

Configural processing behaviour in humans



Configural difference

200ms



Featural difference

Patients had 62-187 days of visual deprivation post-birth due to dense cataracts

Table 1 Detection of facial configural and featural differences

Group	N	*Mean age (range)*	*Configural*		*Featural*	
			Upright	*Inverted*	*Upright*	*Inverted*
Adults	26	19 (18–22)	80 (1.9)	63 (2.2)	80 (1.4)	81 (1.8)
Controls	14	14 (9–21)	81 (2.7)	59 (1.8)	79 (1.9)	80 (2.9)
Patients	14	14 (9–21)	62 (3.2)	55 (3.0)	80 (2.4)	78 (3.1)

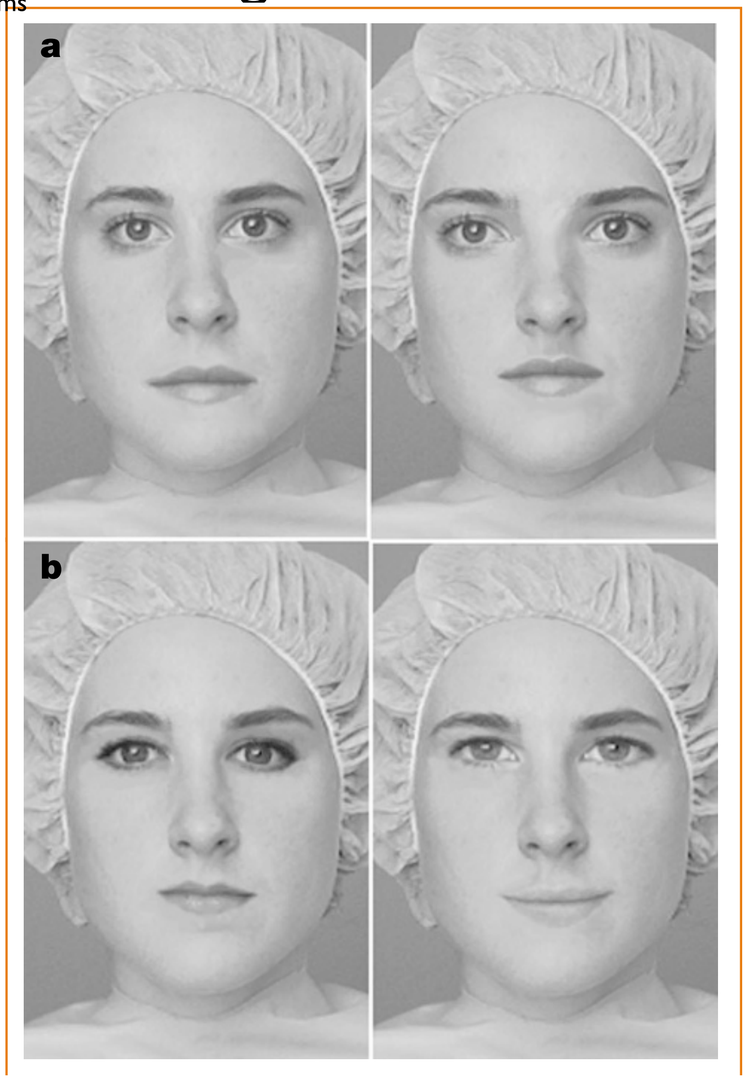
Le Grand et al. Nature 2001; Mondloch et al. Dev. Sci. 2013; Le Grand et al. Psych Sci. 2004; Vogelsang et al. 2024

Configural processing behaviour in humans



Configural difference

200ms



Featural difference

Patients had 62-187 days of visual deprivation post-birth due to dense cataracts

Table 1 Detection of facial configural and featural differences

Group	N	*Mean age (range)*	*Configural*		*Featural*	
			Upright	*Inverted*	*Upright*	*Inverted*
Adults	26	19 (18–22)	80 (1.9)	63 (2.2)	80 (1.4)	81 (1.8)
Controls	14	14 (9–21)	81 (2.7)	59 (1.8)	79 (1.9)	80 (2.9)
Patients	14	14 (9–21)	62 (3.2)	55 (3.0)	80 (2.4)	78 (3.1)

Even after years of experience, configural face processing specifically was poorer.

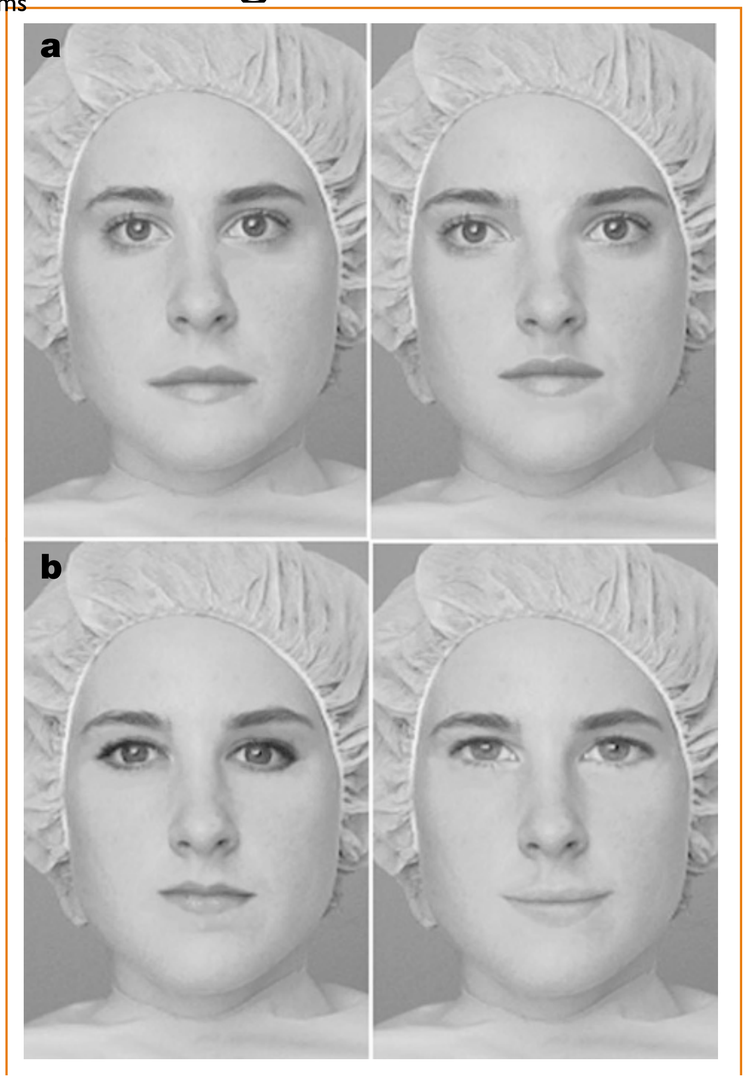
Le Grand et al. Nature 2001; Mondloch et al. Dev. Sci. 2013; Le Grand et al. Psych Sci. 2004; Vogelsang et al. 2024

Configural processing behaviour in humans



Configural difference

200ms



Featural difference

Patients had 62-187 days of visual deprivation post-birth due to dense cataracts

Table 1 Detection of facial configural and featural differences

Group	N	*Mean age (range)*	*Configural*		*Featural*	
			Upright	*Inverted*	*Upright*	*Inverted*
Adults	26	19 (18–22)	80 (1.9)	63 (2.2)	80 (1.4)	81 (1.8)
Controls	14	14 (9–21)	81 (2.7)	59 (1.8)	79 (1.9)	80 (2.9)
Patients	14	14 (9–21)	62 (3.2)	55 (3.0)	80 (2.4)	78 (3.1)

Even after years of experience, configural face processing specifically was poorer.

Theory—*Adaptive Initial Degradation:*

Blurry vision has low spatial frequencies which, for faces, specifies the global contour and location of features but little of their detail.

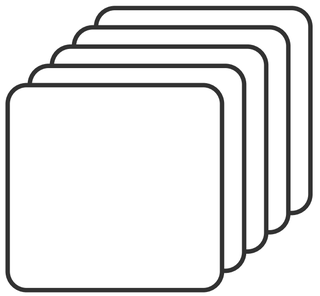
Le Grand et al. Nature 2001; Mondloch et al. Dev. Sci. 2013; Le Grand et al. Psych Sci. 2004; Vogelsang et al. 2024

Theory instantiation: Developmental visual diet (DVD) for ANNs

…generalised to objects

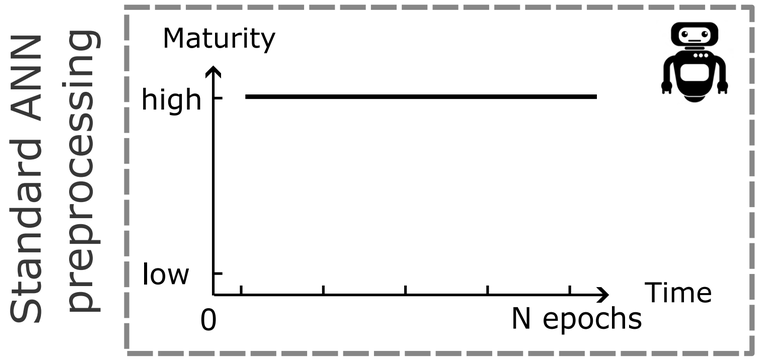
Image dataset

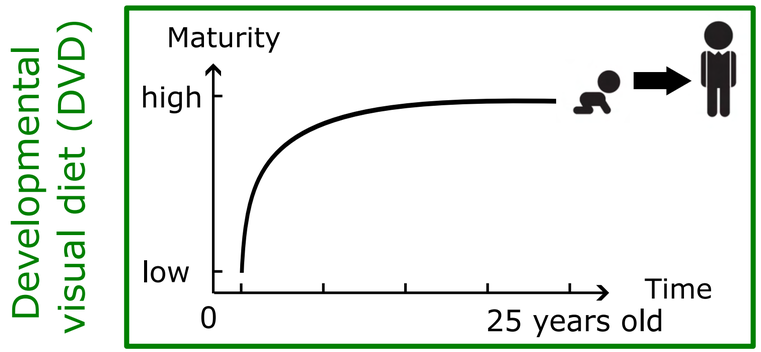
Dataset preprocessing



ecoset

IMAGENET





Mehrer et al. 2018

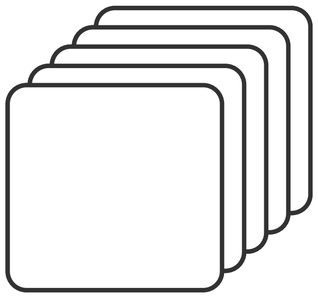
Theory instantiation: Developmental visual diet (DVD) for ANNs

…generalised to objects

Image dataset

Dataset preprocessing

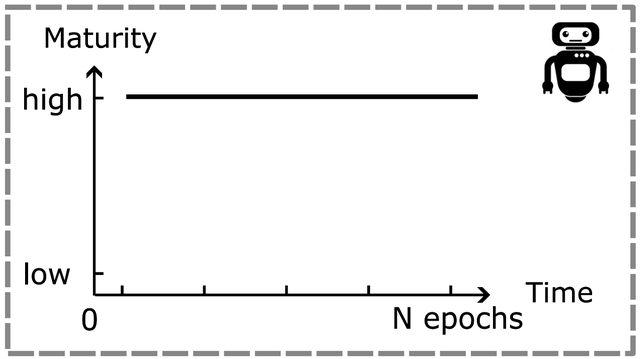
ANN training



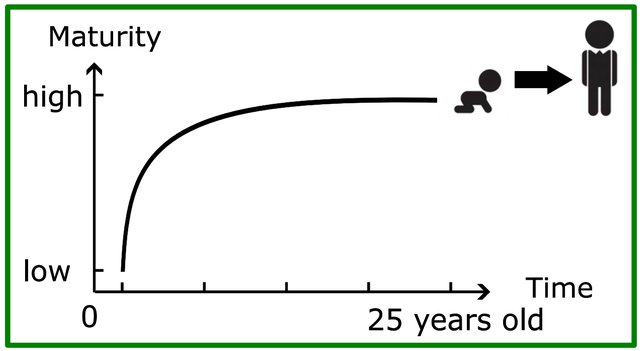
ecoset

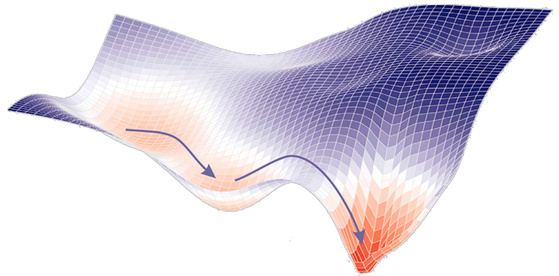
IMAGENET

Standard ANN preprocessing



Developmental visual diet (DVD)





- Alexnet
- ResNext-50
- VGG
- ViT
- MobileNet
- DeiT
- ResNet-50
- …

Mehrer et al. 2018

Theory instantiation: Developmental visual diet (DVD) for ANNs

…generalised to objects

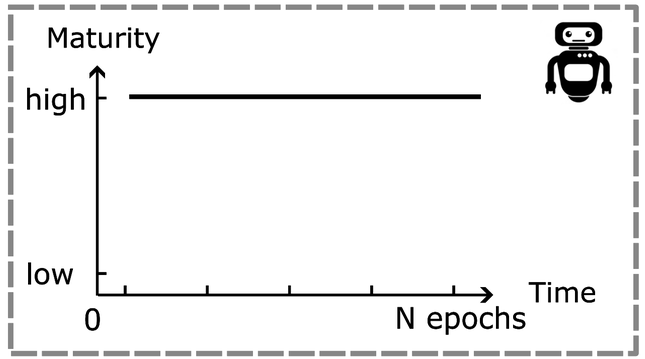
Image dataset

ecoset

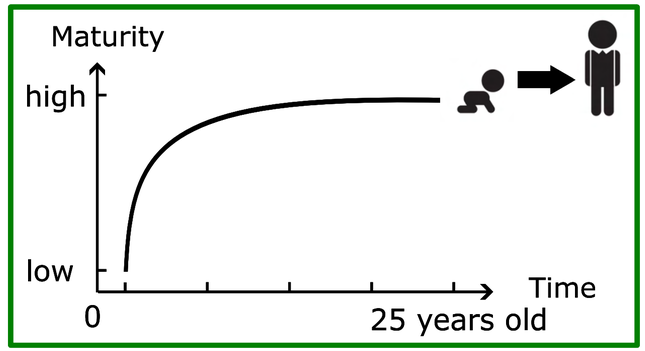
IMAGENET

Dataset preprocessing

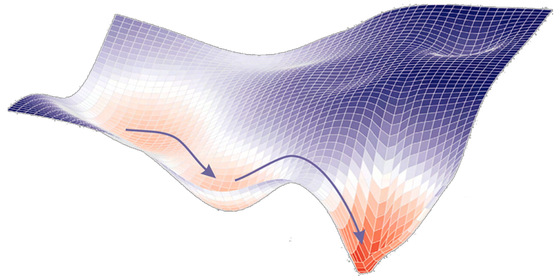
Standard ANN preprocessing



Developmental visual diet (DVD)



