Zero-Shot Visual Numerical Reasoning in Dual-Stream Neural Networks

Jessica A.F. Thompson[1], Hannah Sheahan[2], Tsvetomira Dumbalska[1], Julian Sandbrink[1], Manuela Piazza[3], Christopher Summerfield[1]

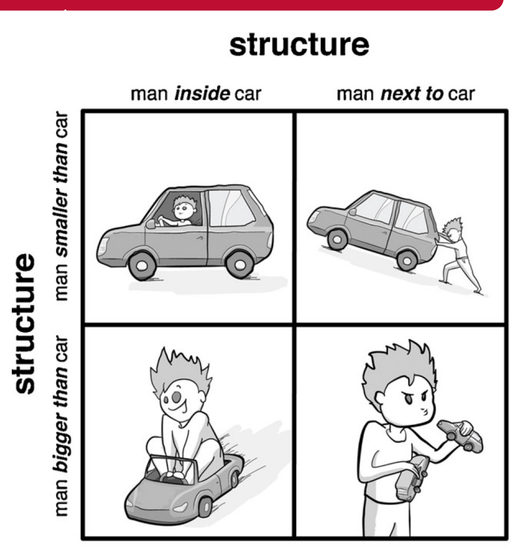
[1]Department of Experimental Psychology, University of Oxford [2]Google DeepMind, London, UK [3]University of Trento, Trento, Italy

Introduction

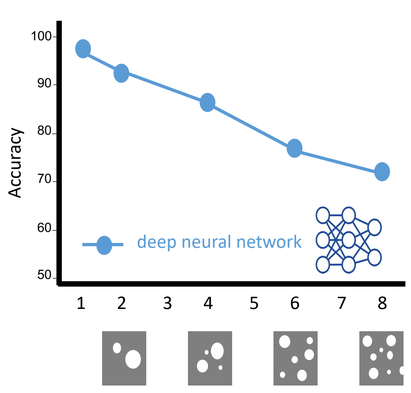
Visual scene understanding requires reasoning about the relations among objects—the "structure" of visual scenes. Here we use numerical reasoning as a testbed to study visual relational reasoning in the primate brain.

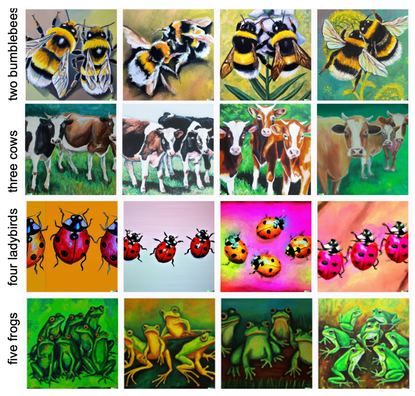
Research Goals:

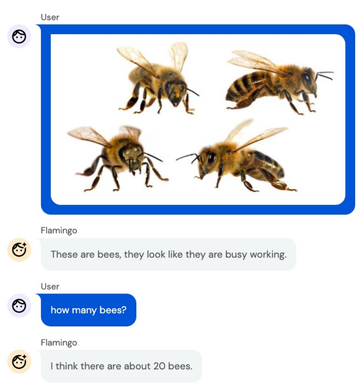
- Formalize theory of primate relational reasoning in a neural network model
- Demonstrate that the model can generalize numerical reasoning zero-shot
- Show that it generalizes *because* of the specific neural-inspired features we built in
- Understand how its function and organization relate to visual numerical reasoning in biology



Zero-shot numerical reasoning challenges modern AI systems







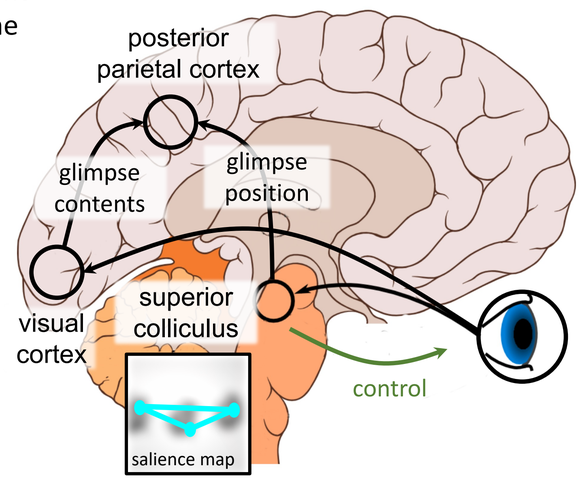
Numerical Reasoning in the Primate Brain

Beyond the ventral stream:

- Patients with damage to parietal regions (e.g., intraparietal sulcus) show deficits in numerical cognition.
- Electrophysiology in monkeys and fMRI in humans have revealed topographic representations of visual number in posterior parietal cortex
- Eye-movements contain content-invariant information about the structure of visual scenes

