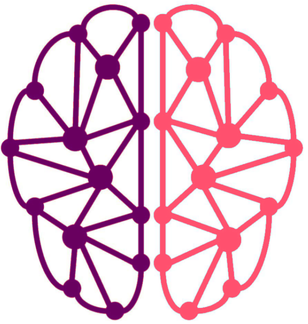
Evaluating alignment between humans and neural network representations in image-based learning tasks

Can Demircan, Tankred Saanum, Leonardo Pettini, Marcel Binz
Blazej M Baczkowski, Christian F Doeller, Mona M Garvert, Eric Schulz



HELMHOLTZ
MUNICH

Human & neural network alignment

Human & neural network alignment

- Humans have rich sensory representations and can generalise effectively.

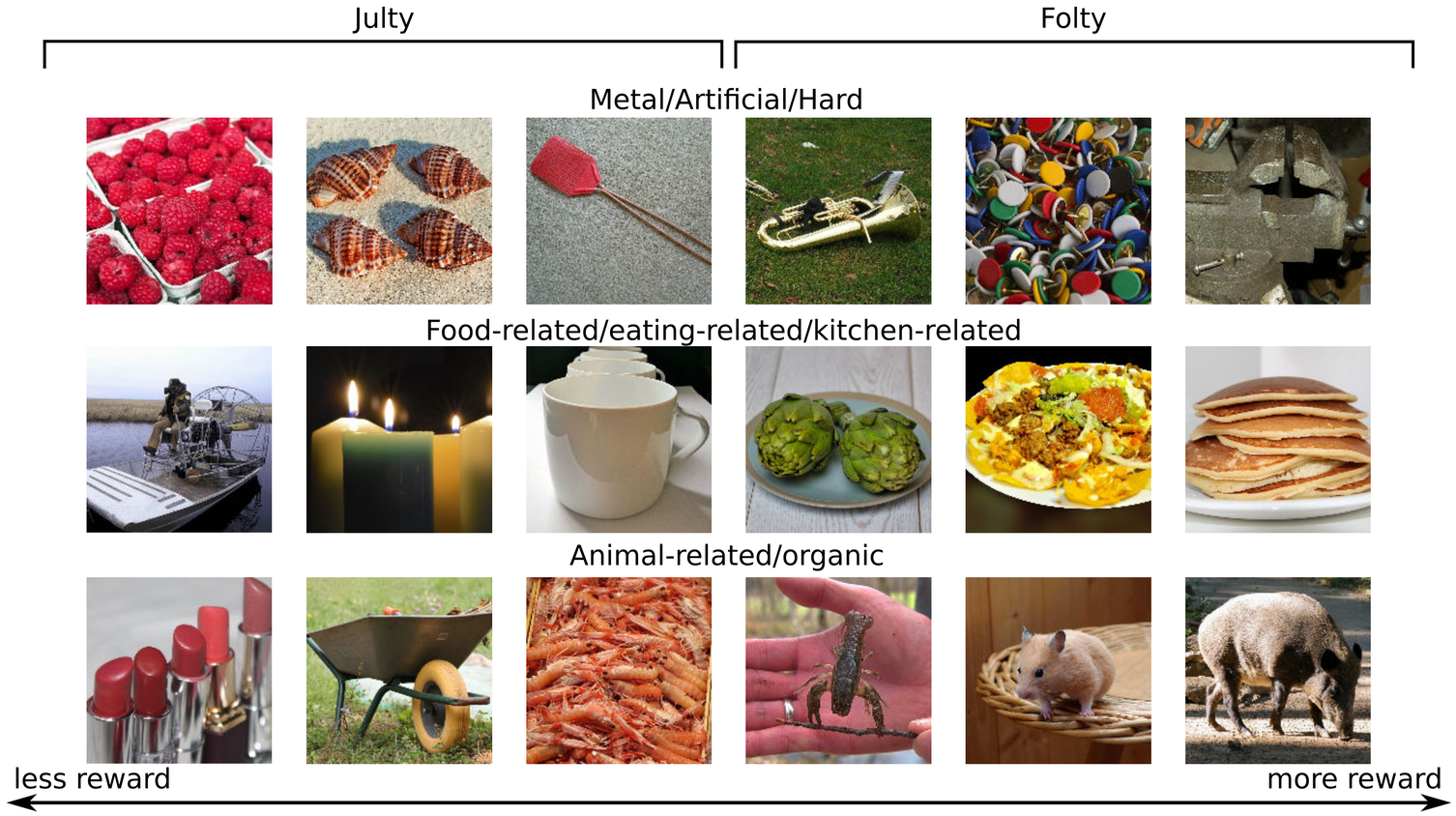
Human & neural network alignment

- Humans have rich sensory representations and can generalise effectively.
- What determines whether a neural network generalises like a human?