ANN training



Alexnet | ResNext-50

VGG | ViT

MobileNet | DeiT

ResNet-50 | …

Robust vision test battery

Shape-texture bias



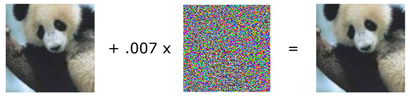
Abstract shape recognition



Degradation robustness

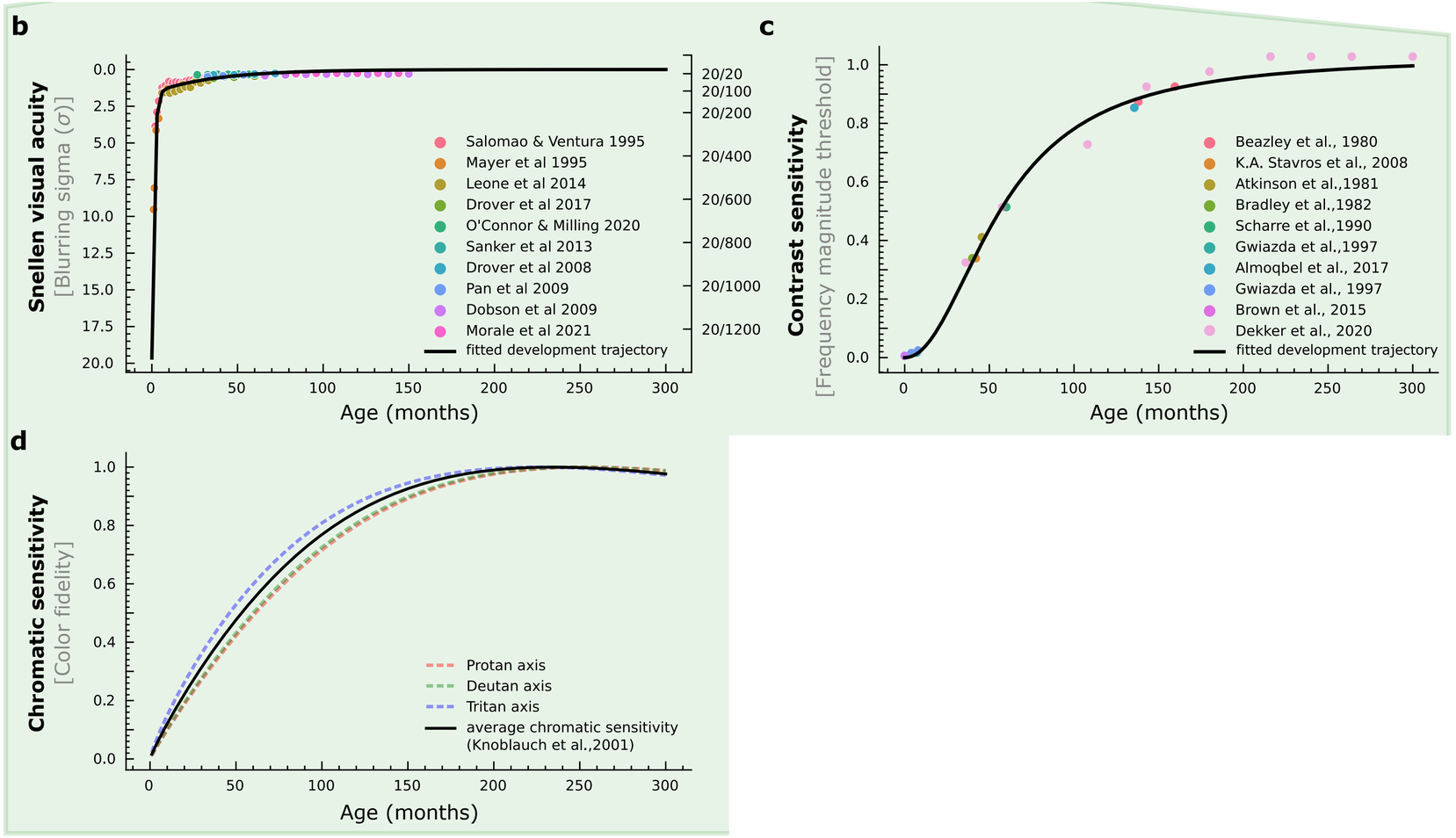


Adversarial robustness



Mehrer et al. 2018

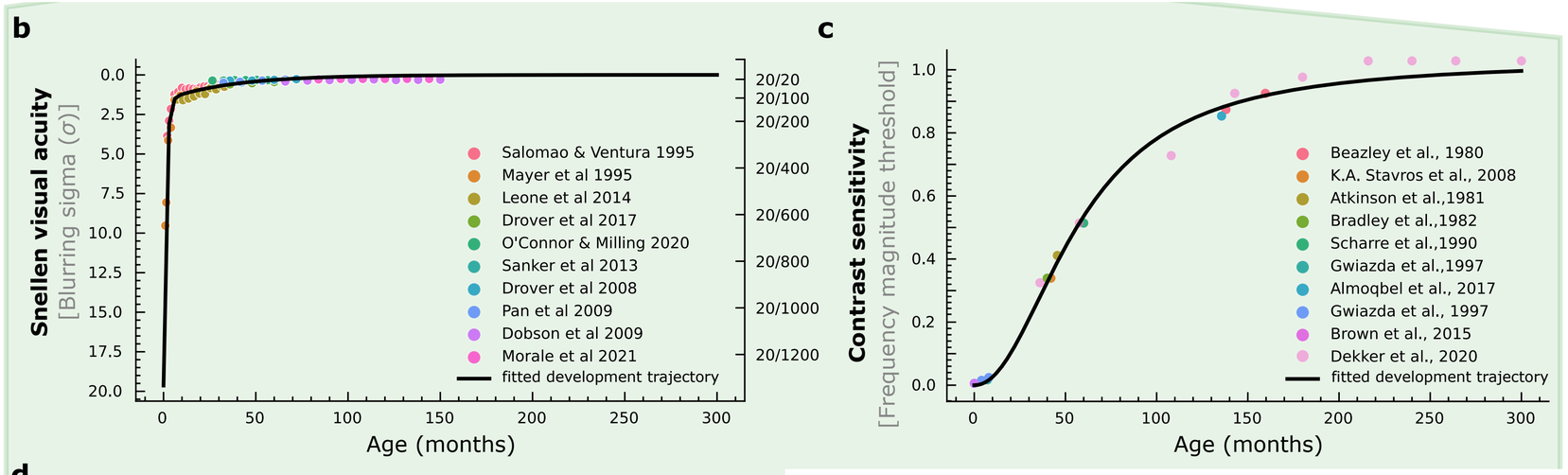
DVD details

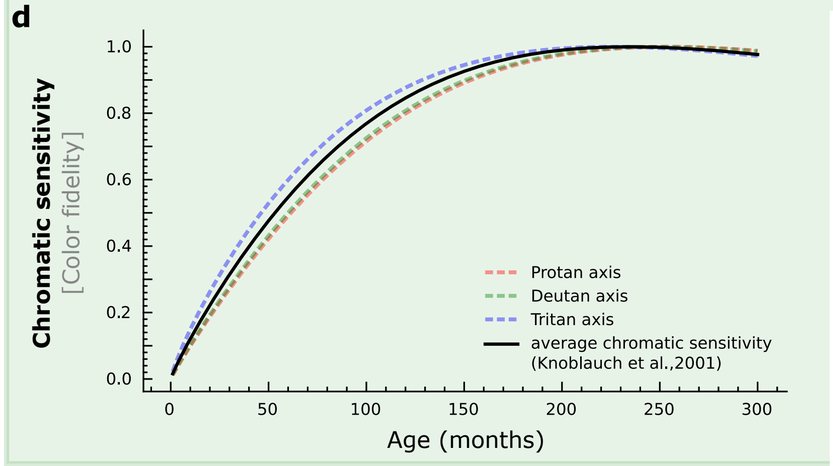


Visual acuity
Chromatic sensitivity
Contrast sensitivity

0-25yrs

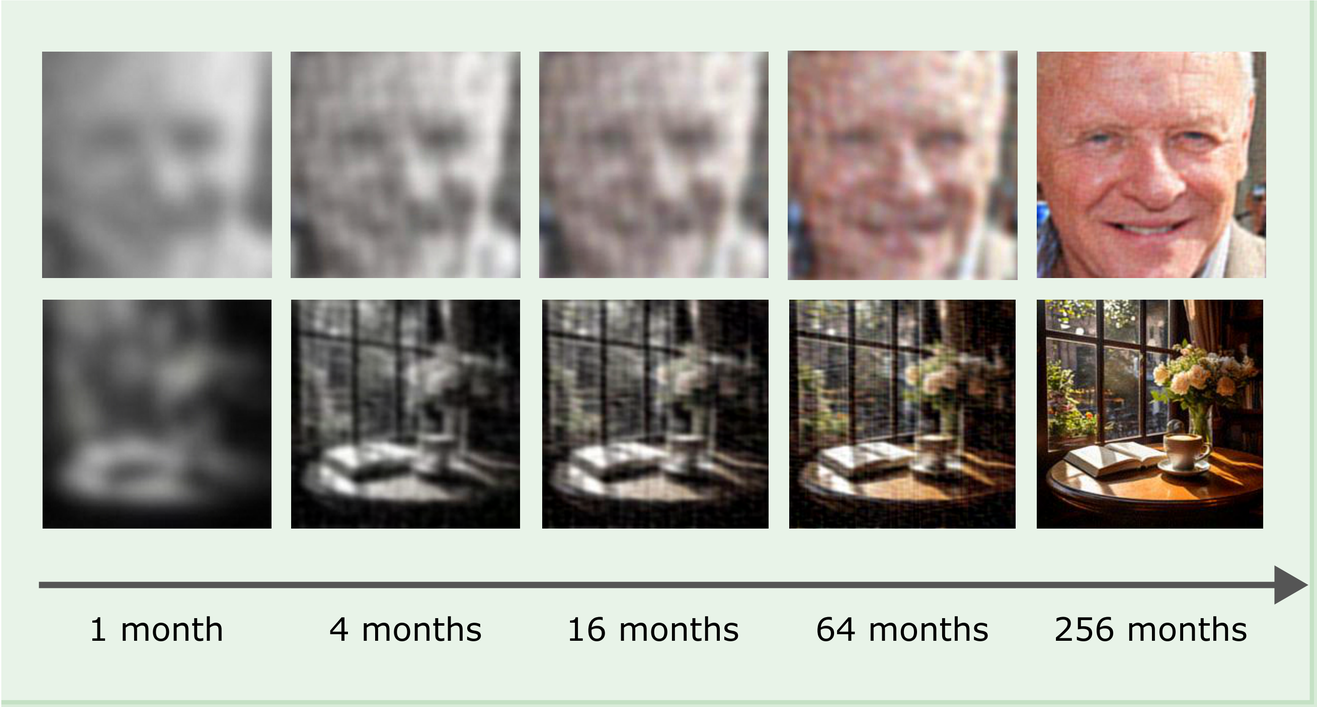
DVD details





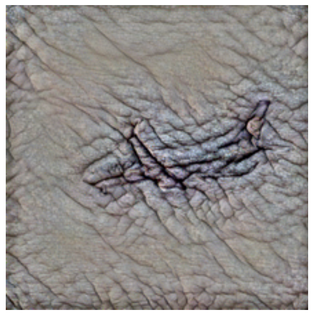
Visual acuity
Chromatic sensitivity
Contrast sensitivity

0-25yrs



Evaluating shape-biased (configural) behaviour

Cue-conflict images



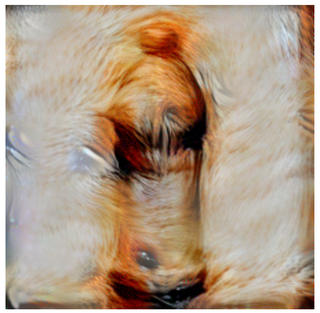



Shape class: airplane
Texture class: elephant

Shape class: bear
Texture class: clock

Shape class: bicycle
Texture class: elephant

Shape class: bird
Texture class: clock



Shape class: boat
Texture class: bottle

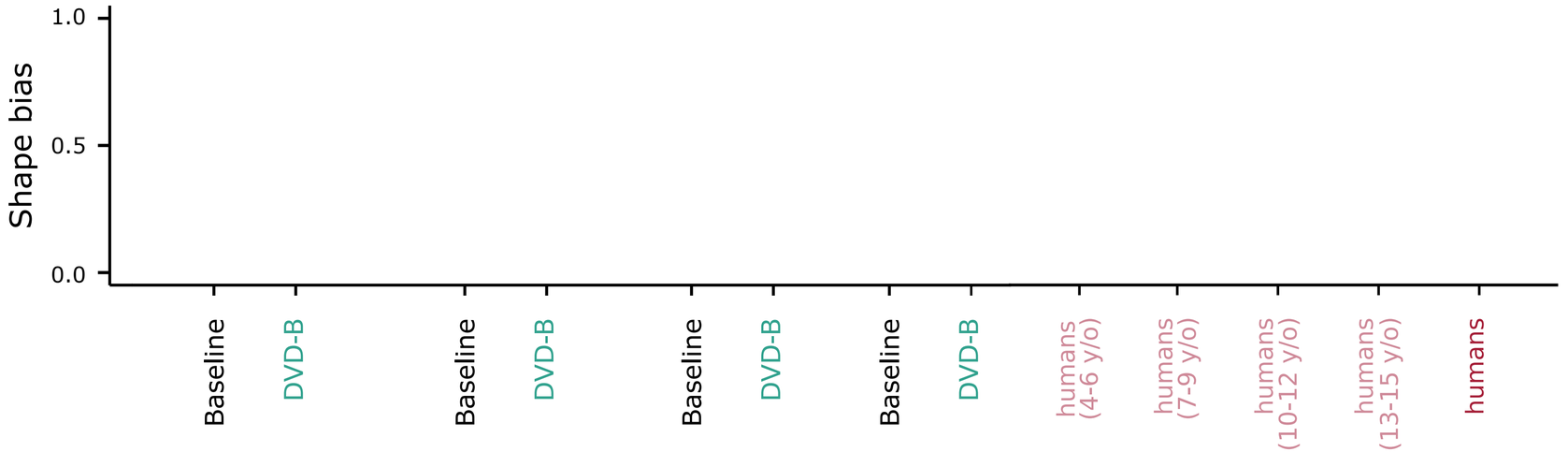
Shape class: bottle
Texture class: dog

Shape class: car
Texture class: clock

Shape class: cat
Texture class: elephant

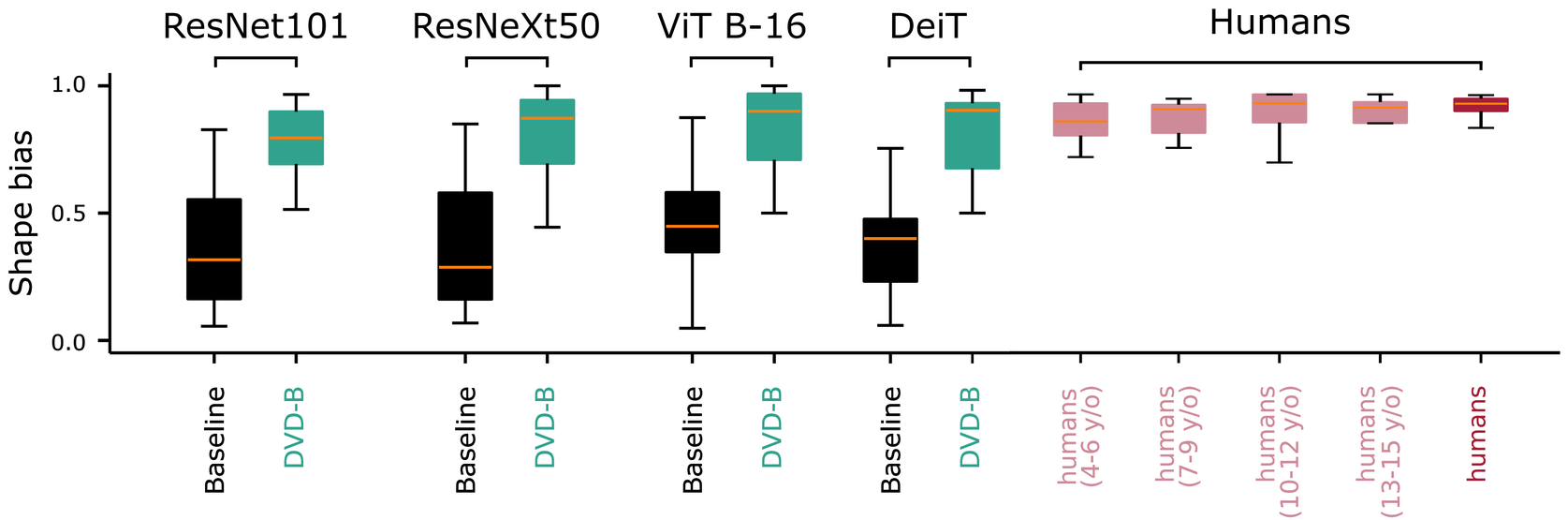
Geirhos et al. 2019

Assessing ANN behaviour



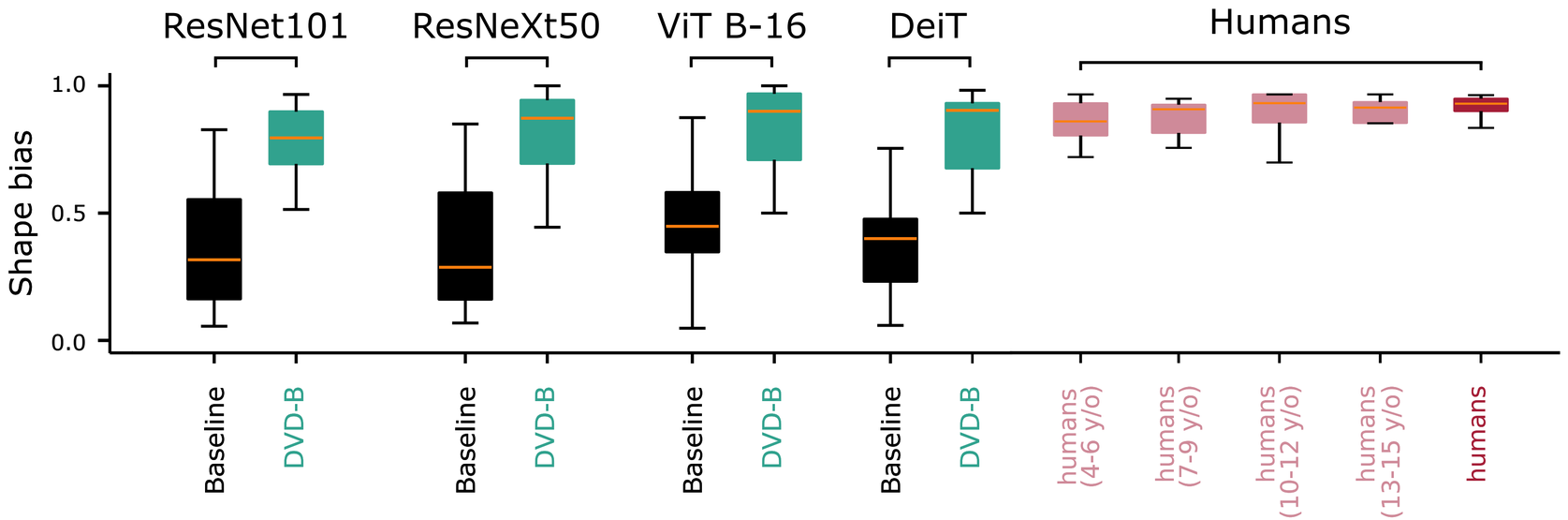
Geirhos et al. 2019; Selvaraju et al. 2017