Hypothesize that relational reasoning enabled by:

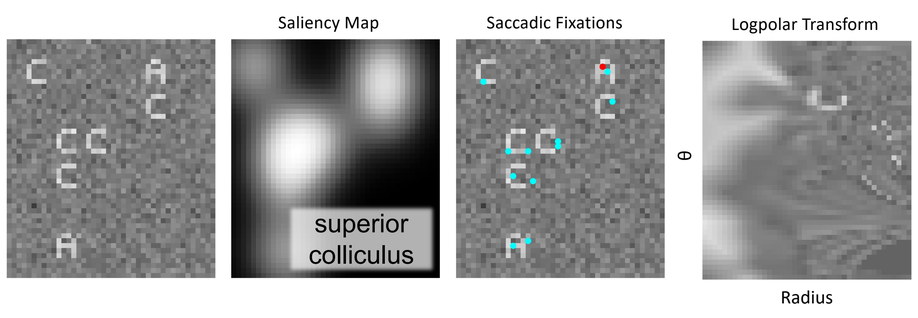
- Factorized representations of scene contents and structure in the parallel visual pathways
- Efferent copies of action-related signals (e.g., eye movements) provide relational information, enabling abstractions grounded in action
- Signal integration in posterior parietal cortex



Model

Simulating Foveated Glimpses

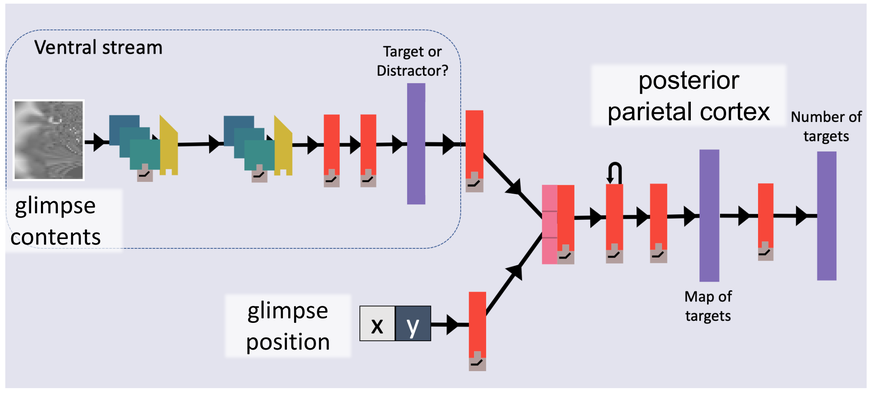
Saccadic targets (fixation points) are sampled from a saliency map of the image, subject to the constraint that all items are glimpsed at least once.



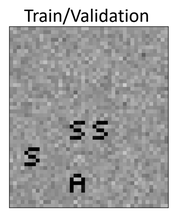
We model the retinal-to-cortical transformation as a log-polar transform centered on the fixation point.

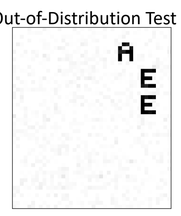
Dual-Stream Recurrent Glimpse Network

Model embodies our hypotheses about how the parallel pathways of the primate visual systems and posterior parietal cortex serve zero-shot visual numerical reasoning.