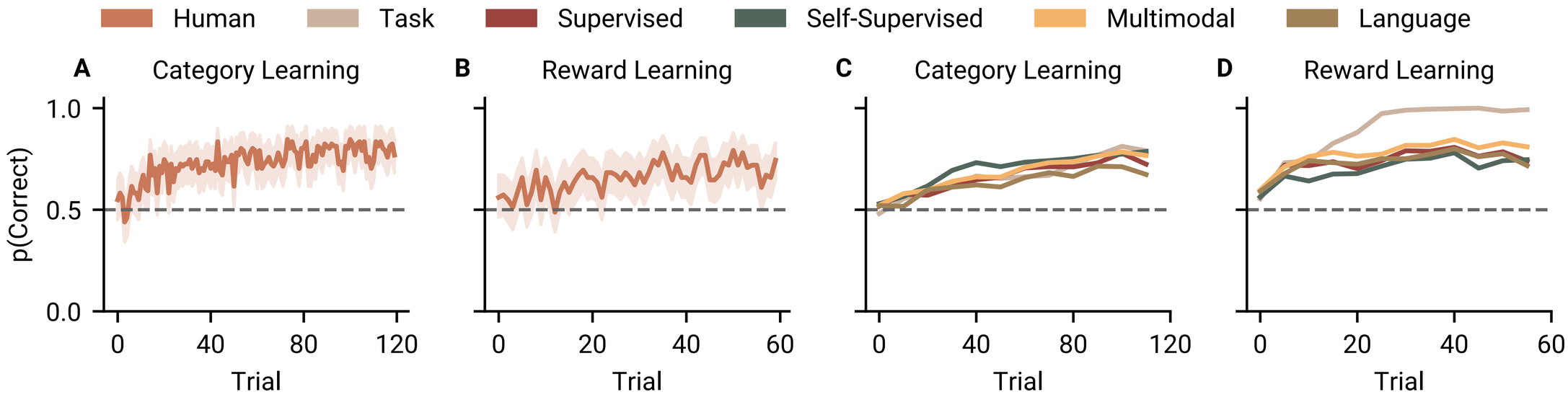
A Experiment 1: Category Learning

B Experiment 2: Reward Learning

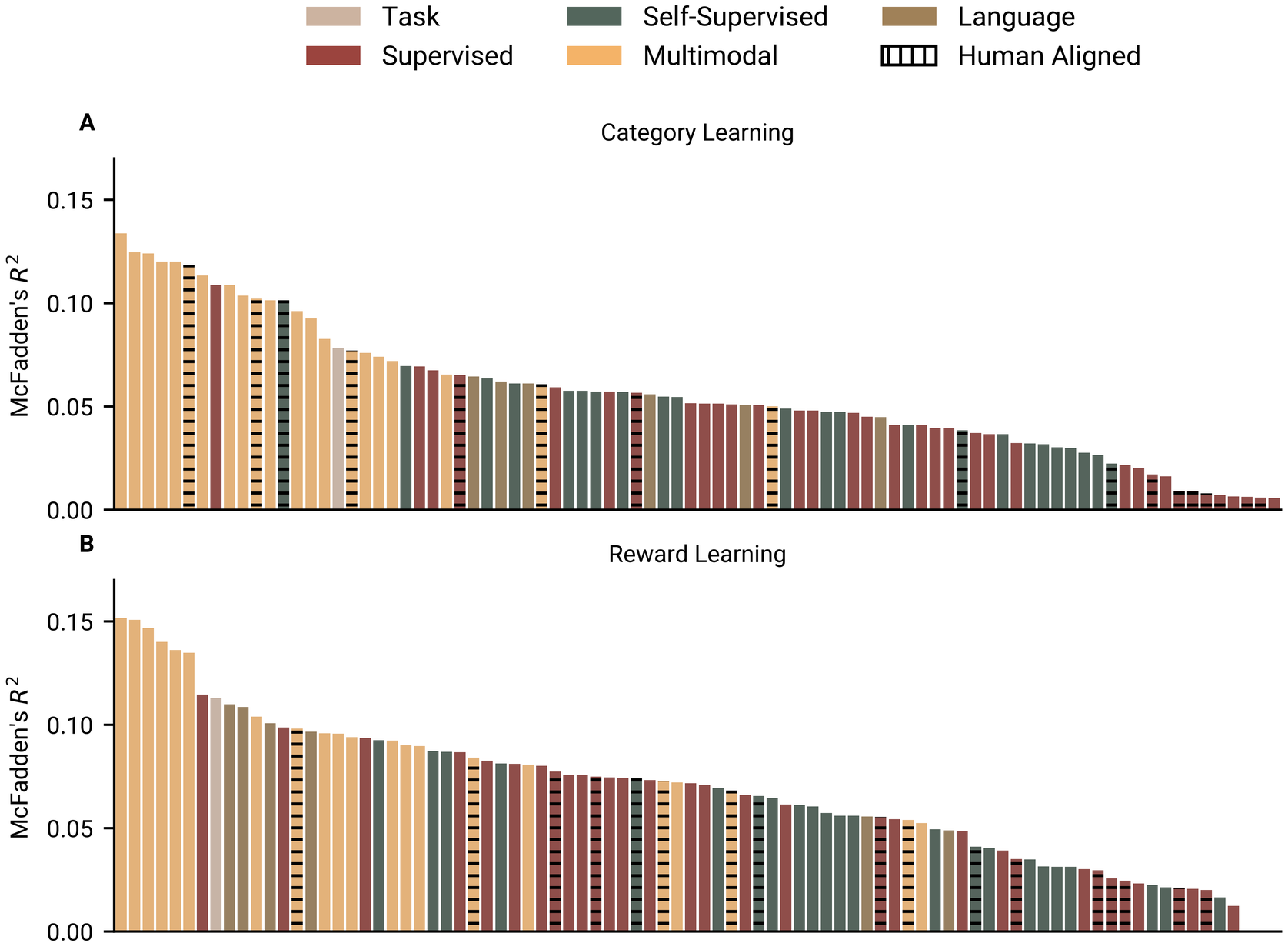
C



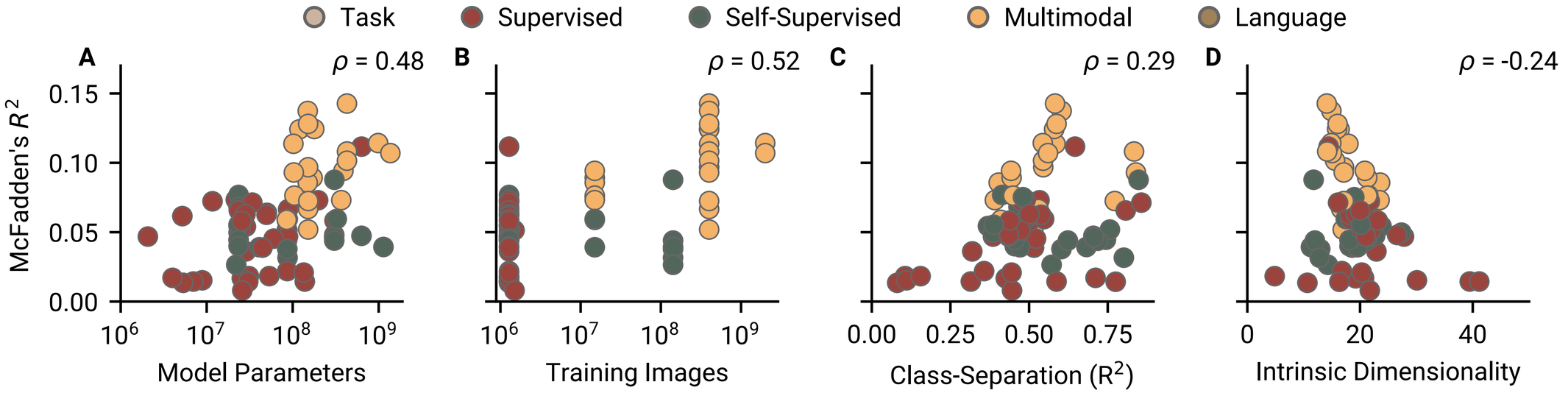
Both humans and neural networks can learn to solve the tasks.



Multimodal models are particularly human-like in how they perform.



Several factors are important for alignment



Future outlook

- For ML: Measuring alignment using semantically rich tasks can help build stronger models.

- For CogSci: Opportunities to study behaviour in more naturalistic settings and leverage pretrained neural networks for cognitive models

Future outlook

- For ML: Measuring alignment using semantically rich tasks can help build stronger models.

- For CogSci: Opportunities to study behaviour in more naturalistic settings and leverage pretrained neural networks for cognitive models

Poster

Evaluating alignment between humans and neural network representations in image-based learning tasks

Can Demircan · Tankred Saanum · Leonardo Pettini · Marcel Binz · Blazej Baczkowski · Christian Doeller · Mona Garvert · Eric Schulz

[Abstract]

Wed 11 Dec 11 a.m. PST — 2 p.m. PST (Bookmark)