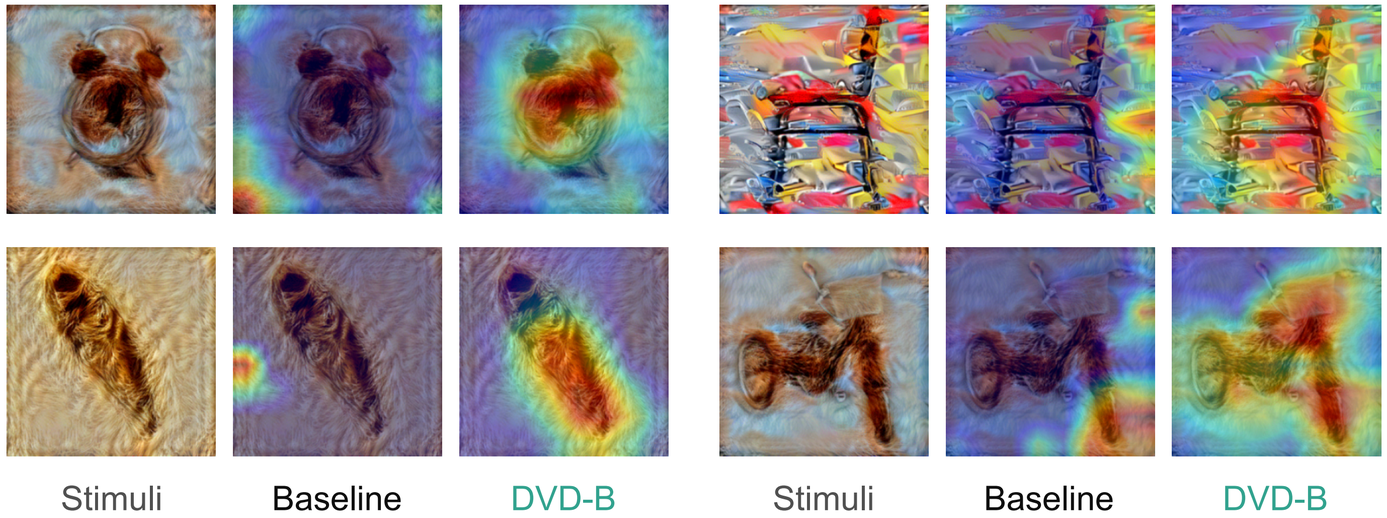
Assessing ANN behaviour



Geirhos et al. 2019; Selvaraju et al. 2017

Assessing ANN behaviour

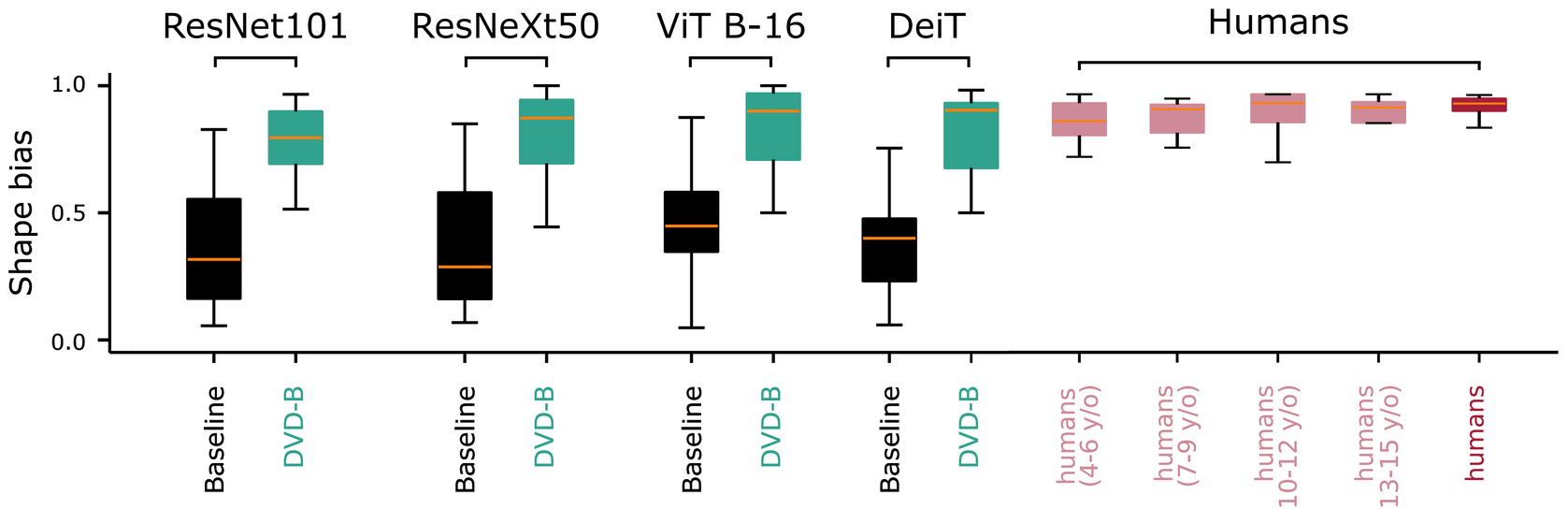


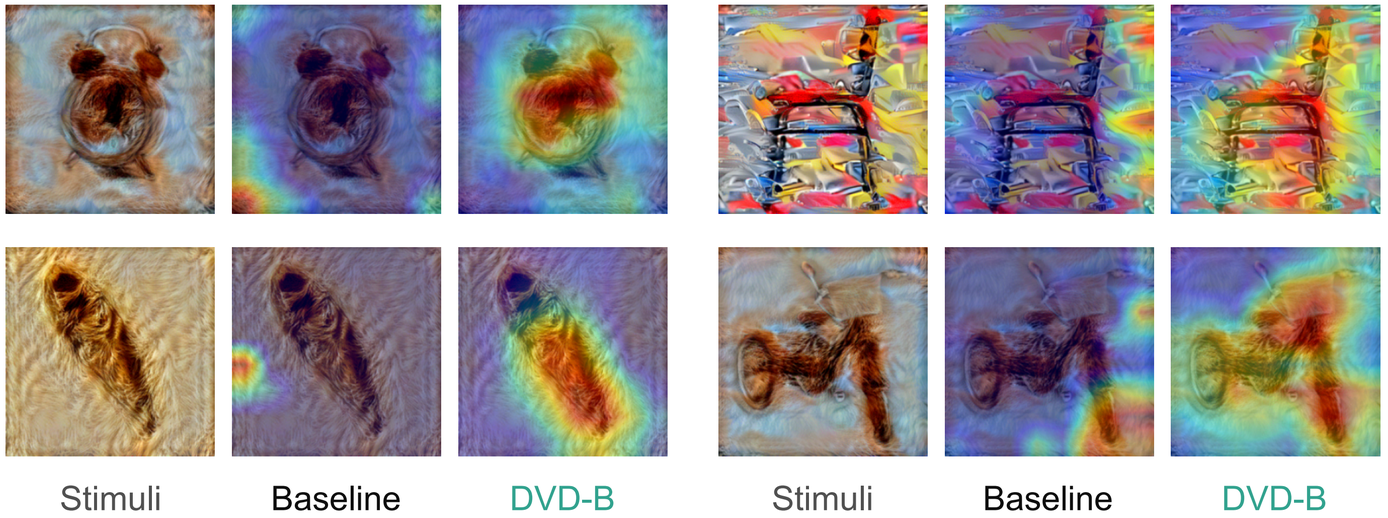


GRAD-CAM "predicted-class maps"

Geirhos et al. 2019; Selvaraju et al. 2017

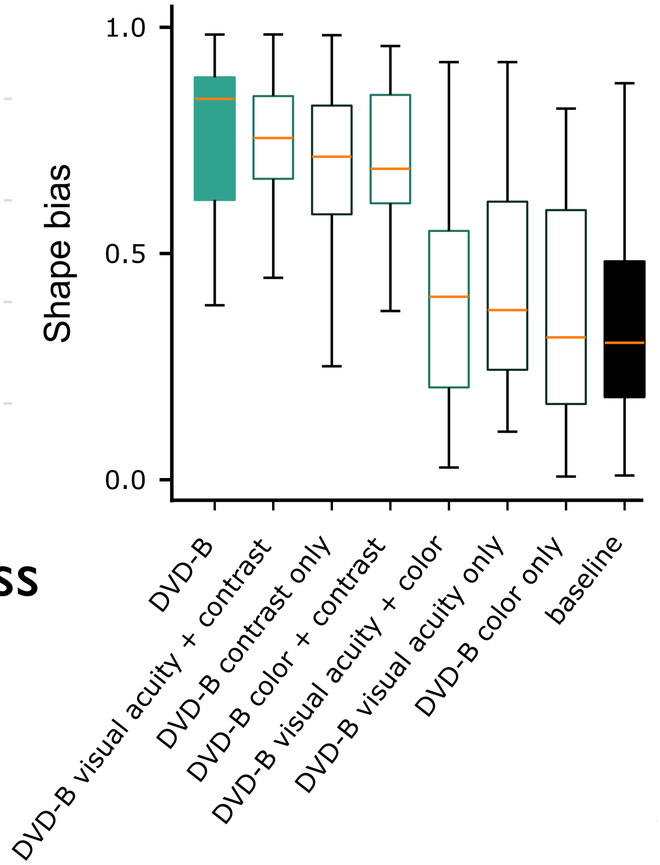
Assessing ANN behaviour





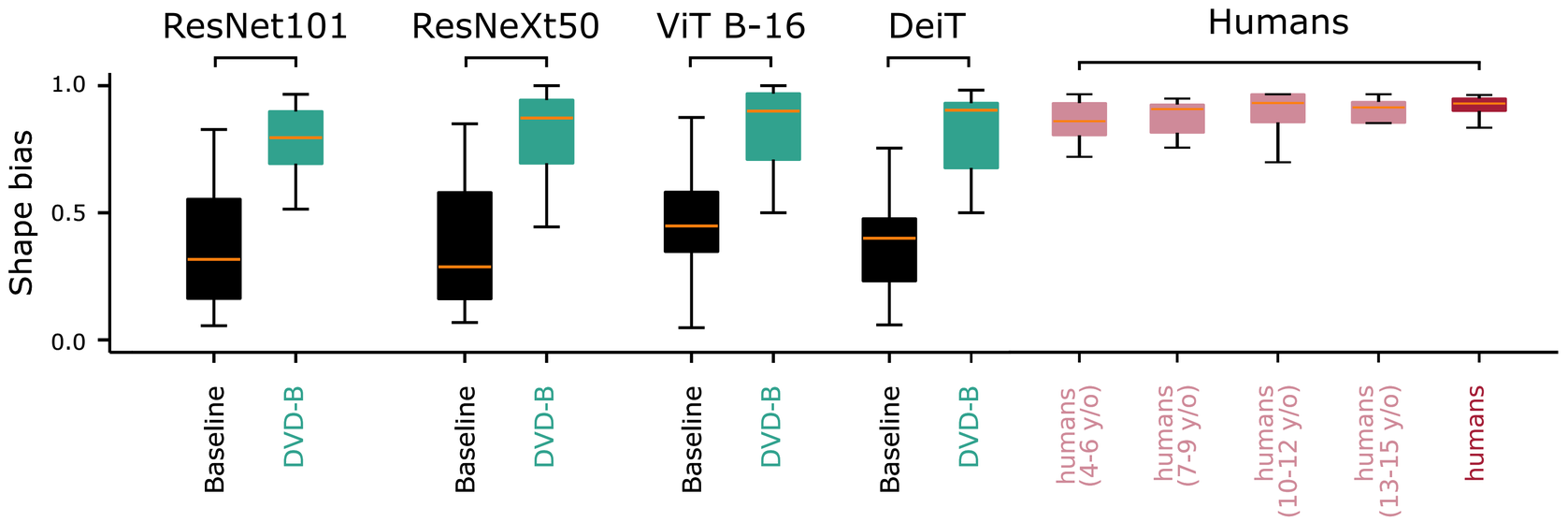
GRAD-CAM "predicted-class maps"

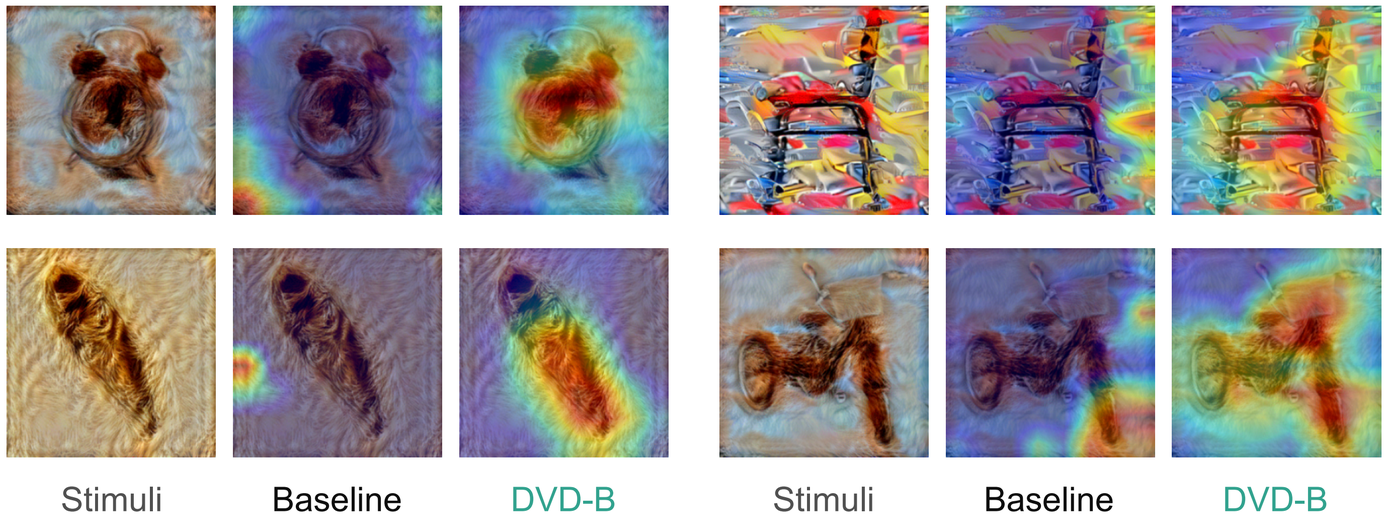
Controlled rearing



Geirhos et al. 2019; Selvaraju et al. 2017

Assessing ANN behaviour

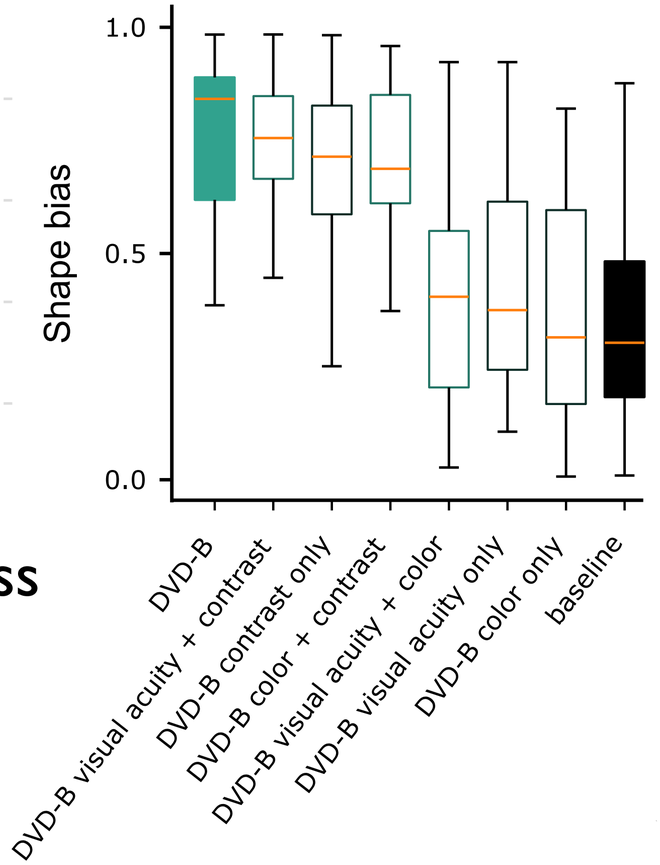




GRAD-CAM
"predicted-class maps"

All 3 factors essential
& *Contrast* >> *Acuity*

Controlled rearing



Geirhos et al. 2019; Selvaraju et al. 2017

Initially-degraded vision indeed leads to a, configural processing-like, shape bias!

✅ Computational demonstration

Initially-degraded vision indeed leads to a, configural processing-like, shape bias!

✅ Computational demonstration

✅ By-product: SOTA shape-biased ANNs

Initially-degraded vision indeed leads to a, configural processing-like, shape bias!

✅ Computational demonstration

✅ By-product: SOTA shape-biased ANNs

Novel hypothesis: Contrast sensitivity, not visual acuity, is the main driver **?**

Initially-degraded vision indeed leads to a, configural processing-like, shape bias!

✅ Computational demonstration

✅ By-product: SOTA shape-biased ANNs

Novel hypothesis: Contrast sensitivity, not visual acuity, is the main driver **?**

Further check: Is the DVD-trained network more brain-aligned? *Longitudinally*?

→ if not, what do we change (architecture, data diet, learning rule) to improve?

2. Representational Drift

"Keep moving: identifying task-relevant subspaces to maximise plasticity for newly learned tasks"
Proceedings of The 3rd Conference on Lifelong Learning Agents, PMLR 274:815-831, 2025.







Daniel Anthes Peter König Tim Kietzmann

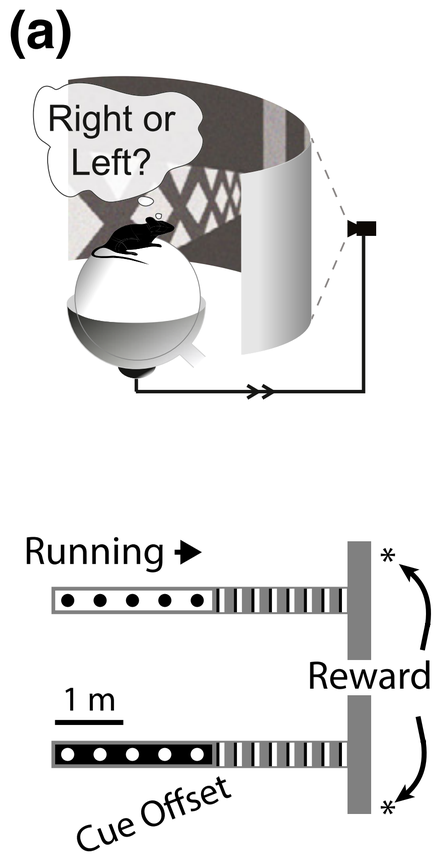
KIETZMANN LAB

European Research Council
Established by the European Commission

UNIVERSITÄT OSNABRÜCK

DFG Deutsche Forschungsgemeinschaft

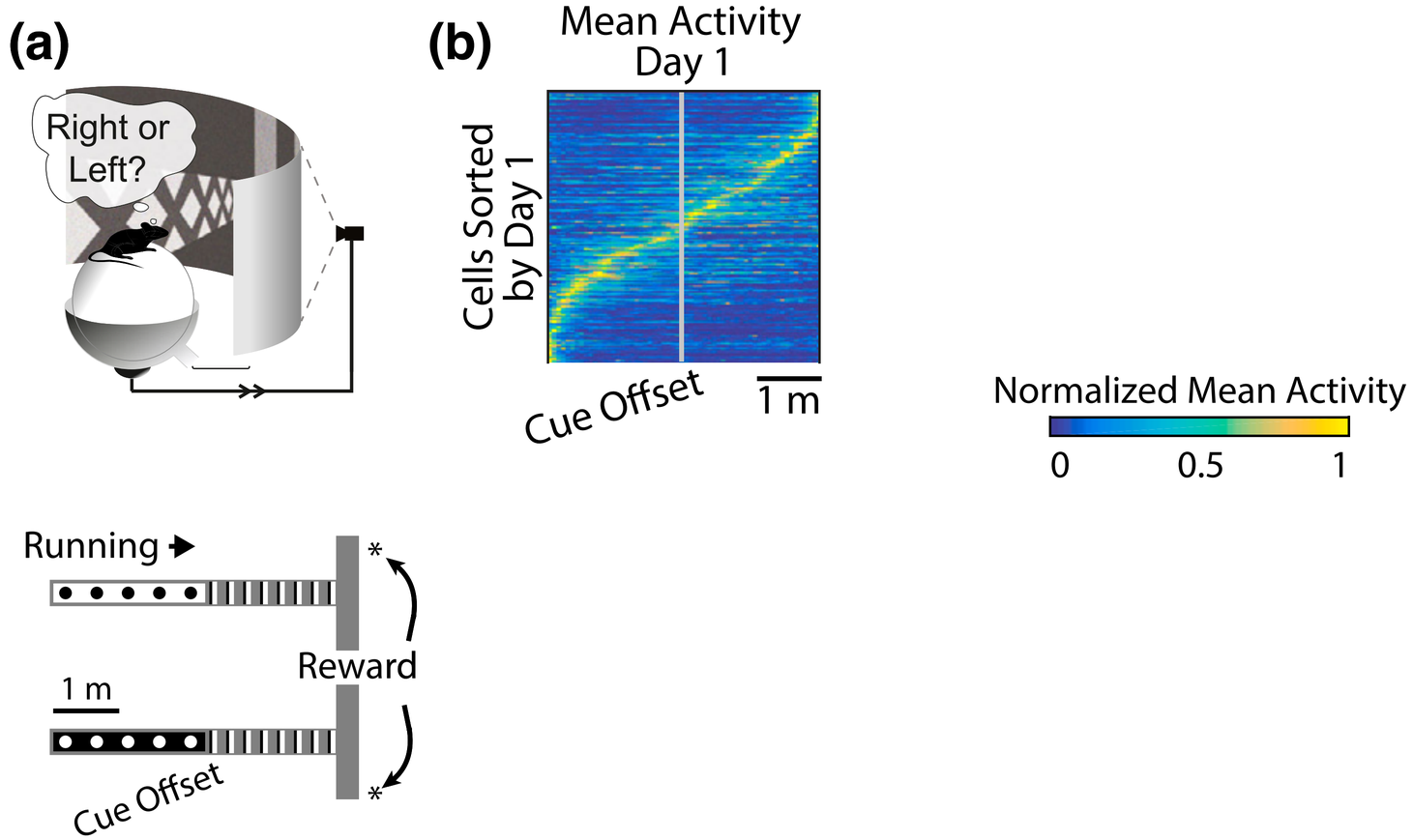
Representational drift in biology



Rule et al. 2019; Driscoll et al. 2017; Schoonover et al. 2021; Driscoll et al. 2022