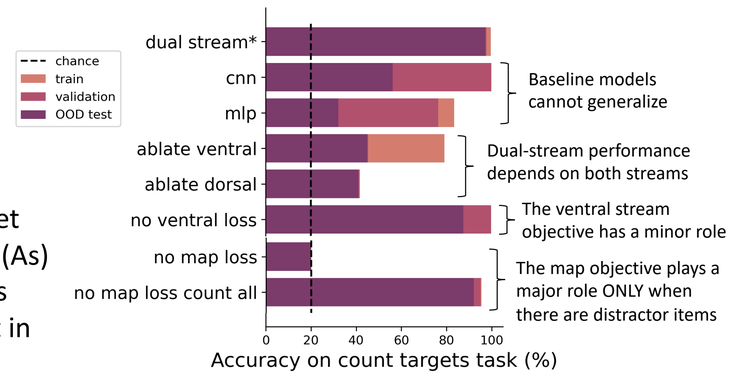


Inspecting Model Performance

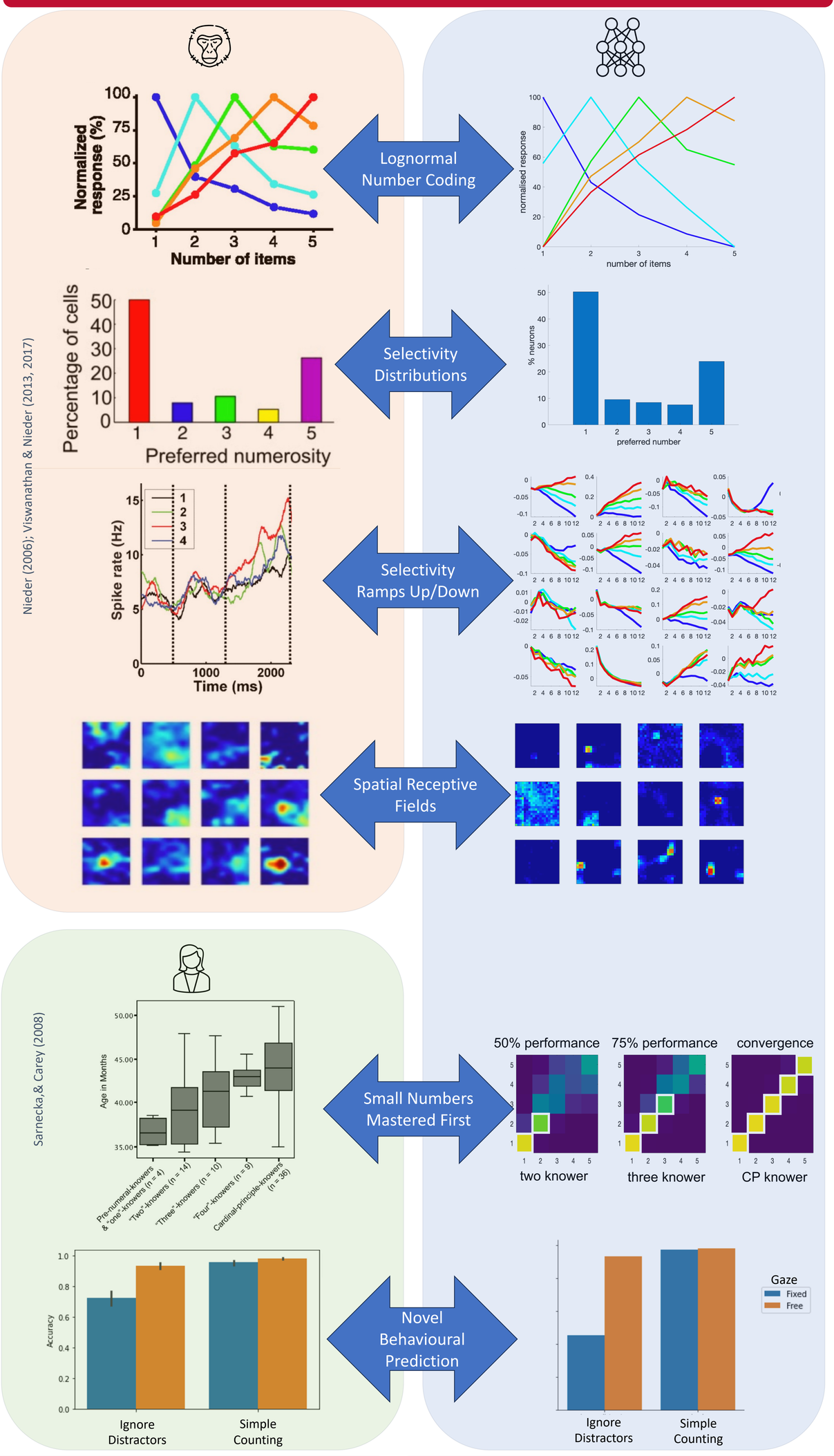




- Training objective: Count all target letters while ignoring distractors (As)
- OOD test set includes new letters and new luminances not present in the training set.



Neural and Behavioural Comparisons



Conclusion

Neuro and cognitively-inspired dual-stream neural network:

- Displays zero-shot numerical reasoning
- Mirrors behavioural and neural signatures of numerical/spatial cognition
- Makes verified predictions about human behaviour

Evidence for a theory of the role of PPC in visual relational reasoning

References

- Summerfield, C., Luyckx, F., & Sheahan, H. (2020). Structure learning and the posterior parietal cortex. *Progress in Neurobiology*.
- Viswanathan, P., & Nieder, A. (2017). Comparison of visual receptive fields in the dorsolateral prefrontal cortex and ventral intraparietal area in macaques. *European Journal of Neuroscience*.
- Viswanathan, P., & Nieder, A. (2013). Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. *PNAS*.
- Nieder, A., Diester, I., & Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Neuroforum*.
- Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*.

Human Brain Project

European Research Council
Established by the European Commission

This work was supported by funding from the European Research Council (ERC Consolidator award 725937) and Special Grant Agreement No. 945539 (Human Brain Project SGA).

Email: Jessica.Thompson@psy.ox.ac.uk