Representational drift in biology

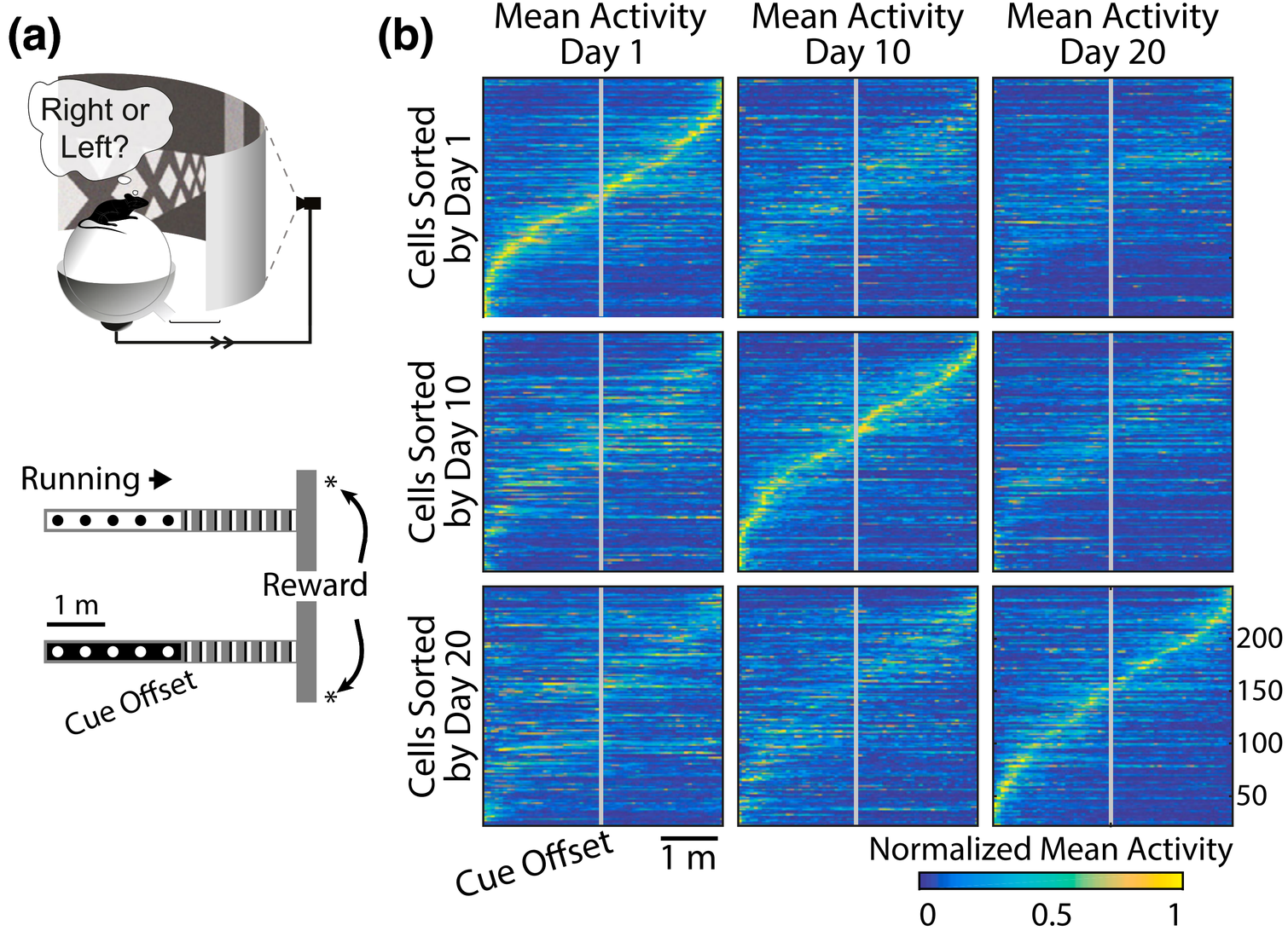
PPC neurons during stem movement



Rule et al. 2019; Driscoll et al. 2017; Schoonover et al. 2021; Driscoll et al. 2022

Representational drift in biology

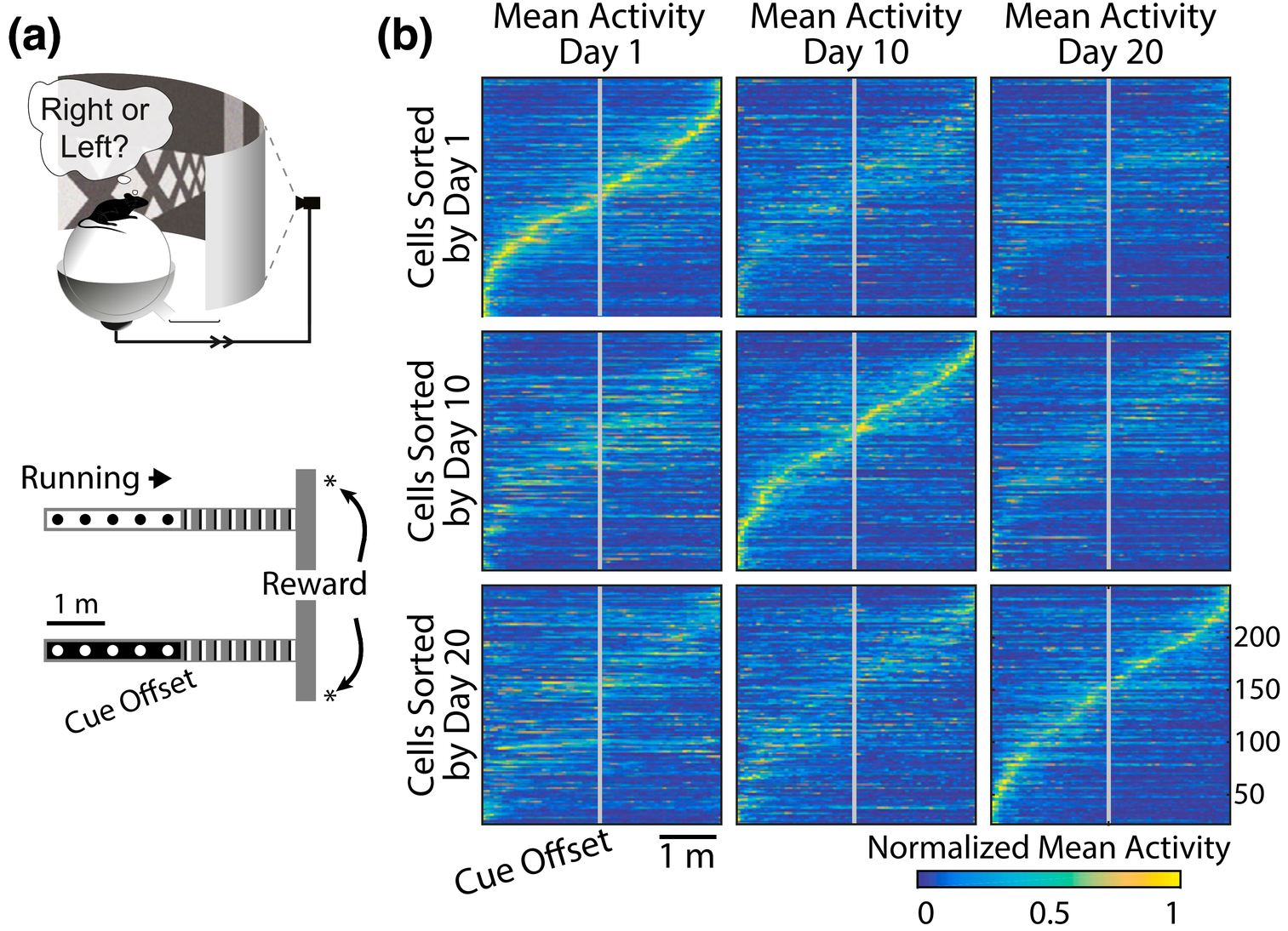
PPC neurons during stem movement

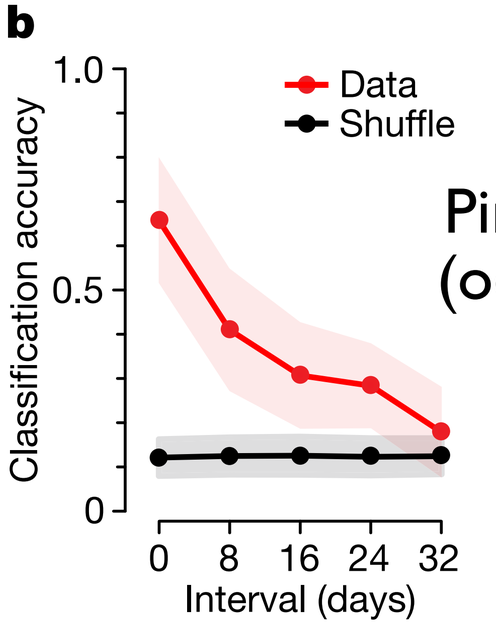


Rule et al. 2019; Driscoll et al. 2017; Schoonover et al. 2021; Driscoll et al. 2022

Representational drift in biology

PPC neurons during stem movement



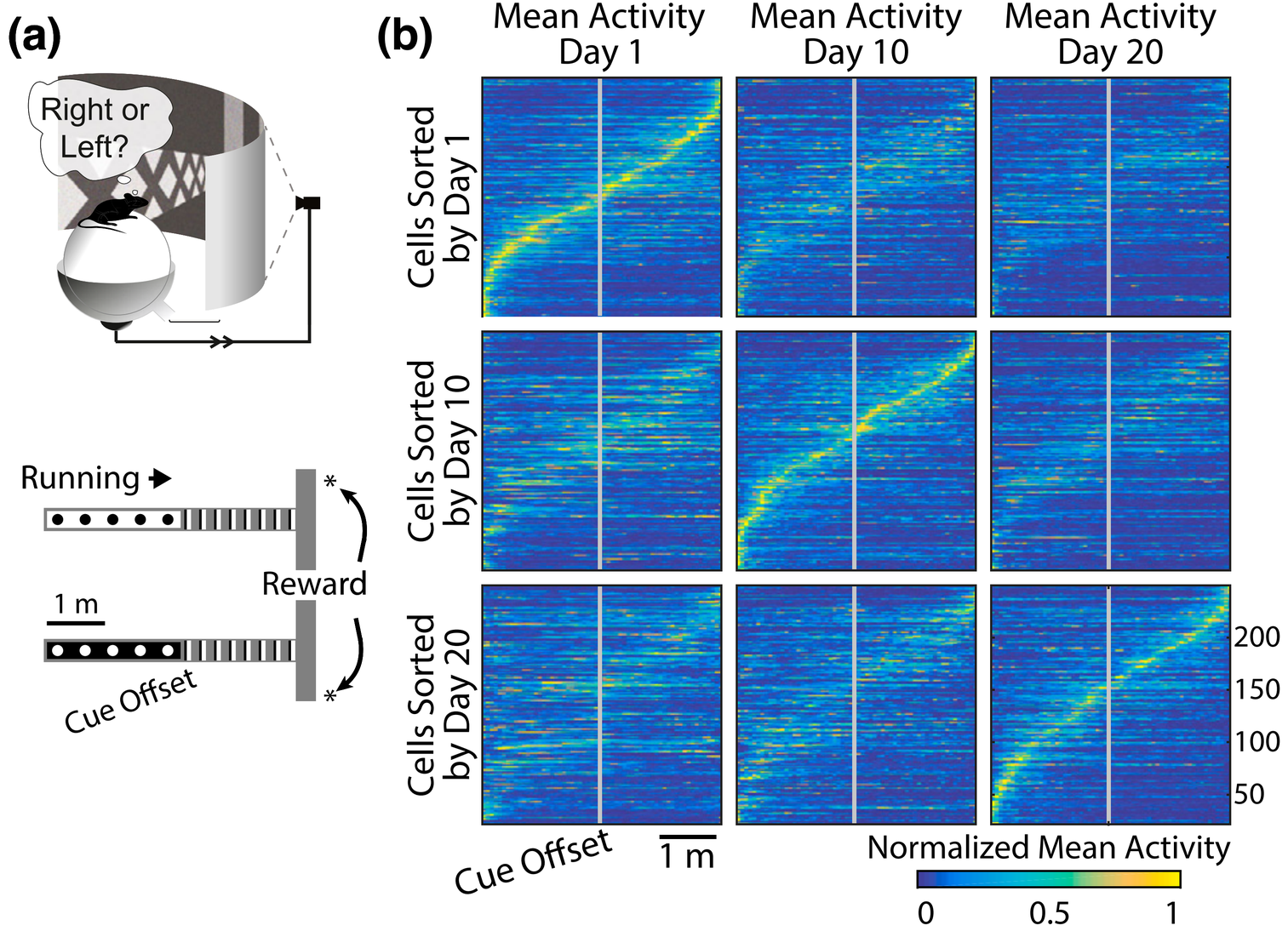


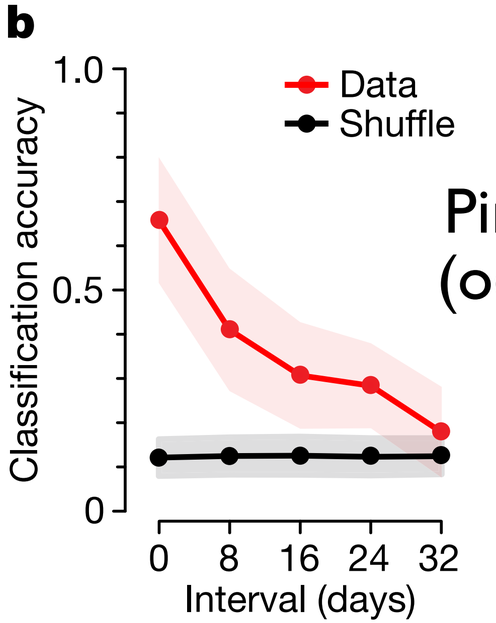
Piriform neurons (odour response)

Rule et al. 2019; Driscoll et al. 2017; Schoonover et al. 2021; Driscoll et al. 2022

Representational drift in biology

PPC neurons during stem movement





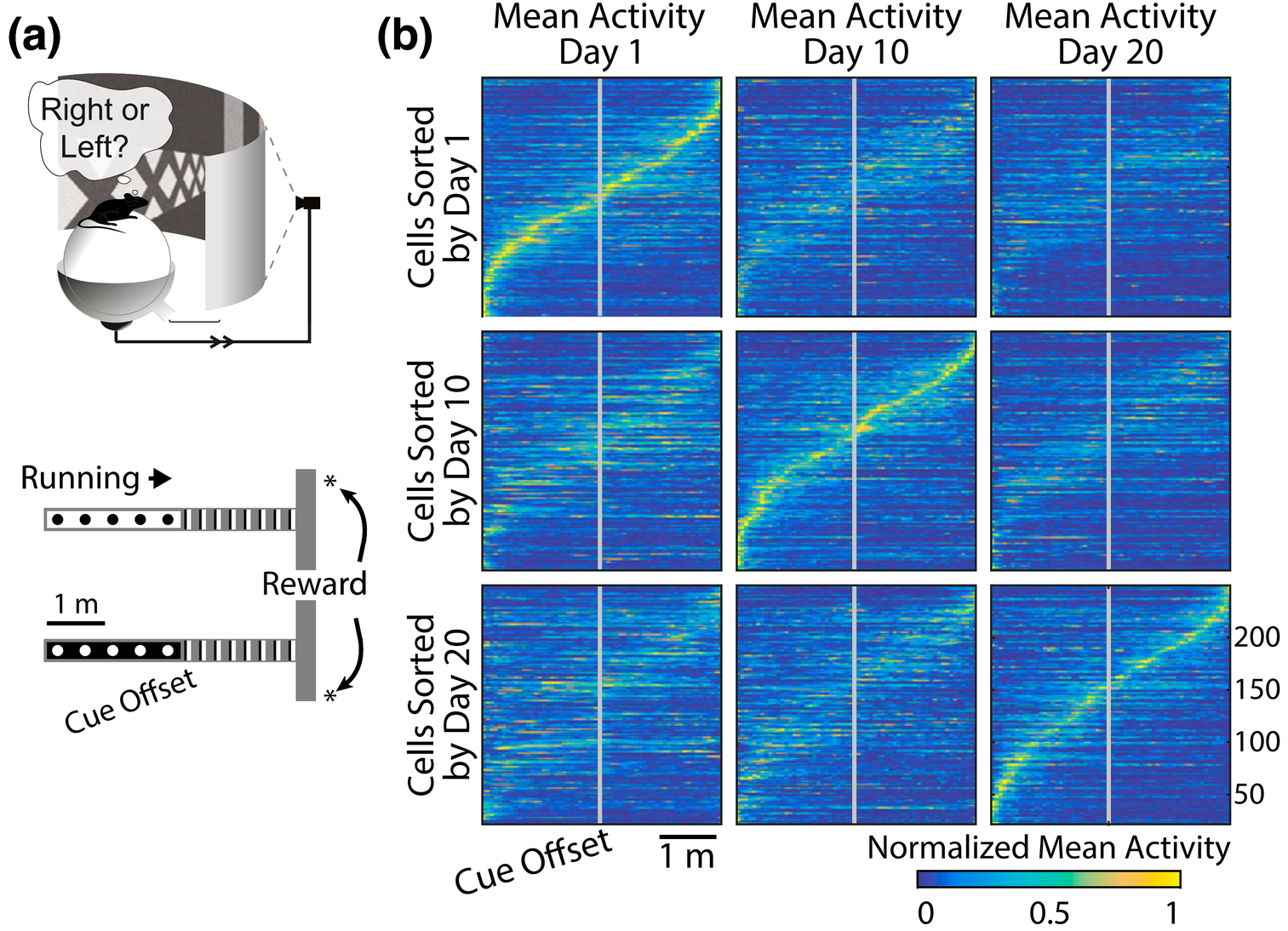
Piriform neurons (odour response)

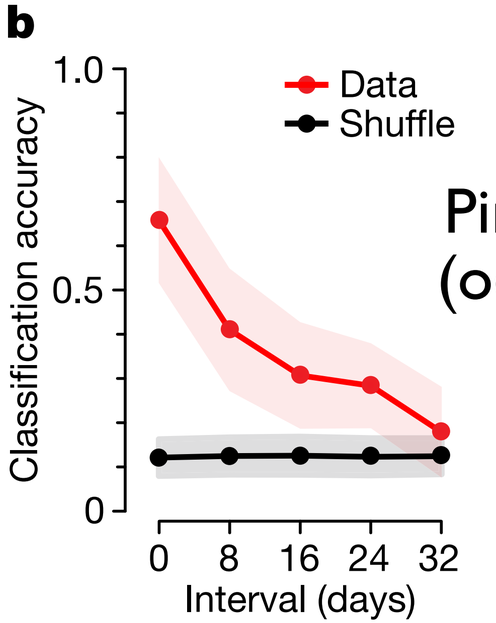
Bug or feature?

Rule et al. 2019; Driscoll et al. 2017; Schoonover et al. 2021; Driscoll et al. 2022

Representational drift in biology

PPC neurons during stem movement





Piriform neurons (odour response)

Bug or feature?

One theory: "*drift reflects learning*"

Rule et al. 2019; Driscoll et al. 2017; Schoonover et al. 2021; Driscoll et al. 2022

Theory instantiation & expansion: Continual learning in ANNs

CIFAR-100 split into 10 tasks; train sequentially

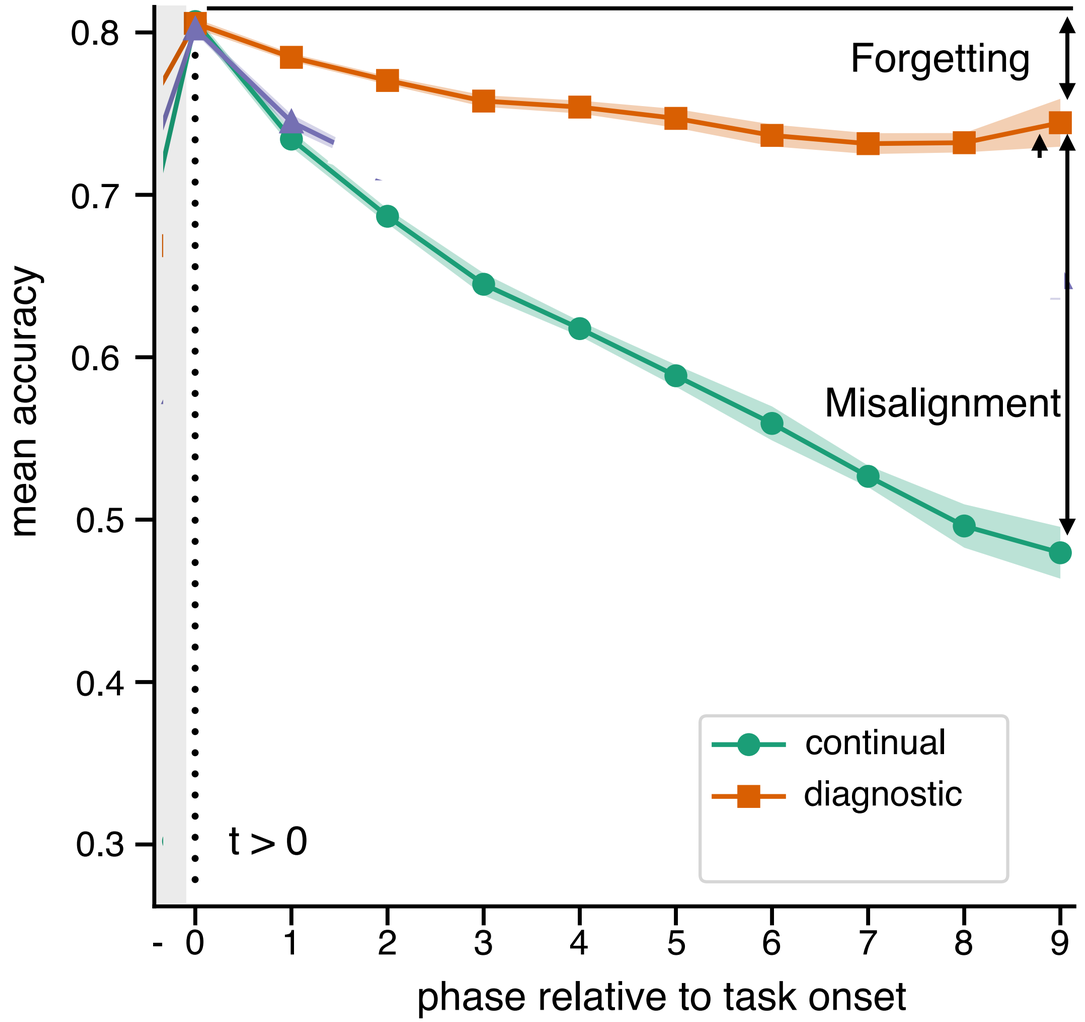


Anthes et al. 2023; Pashakhanloo & Koulakov, 2023

Theory instantiation & expansion: Continual learning in ANNs

CIFAR-100 split into 10 tasks; train sequentially





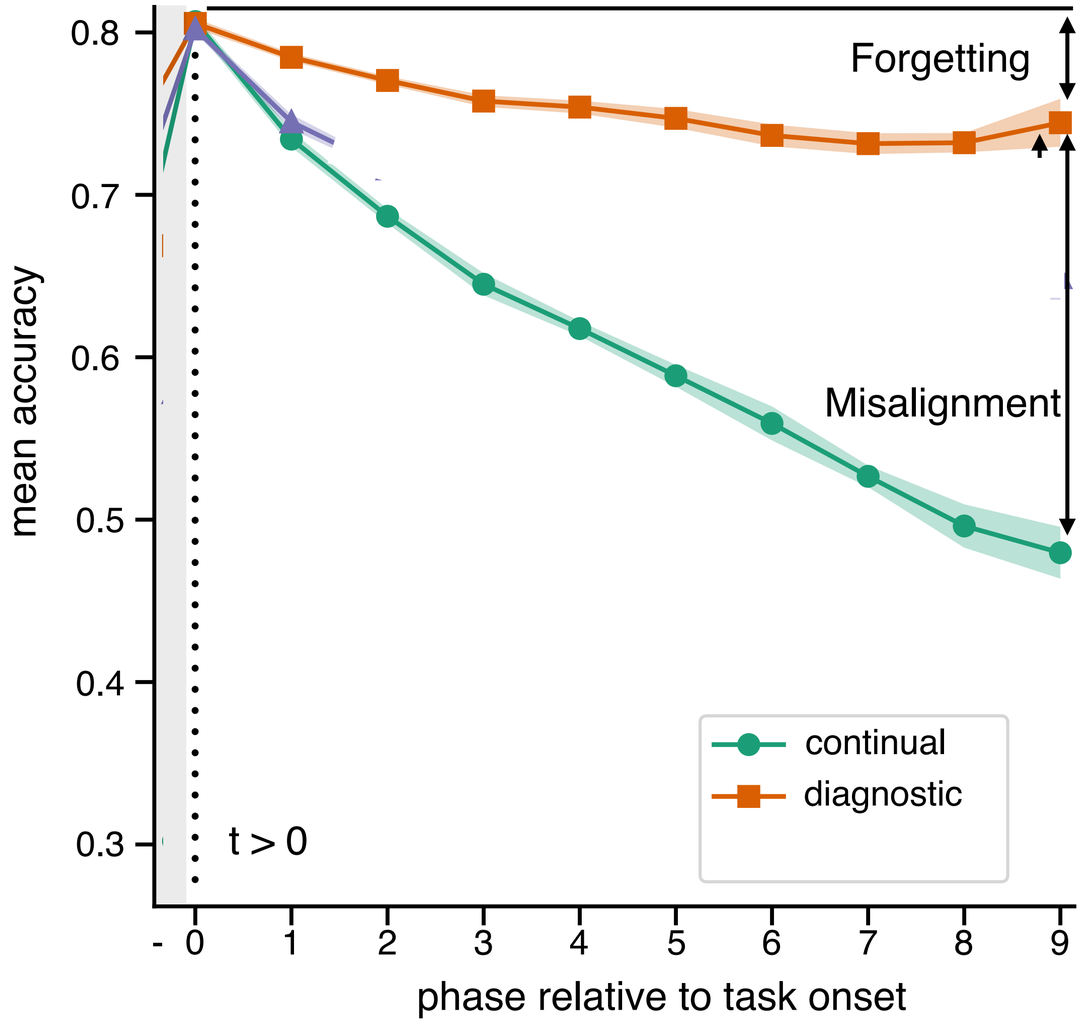
Anthes et al. 2023; Pashakhanloo & Koulakov, 2023

Theory instantiation & expansion: Continual learning in ANNs

CIFAR-100 split into 10 tasks; train sequentially



Learning indeed leads to representational drift!



Anthes et al. 2023; Pashakhanloo & Koulakov, 2023

Theory instantiation & expansion: Continual learning in ANNs

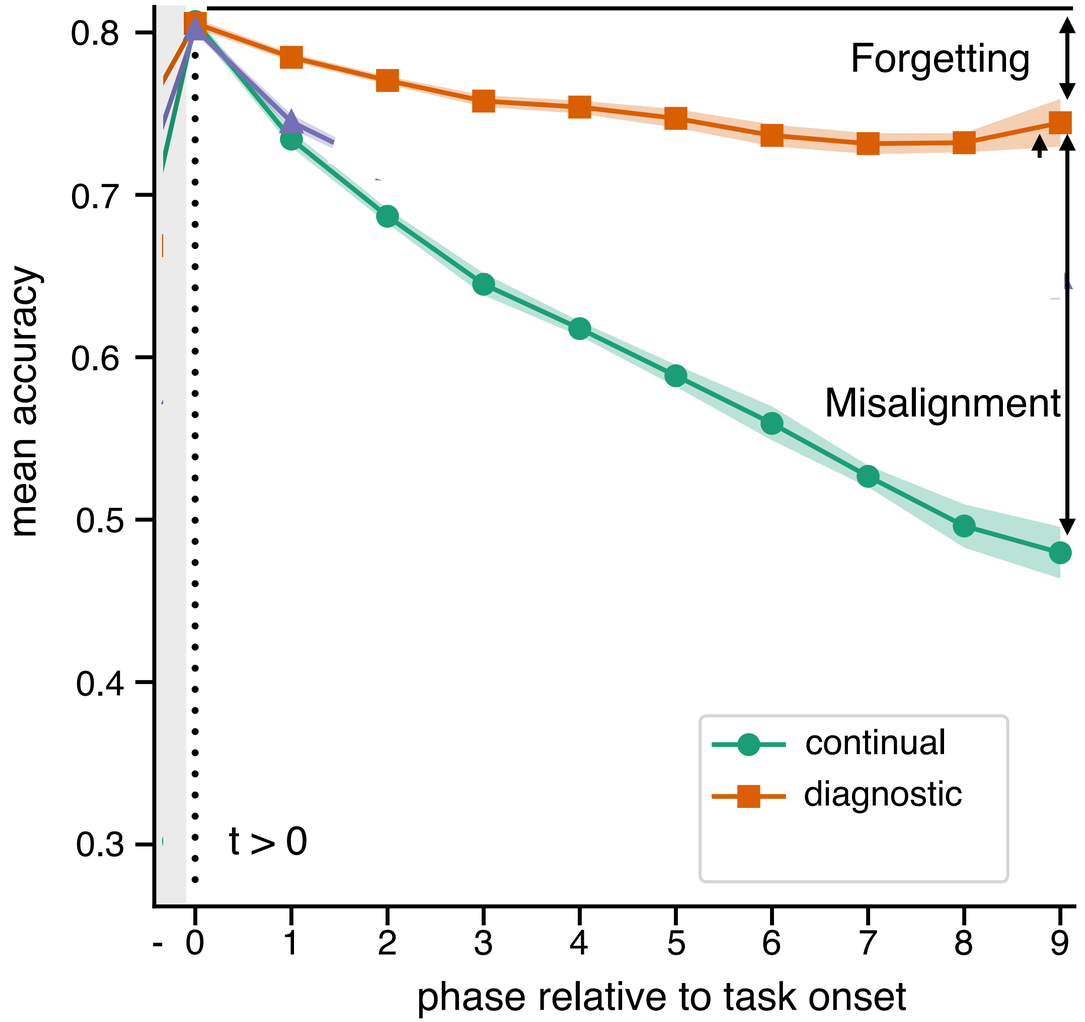
CIFAR-100 split into 10 tasks; train sequentially



Learning indeed leads to representational drift!

Two questions:

1. Is drift inevitable in learning systems?
2. How can the brain maintain stable behavioural readout?



Anthes et al. 2023; Pashakhanloo & Koulakov, 2023

Taming drift as a continual learning solution

Consider a 1-hidden-layer linear NN: $o^1 = W_R^1 W_H x^1$

Anthes et al. CoLLAs 2024; Anthes et al. CCN 2024

Taming drift as a continual learning solution

Consider a 1-hidden-layer linear NN: $o^1 = W_R^1 W_H x^1$

$W_R^1(h + A\Delta h)$

For any gradient ΔW_H on task N, we want to ensure: $o^1 = W_R^1(W_H + A\Delta W_H)x^1 \implies W_R^1 A = 0$

Anthes et al. CoLLAs 2024; Anthes et al. CCN 2024

Taming drift as a continual learning solution

Consider a 1-hidden-layer linear NN: $o^1 = W_R^1 W_H x^1$

$W_R^1(h + A\Delta h)$

For any gradient ΔW_H on task N, we want to ensure: $o^1 = W_R^1(W_H + A\Delta W_H)x^1 \implies W_R^1 A = 0$

$A = NN^T, I = CC^T + NN^T$

Anthes et al. CoLLAs 2024; Anthes et al. CCN 2024

Taming drift as a continual learning solution

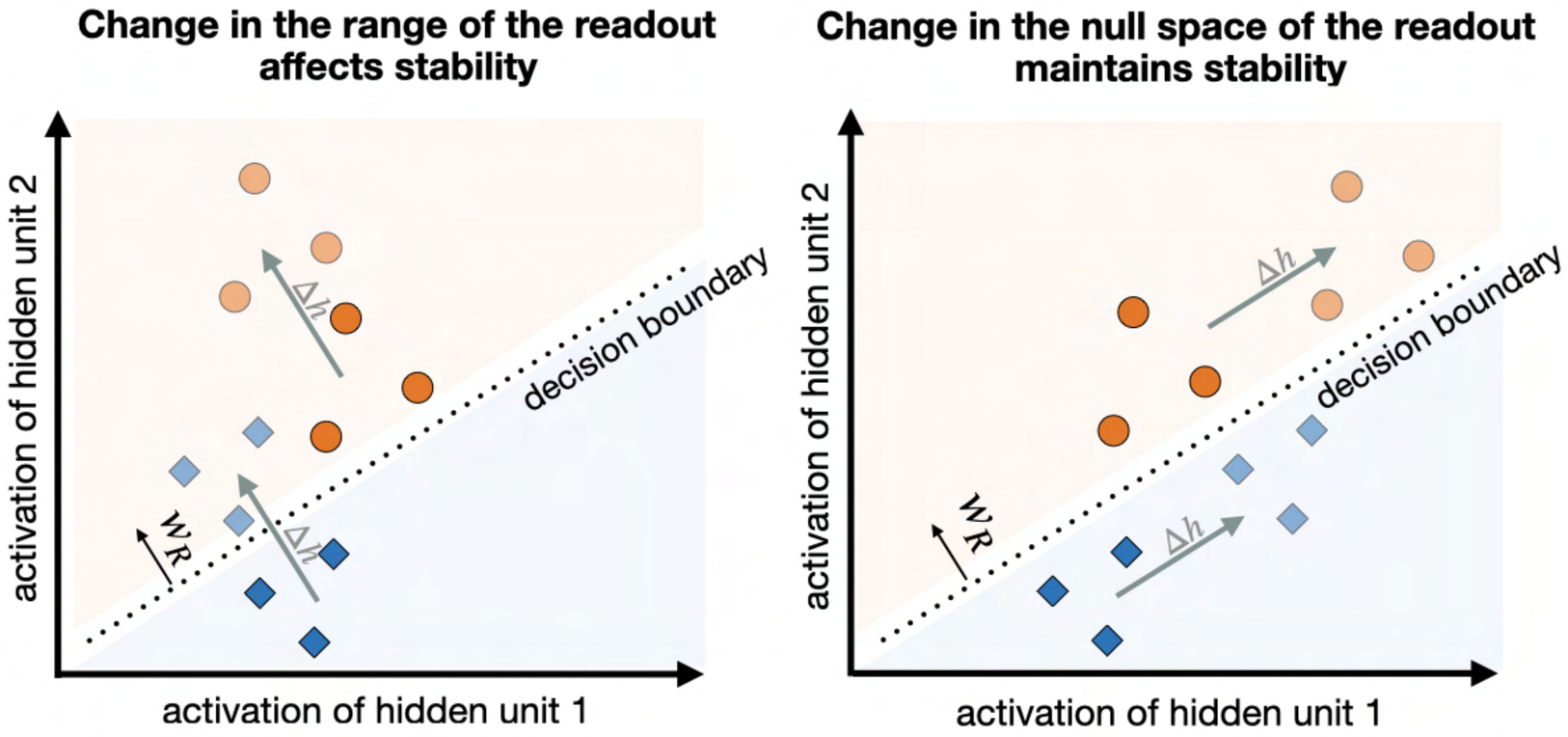
Consider a 1-hidden-layer linear NN: $o^1 = W_R^1 W_H x^1$

$W_R^1(h + A\Delta h)$

For any gradient ΔW_H on task N, we want to ensure: $o^1 = W_R^1(W_H + A\Delta W_H)x^1 \implies W_R^1 A = 0$

i.e., no representational changes in the range of W_R^1 but yes in null!

$A = NN^T, I = CC^T + NN^T$



Anthes et al. CoLLAs 2024; Anthes et al. CCN 2024

Taming drift as a continual learning solution

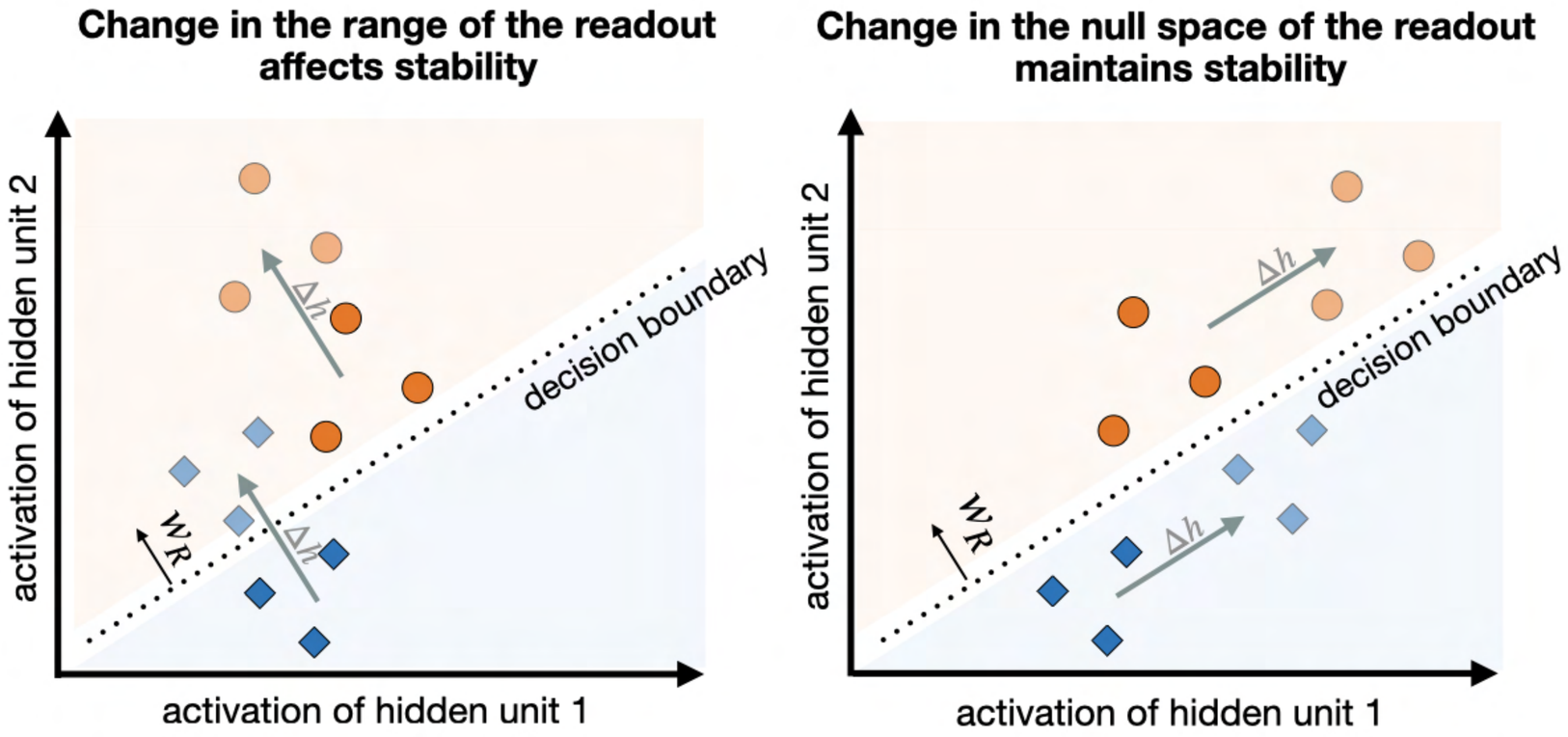
Consider a 1-hidden-layer linear NN: $o^1 = W_R^1 W_H x^1$

$W_R^1(h + A\Delta h)$

For any gradient ΔW_H on task N, we want to ensure: $o^1 = W_R^1(W_H + A\Delta W_H)x^1 \implies W_R^1 A = 0$

i.e., no representational changes in the range of W_R^1 but yes in null!

$A = NN^T, I = CC^T + NN^T$



Drift is essential, best directed into the nullspace of previous tasks' readouts!

Anthes et al. CoLLAs 2024; Anthes et al. CCN 2024

Taming drift as a continual learning solution

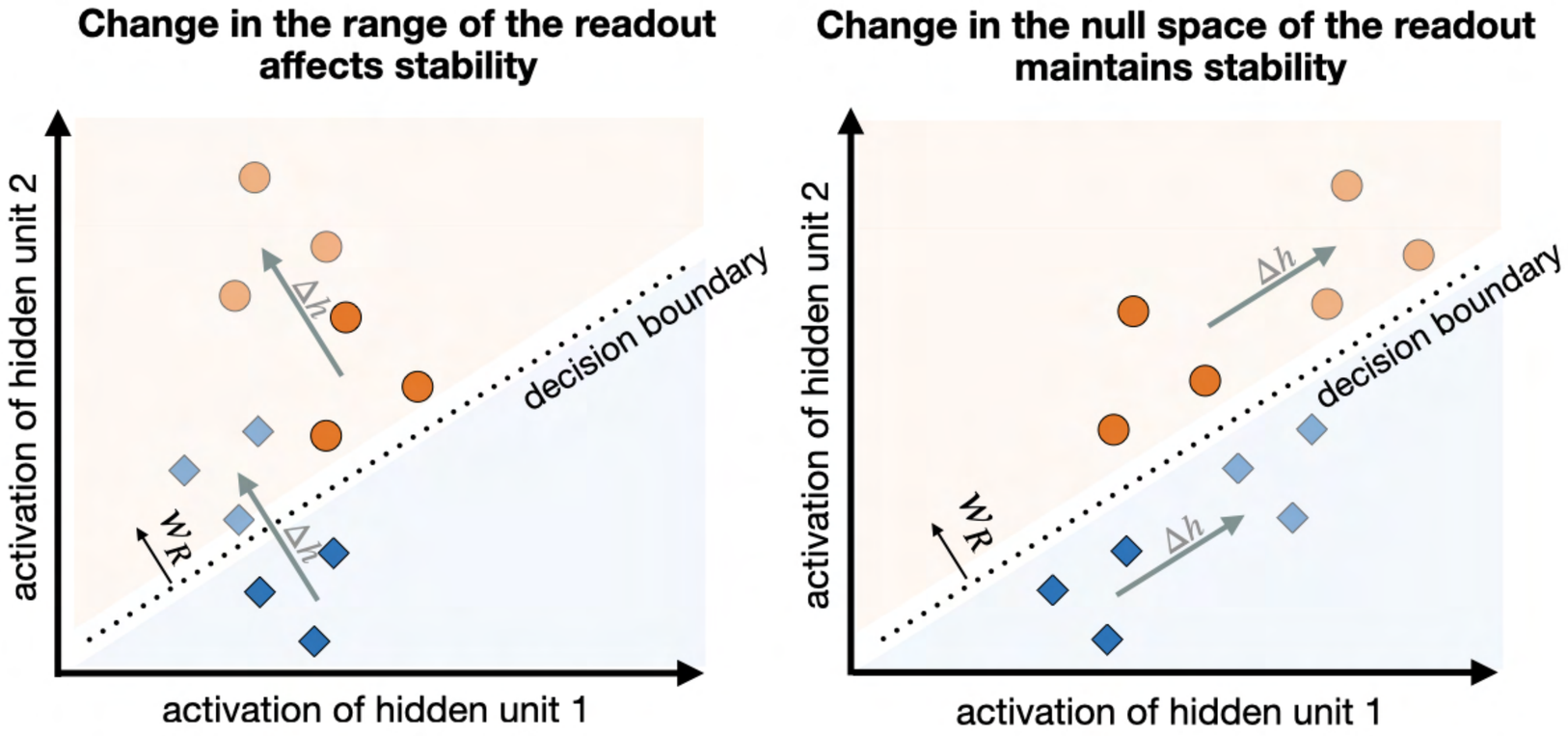
Consider a 1-hidden-layer linear NN: $o^1 = W_R^1 W_H x^1$

$W_R^1(h + A\Delta h)$

For any gradient ΔW_H on task N, we want to ensure: $o^1 = W_R^1(W_H + A\Delta W_H)x^1 \implies W_R^1 A = 0$

i.e., no representational changes in the range of W_R^1 but yes in null!

$A = NN^T, I = CC^T + NN^T$



Drift is essential, best directed into the nullspace of previous tasks' readouts!

CL algorithms implicitly do this!

Anthes et al. CoLLAs 2024; Anthes et al. CCN 2024

Taming drift as a continual learning solution

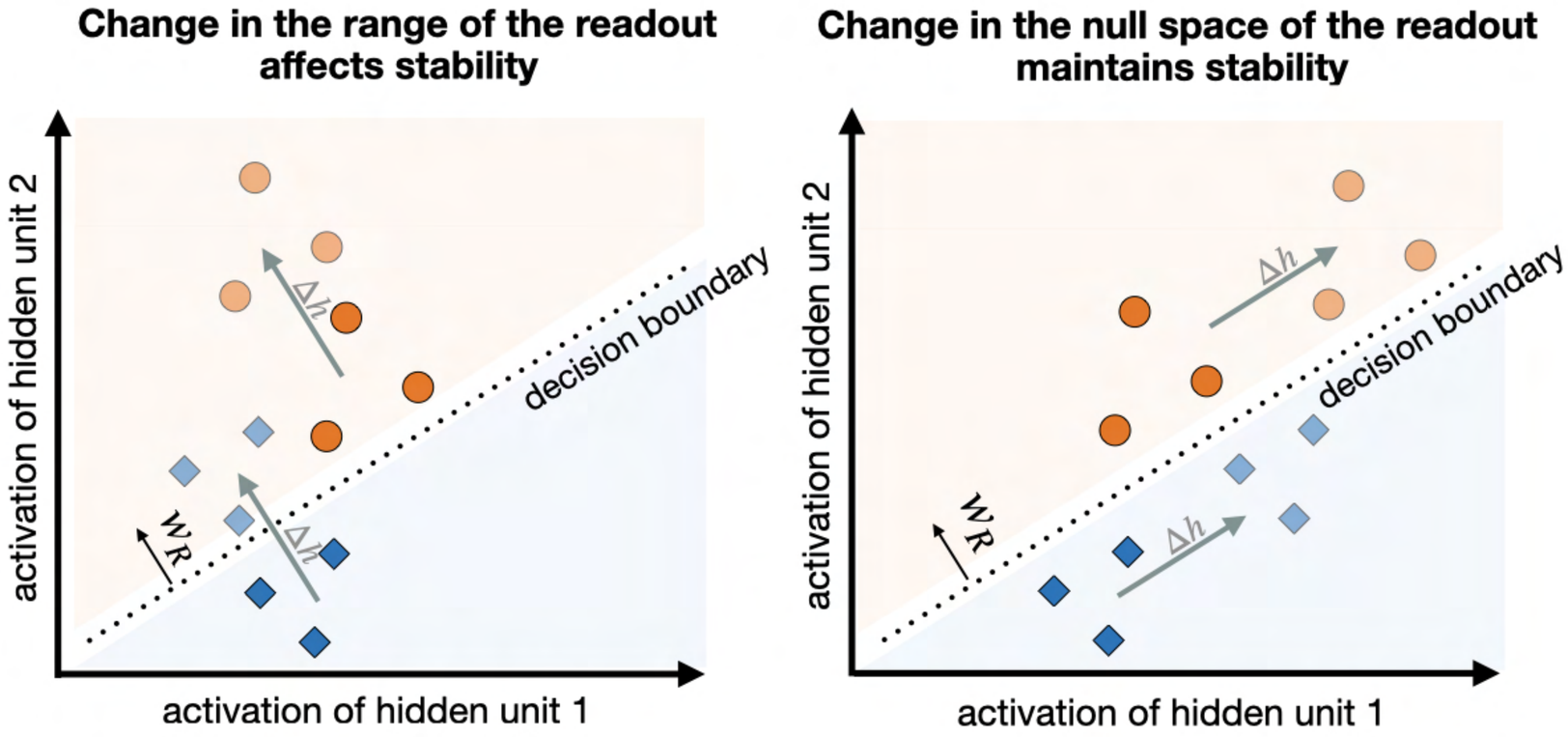
Consider a 1-hidden-layer linear NN: $o^1 = W_R^1 W_H x^1$

$W_R^1(h + A\Delta h)$

For any gradient ΔW_H on task N, we want to ensure: $o^1 = W_R^1(W_H + A\Delta W_H)x^1 \implies W_R^1 A = 0$

i.e., no representational changes in the range of W_R^1 but yes in null!

$A = NN^T, I = CC^T + NN^T$



Drift is essential, best directed into the nullspace of previous tasks' readouts!

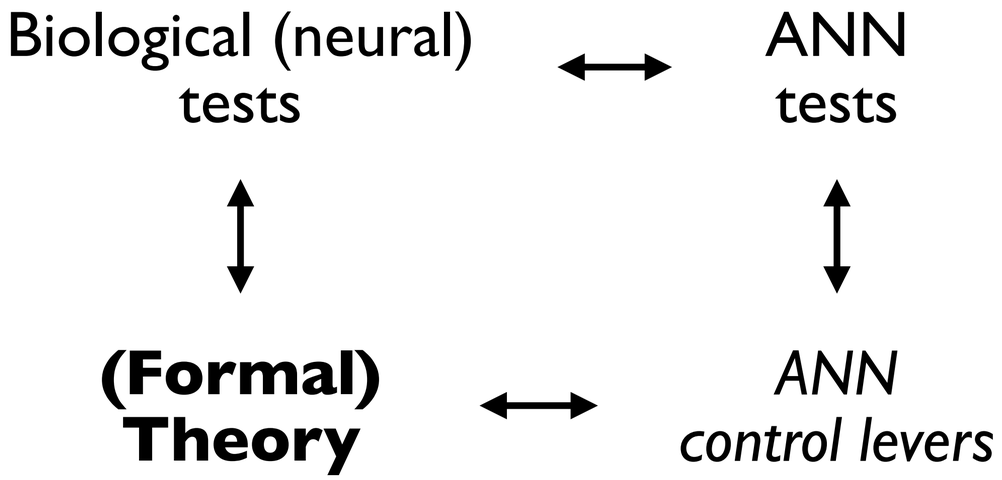
CL algorithms implicitly do this!

Question:

<u>Can we find drift restricted to readout nullspace in brains?</u>

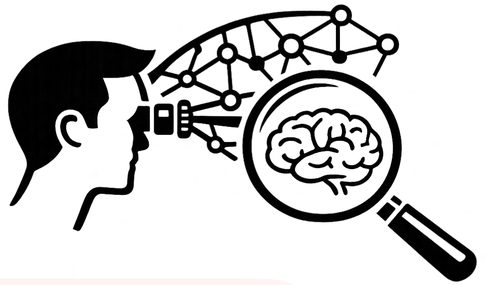
<u>Anthes et al. CoLLAs 2024</u>; <u>Anthes et al. CCN 2024</u>

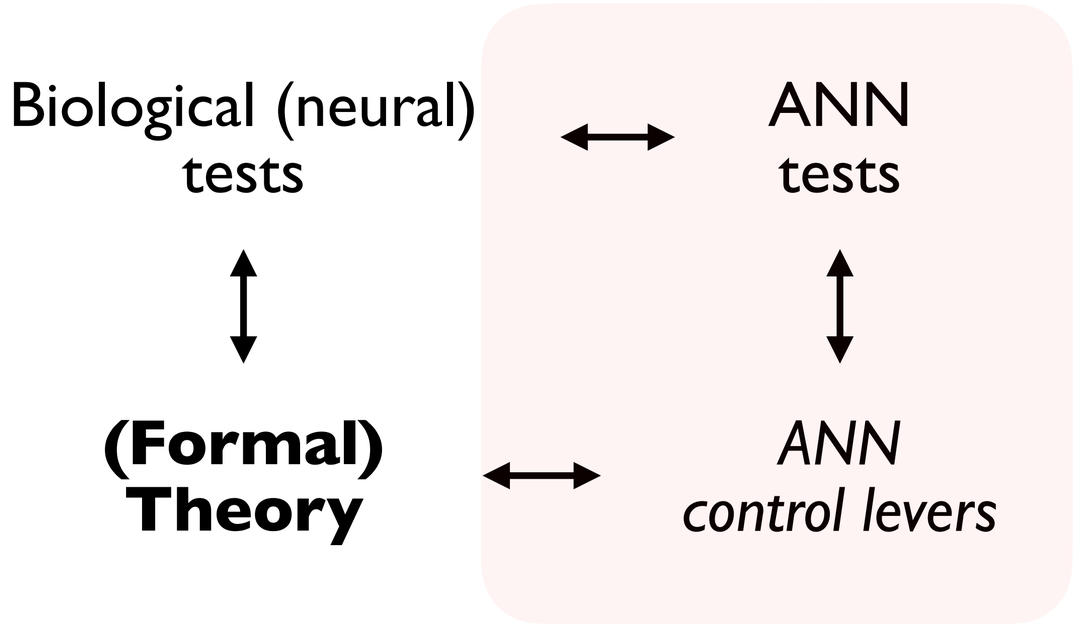
Conclusion



Yamins & DiCarlo, 2016; Kell & McDermott, 2019; Whittington et al. 2020; Ricci et al. 2021

Conclusion

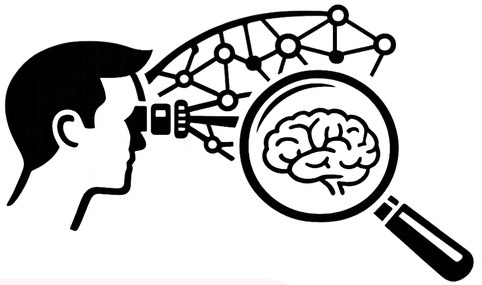


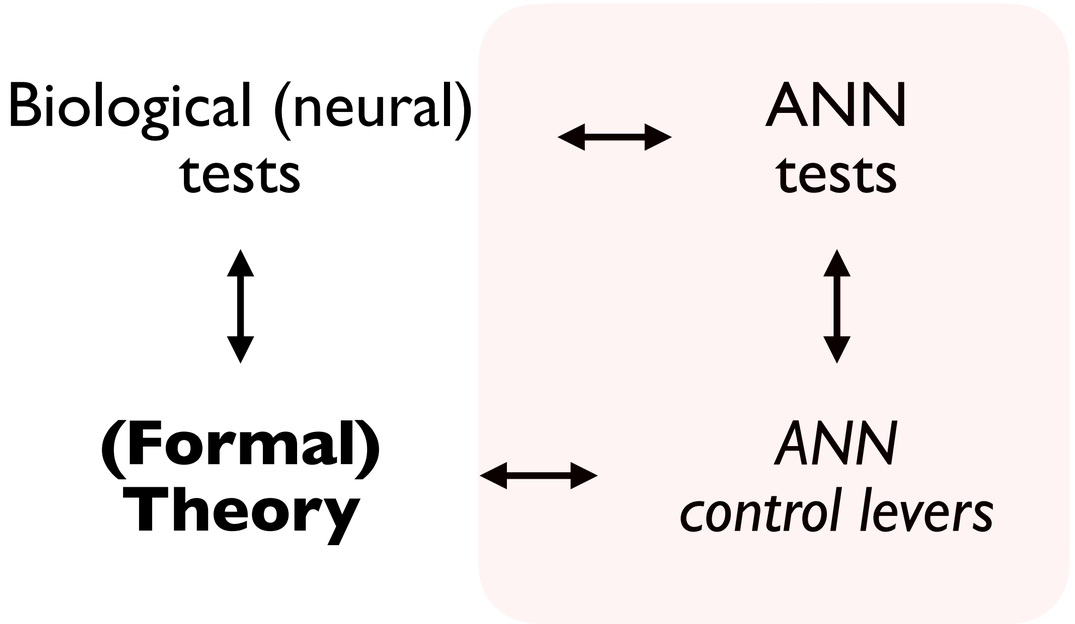


Yamins & DiCarlo, 2016; Kell & McDermott, 2019; Whittington et al. 2020; Ricci et al. 2021

Conclusion

Don't mind the "AGI",
there's a **lot** more to be done!



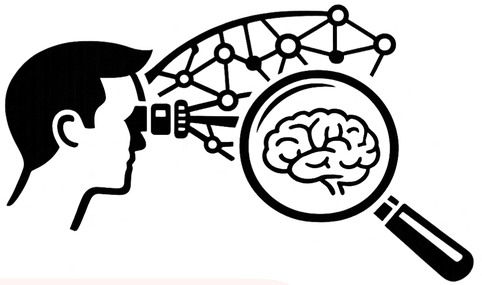


Yamins & DiCarlo, 2016; Kell & McDermott, 2019; Whittington et al. 2020; Ricci et al. 2021

Conclusion

Don't mind the "AGI",
there's a **lot** more to be done!

How much do we neuroscientists
truly know the brain?
VQA? How do we do it?



Biological (neural) tests ↔ ANN tests

↕ ↕

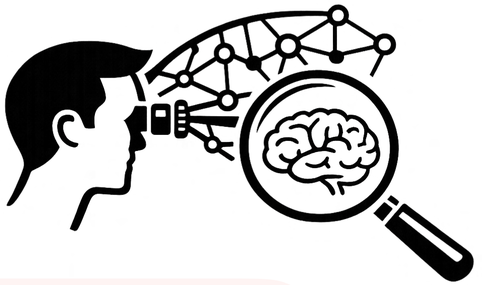
(Formal) Theory ↔ *ANN control levers*

Yamins & DiCarlo, 2016; Kell & McDermott, 2019; Whittington et al. 2020; Ricci et al. 2021

Conclusion

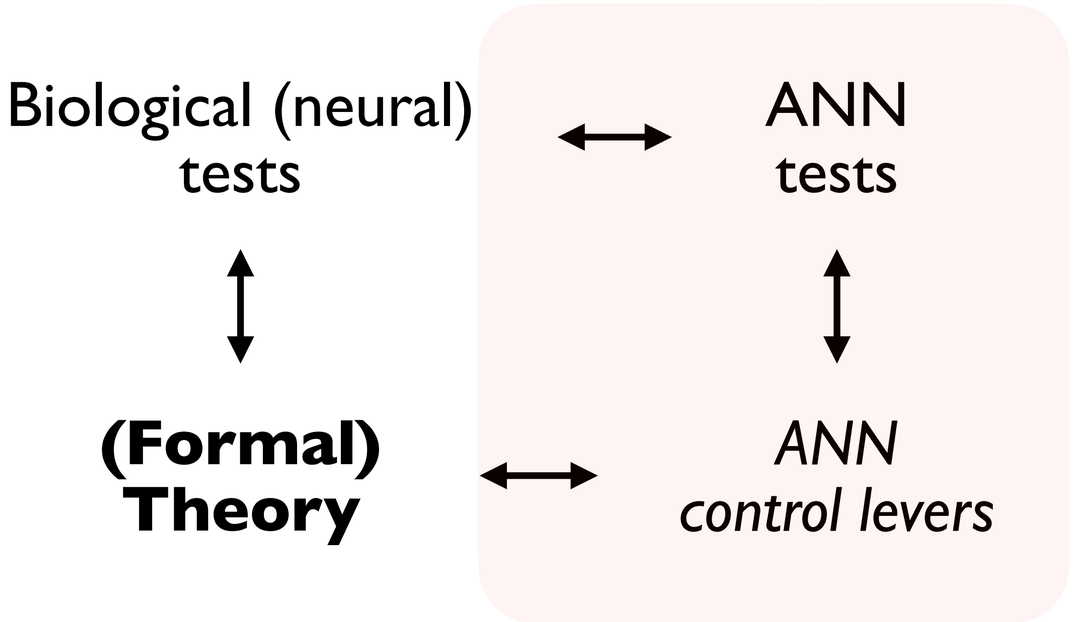
Don't mind the "AGI",
there's a **lot** more to be done!

How much do we neuroscientists truly know the brain?
VQA? How do we do it?



Leveraging the new toys:

1. Simulate visual cortex devo
    1. Ego-video + Gaze (JEPA)
2. Simulate memory system coordination (Long-horizon AI)
3. Simulate learning across specialised systems (Multi-agent orchestration)



Yamins & DiCarlo, 2016; Kell & McDermott, 2019; Whittington et al. 2020; Ricci et al. 2021

Supplementary

What is Cognition?

See more in course slides: https://doi.org/10.6084/m9.figshare.23537094.v2

[1]https://en.wikipedia.org/wiki/Cognition; [2]https://www.gutenberg.org/files/57628/57628-h/57628-h.htm; [3]https://plato.stanford.edu/entries/cognition-animal/;
[4]https://www.scientificamerican.com/article/the-mind-of-an-octopus/; [5]Calvo et al., Annals of Botany, 2019; [6]Clark, BBS 2013

What is Cognition?

"The ***mental action or process*** of acquiring knowledge and understanding through thought, experience, and the senses"[1] (-> *is a process cognitive?*)

Origin of the term:
"cognoscere" (Latin; to get to know) -> "cognitio" (Latin; knowledge/perception)

See more in course slides: https://doi.org/10.6084/m9.figshare.23537094.v2

[1]https://en.wikipedia.org/wiki/Cognition; [2]https://www.gutenberg.org/files/57628/57628-h/57628-h.htm; [3]https://plato.stanford.edu/entries/cognition-animal/;
[4]https://www.scientificamerican.com/article/the-mind-of-an-octopus/; [5]Calvo et al., Annals of Botany, 2019; [6]Clark, BBS 2013

What is Cognition?

"The ***mental action or process*** of acquiring knowledge and understanding through thought, experience, and the senses"[1] (-> *is a process cognitive?*)

Origin of the term:
"cognoscere" (Latin; to get to know) -> "cognitio" (Latin; knowledge/perception)

… seems like a broad *philosophical* concept[2] that might be best understood by looking at example processes covered under it; *is a process cognitive?*

See more in course slides: https://doi.org/10.6084/m9.figshare.23537094.v2

[1]https://en.wikipedia.org/wiki/Cognition; [2]https://www.gutenberg.org/files/57628/57628-h/57628-h.htm; [3]https://plato.stanford.edu/entries/cognition-animal/;
[4]https://www.scientificamerican.com/article/the-mind-of-an-octopus/; [5]Calvo et al., Annals of Botany, 2019; [6]Clark, BBS 2013

What is Cognition?

"The ***mental action or process*** of acquiring knowledge and understanding through thought, experience, and the senses"[1] (-> *is a process cognitive?*)

Origin of the term:
"cognoscere" (Latin; to get to know) -> "cognitio" (Latin; knowledge/perception)

… seems like a broad *philosophical* concept[2] that might be best understood by looking at example processes covered under it; *is a process cognitive?*

… as opposed to autonomic processes (heart rate, breathing), reflexes, motor execution, and earliest stages of sensory perception; so *high-level*

See more in course slides: https://doi.org/10.6084/m9.figshare.23537094.v2

[1]https://en.wikipedia.org/wiki/Cognition; [2]https://www.gutenberg.org/files/57628/57628-h/57628-h.htm; [3]https://plato.stanford.edu/entries/cognition-animal/;
[4]https://www.scientificamerican.com/article/the-mind-of-an-octopus/; [5]Calvo et al., Annals of Botany, 2019; [6]Clark, BBS 2013

What is Cognition?

"The ***mental action or process*** of acquiring knowledge and understanding through thought, experience, and the senses"[1] (-> *is a process cognitive?*)

Origin of the term:
"cognoscere" (Latin; to get to know) -> "cognitio" (Latin; knowledge/perception)

… seems like a broad *philosophical* concept[2] that might be best understood by looking at example processes covered under it; *is a process cognitive?*

… as opposed to autonomic processes (heart rate, breathing), reflexes, motor execution, and earliest stages of sensory perception; so *high-level*

… animals have cognition[3], so does the octopus[4], and… plants[5]?!

See more in course slides: https://doi.org/10.6084/m9.figshare.23537094.v2

[1]https://en.wikipedia.org/wiki/Cognition; [2]https://www.gutenberg.org/files/57628/57628-h/57628-h.htm; [3]https://plato.stanford.edu/entries/cognition-animal/;
[4]https://www.scientificamerican.com/article/the-mind-of-an-octopus/; [5]Calvo et al., Annals of Botany, 2019; [6]Clark, BBS 2013

What is Cognition?

"The ***mental action or process*** of acquiring knowledge and understanding through thought, experience, and the senses"[1] (-> *is a process cognitive?*)

Origin of the term:
"cognoscere" (Latin; to get to know) -> "cognitio" (Latin; knowledge/perception)

… seems like a broad *philosophical* concept[2] that might be best understood by looking at example processes covered under it; *is a process cognitive?*

… as opposed to autonomic processes (heart rate, breathing), reflexes, motor execution, and earliest stages of sensory perception; so *high-level*

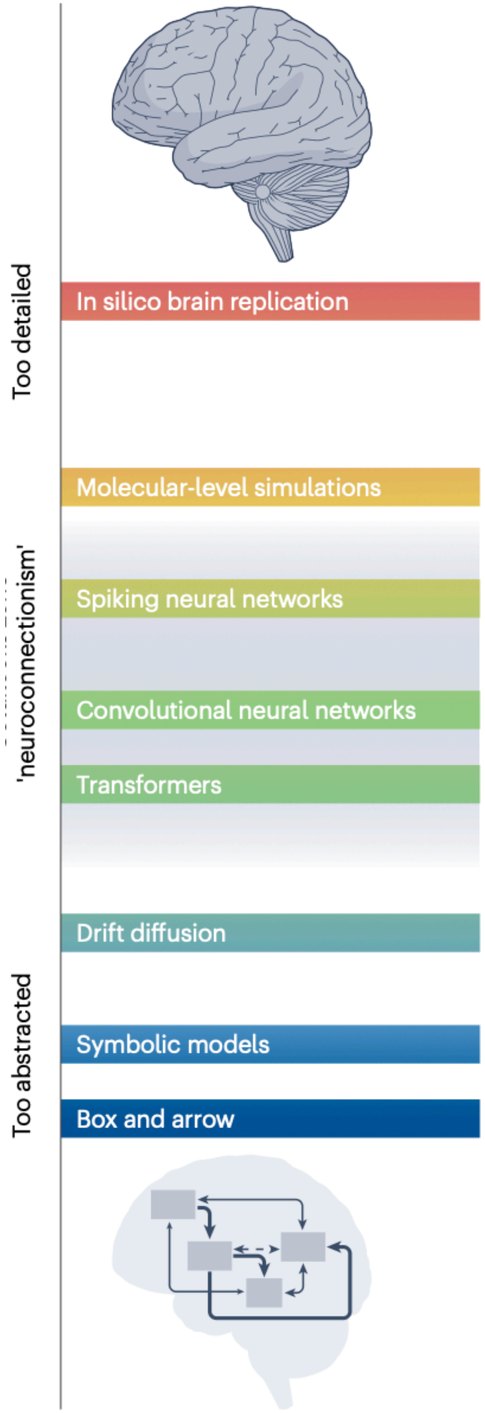
… animals have cognition[3], so does the octopus[4], and… plants[5]?!

… **acquiring, adapting, and rolling-out predictive world models is key**[6]?

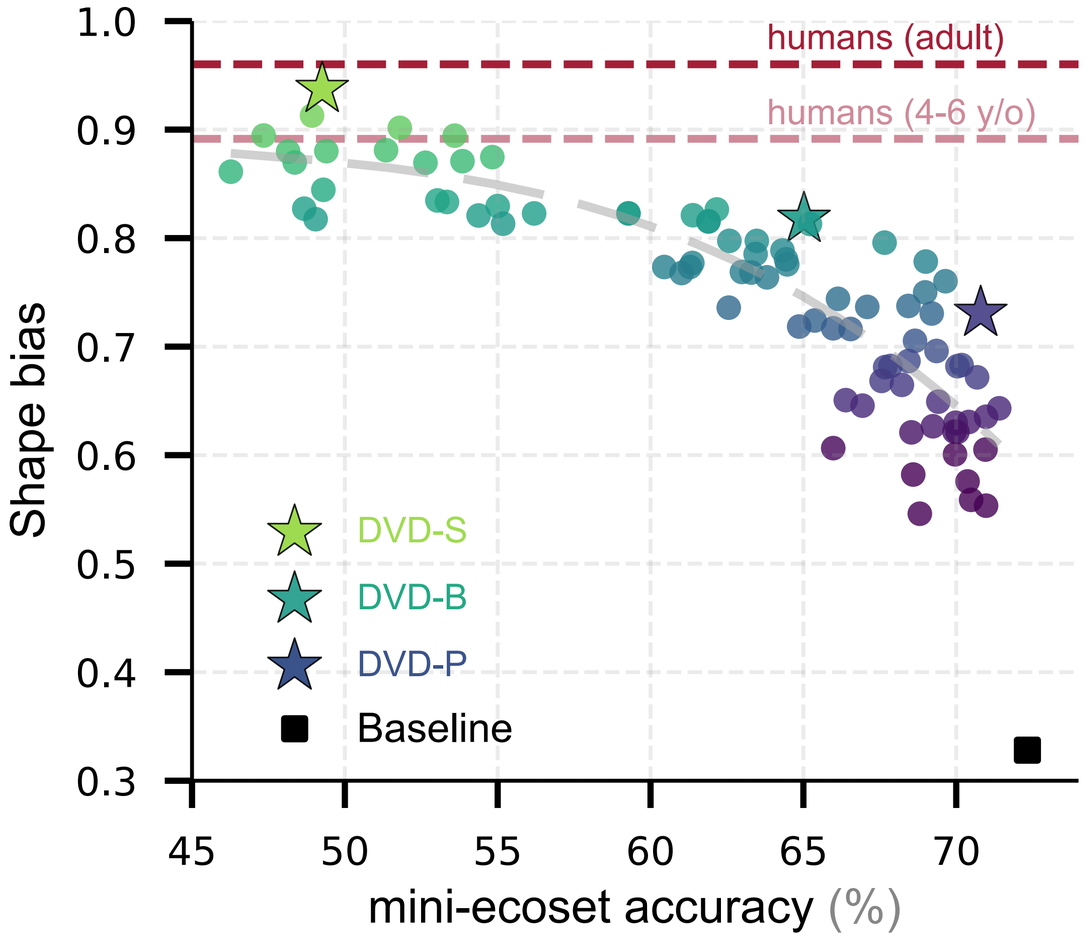
See more in course slides: https://doi.org/10.6084/m9.figshare.23537094.v2

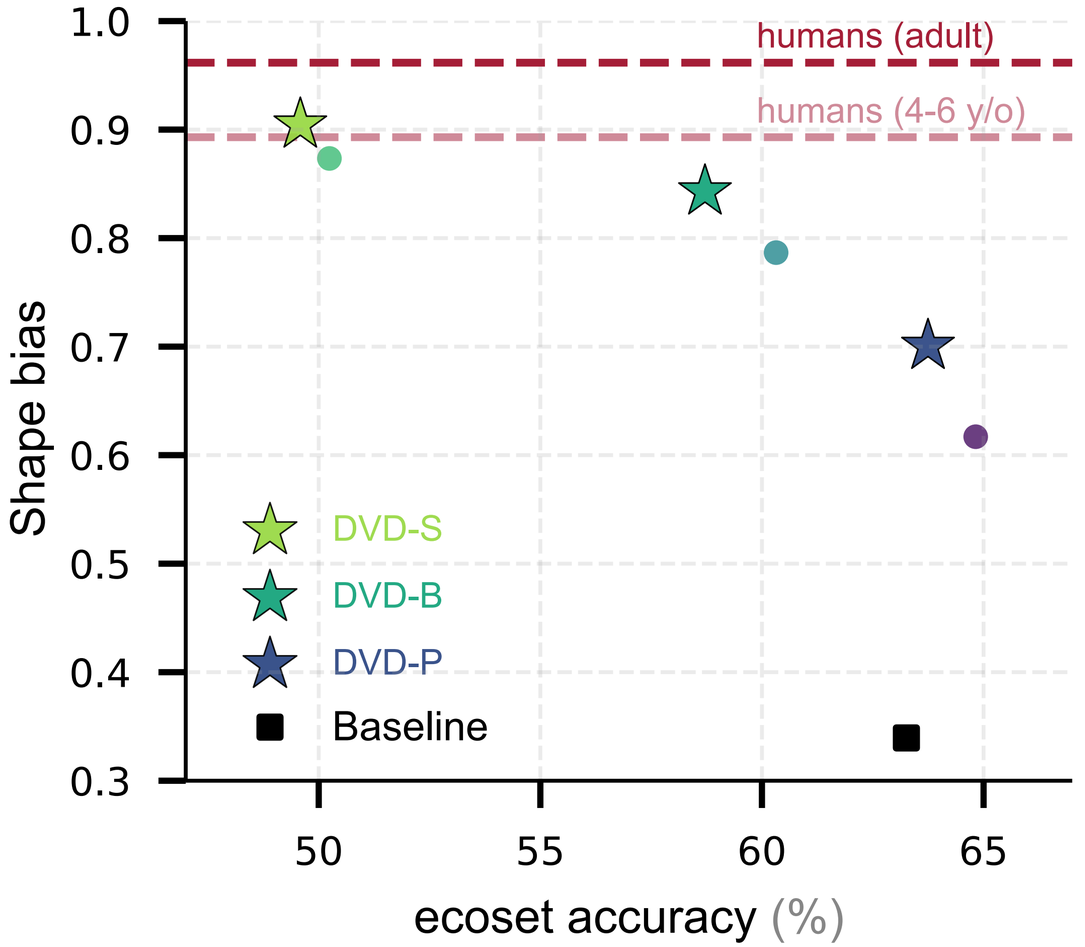
[1]https://en.wikipedia.org/wiki/Cognition; [2]https://www.gutenberg.org/files/57628/57628-h/57628-h.htm; [3]https://plato.stanford.edu/entries/cognition-animal/;
[4]https://www.scientificamerican.com/article/the-mind-of-an-octopus/; [5]Calvo et al., Annals of Botany, 2019; [6]Clark, BBS 2013

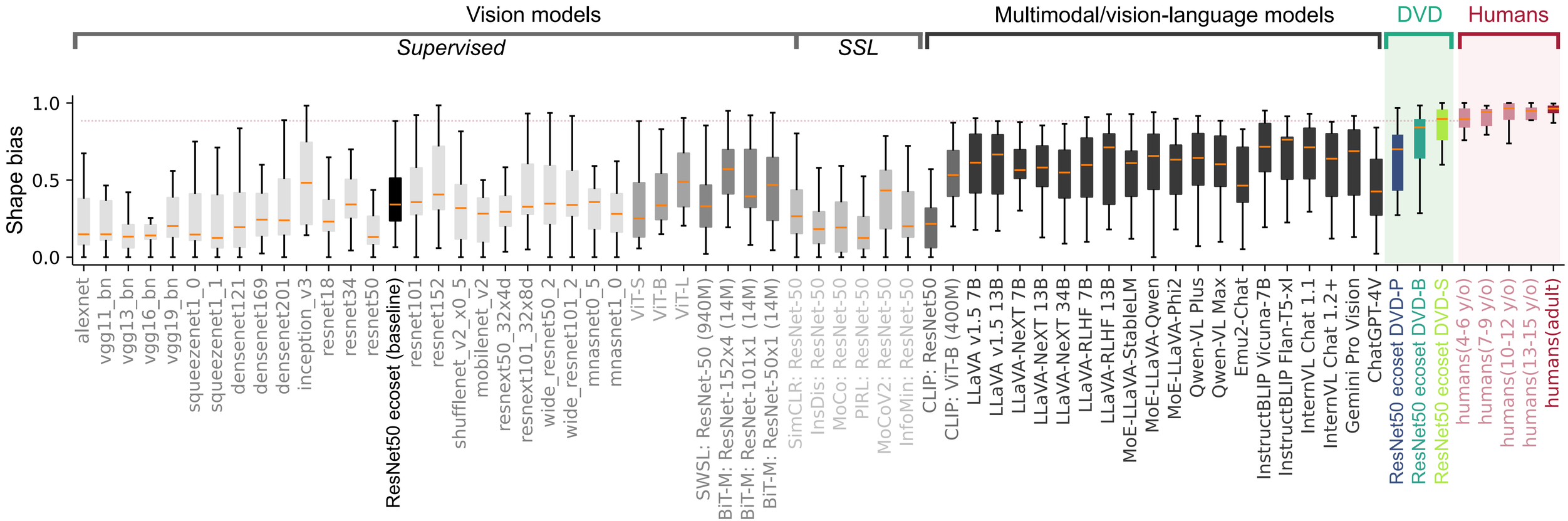
Goldilocks zone of biological abstraction



b Shape bias - accuracy tradeoff



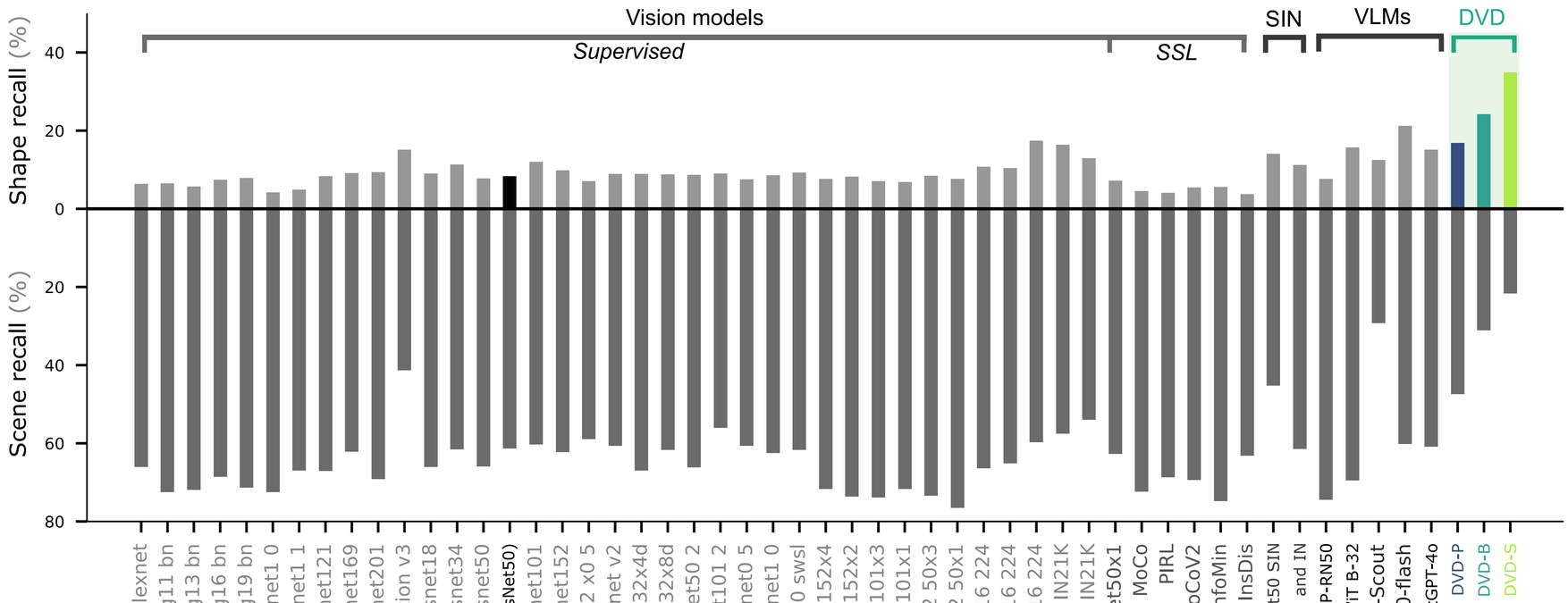


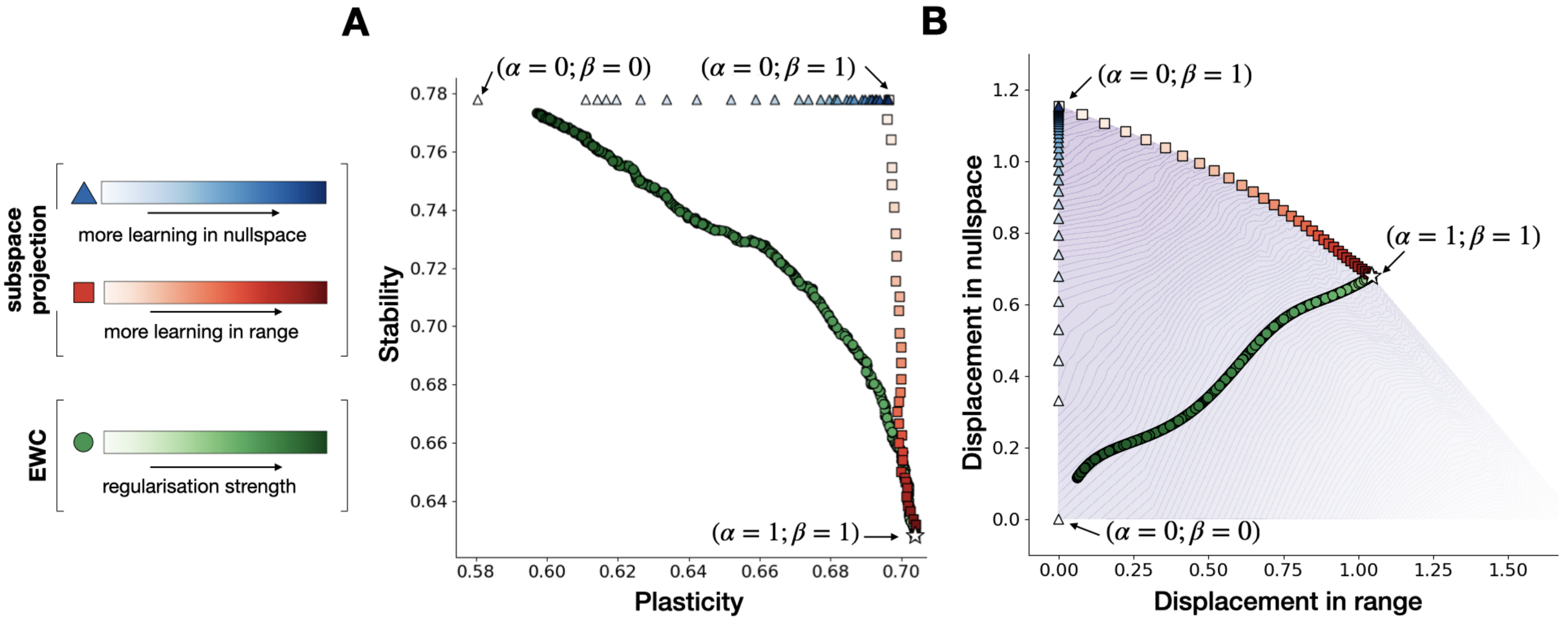

Vision models
Multimodal/vision-language models
DVD
Humans
Supervised
SSL
Shape bias
1.0
0.5
0.0
alexnet
vgg11_bn
vgg13_bn
vgg16_bn
vgg19_bn
squeezenet1_0
squeezenet1_1
densenet121
densenet169
densenet201
inception_v3
resnet18
resnet34
resnet50
ResNet50 ecoset (baseline)
resnet101
resnet152
shufflenet_v2_x0_5
mobilenet_v2
resnext50_32x4d
resnext101_32x8d
wide_resnet50_2
wide_resnet101_2
mnasnet0_5
mnasnet1_0
ViT-S
ViT-B
ViT-L
SWSL: ResNet-50 (940M)
BiT-M: ResNet-152x4 (14M)
BiT-M: ResNet-101x1 (14M)
BiT-M: ResNet-50x1 (14M)
SimCLR: ResNet-50
InsDis: ResNet-50
MoCo: ResNet-50
PIRL: ResNet-50
MoCoV2: ResNet-50
InfoMin: ResNet-50
CLIP: ResNet50
CLIP: ViT-B (400M)
LLaVA v1.5 7B
LLaVA v1.5 13B
LLaVA-NeXT 7B
LLaVA-NeXT 13B
LLaVA-NeXT 34B
LLaVA-RLHF 7B
LLaVA-RLHF 13B
MoE-LLaVA-StableLM
MoE-LLaVA-Qwen
MoE-LLaVA-Phi2
Qwen-VL Plus
Qwen-VL Max
Emu2-Chat
InstructBLIP Vicuna-7B
InstructBLIP Flan-T5-xl
InternVL Chat 1.1
InternVL Chat 1.2+
Gemini Pro Vision
ChatGPT-4V
ResNet50 ecoset DVD-P
ResNet50 ecoset DVD-B
ResNet50 ecoset DVD-S
humans(4-6 y/o)
humans(7-9 y/o)
humans(10-12 y/o)
humans(13-15 y/o)
humans(adult)

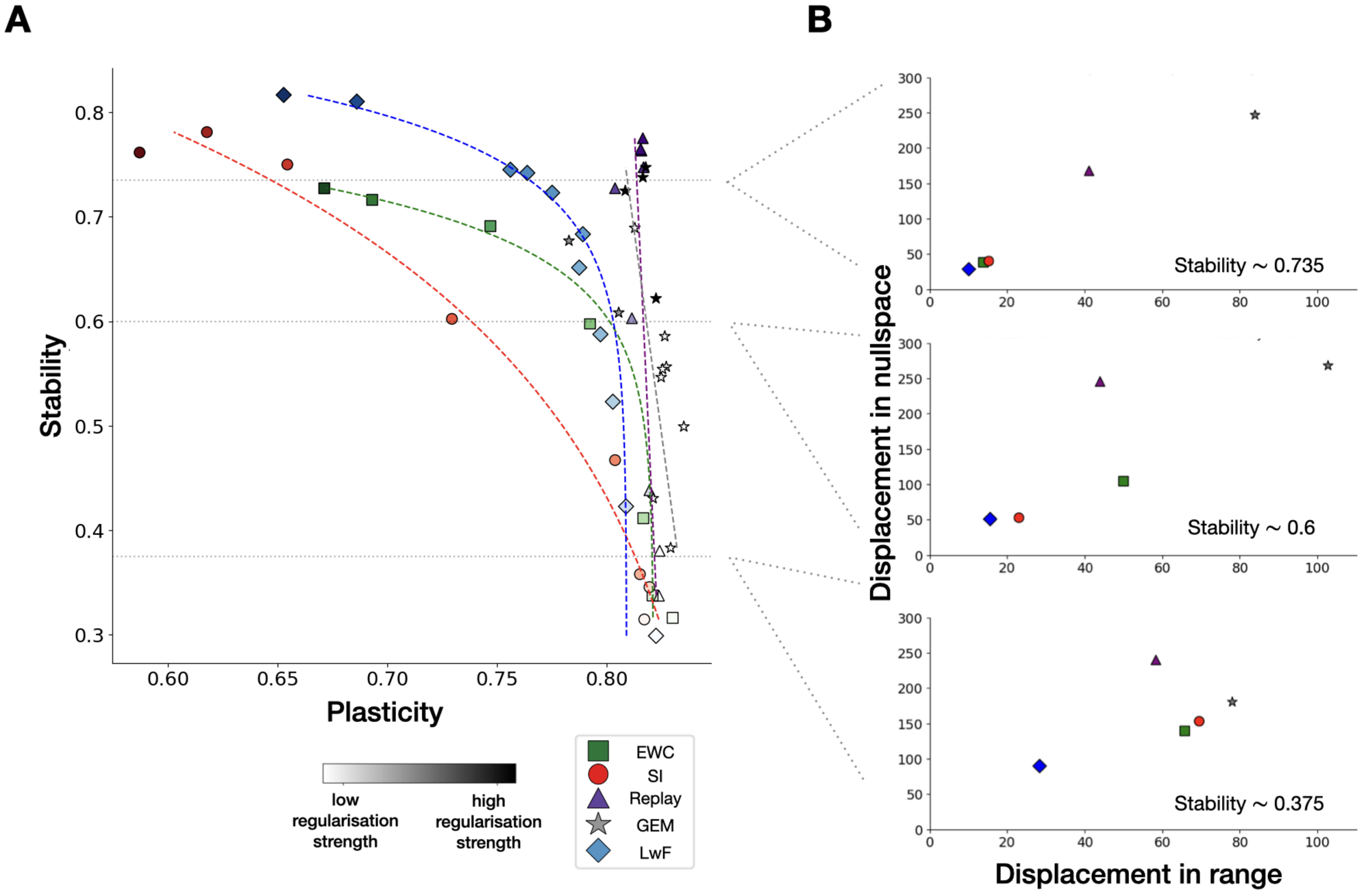
a Examples of abstract shapes hidden in scenes



b Shape recall vs. scene recall benchmark




A
B
subspace projection
more learning in nullspace
more learning in range
EWC
regularisation strength
Stability
Plasticity
$(\alpha = 0; \beta = 0)$
$(\alpha = 0; \beta = 1)$
$(\alpha = 1; \beta = 1)$
0.78
0.76
0.74
0.72
0.70
0.68
0.66
0.64
0.58
0.60
0.62
0.64
0.66
0.68
0.70
Displacement in nullspace
Displacement in range
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50


A
B
Stability
Plasticity
0.8
0.7
0.6
0.5
0.4
0.3
0.60
0.65
0.70
0.75
0.80
low regularisation strength
high regularisation strength
EWC
SI
Replay
GEM
LwF
Displacement in nullspace
Displacement in range
Stability ~ 0.735
Stability ~ 0.6
Stability ~ 0.375
300
250
200
150
100
50
0
0
20
40
60
80
100