Human and Machine Cognition Lab

What makes humans so uniquely intelligent?
How do people make the best use of limited cognitive resources?
What are the unique algorithms we use to learn from other people?

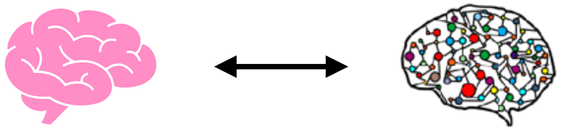
Lab Rotations and BSc/MSc Thesis Projects

hmc-lab.com

Dr. Charley Wu
Group Leader
charley.wu@uni-tuebingen.de

About the HMC Lab

The HMC Lab is an Independent Research Group led by Dr. Charley Wu, with the goal of understanding the gap between human and machine learning.

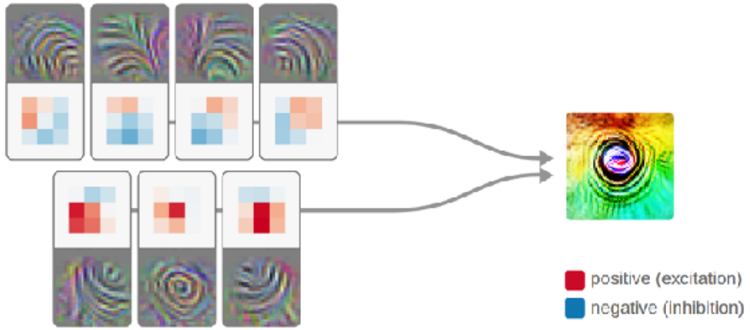


Our research methods include:

- online experiments (commonly in the form of interactive games)
- lab-based virtual reality experiments
- computational modeling of behavior (e.g., decisions, search trajectories, and reaction times)
- evolutionary models and simulations
- developmental studies (comparing children and adults)
- neuroimaging using fMRI/EEG
- analyzing large scale real-world datasets

We also have a rich collaboration network of researchers from Harvard, Princeton, UCL, and several Max Planck Institutes around Germany. To find out more, visit the lab website at **www.hmc-lab.com**

Project 1: Revealing the mechanisms of neural algorithms through *hyperSVD*



What this project is about:

Neural networks don't need to be black boxes; we just need methods that decompose their internal algorithms into constituent parts so that we can understand their mechanisms.

If networks were linear we could decompose their computations using Singular Value Decomposition (SVD). But they're non-linear, so vanilla SVD can't help.

This project develops a simple but very general method (hyperSVD) to decompose nonlinear computations into understandable parts.

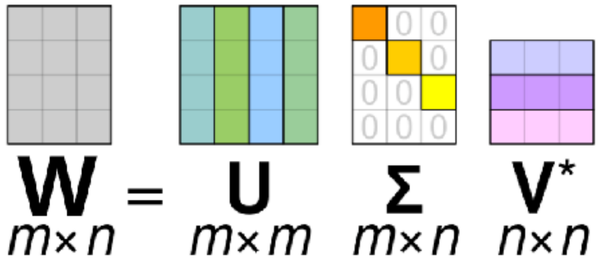
Why it's important:

Our methods for analysing high dimensional nonlinear systems are very poor. Better methods like hyperSVD are important for:

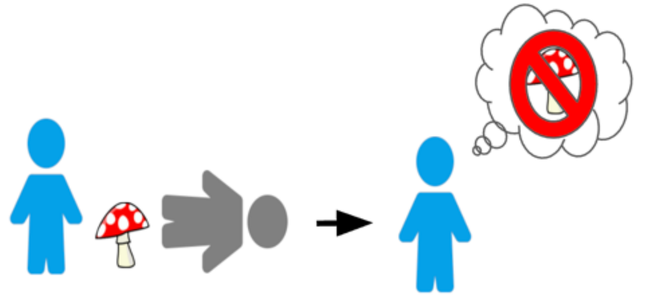
- AI Safety: Actually understand what networks have learned to ensure they are safe, fair, and robust to distributional shift and adversarial examples.
- Neuroscience & cognitive science: Study artificial networks to generate hypotheses about what biological neural networks are doing
- Dynamical systems: Analyse any high dimensional dynamical system, something we're currently very bad at doing!

Skills that you will use/develop:

- Linear Algebra
- Interpretability/explainability
- Neural network training (Pytorch)
- Hypernetworks
- High dimensional neural data analysis methods
- Adversarial examples



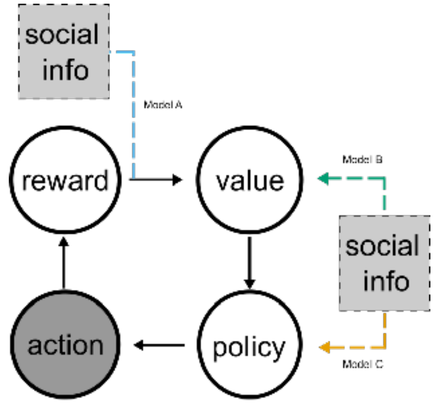
Project 2: Social learning strategies



Social learning is a powerful ability we use almost constantly in our everyday lives. Yet understanding of the computational mechanisms are still limited, preventing us from developing human-like social learning in artificial agents

In this project, you would be working on understanding how humans use social learning so effectively (Wu et al., 2022)

Q1: How is social information integrated into our decision making process?



Q2: How do humans trade-off between computational complexity and informational gain in social learning?

You could either gain experience with computational and evolutionary simulations (on the basis of existing code) or design your own online experiment

Project 3: Taming the cost of control by disentangling complex representations

Research Question

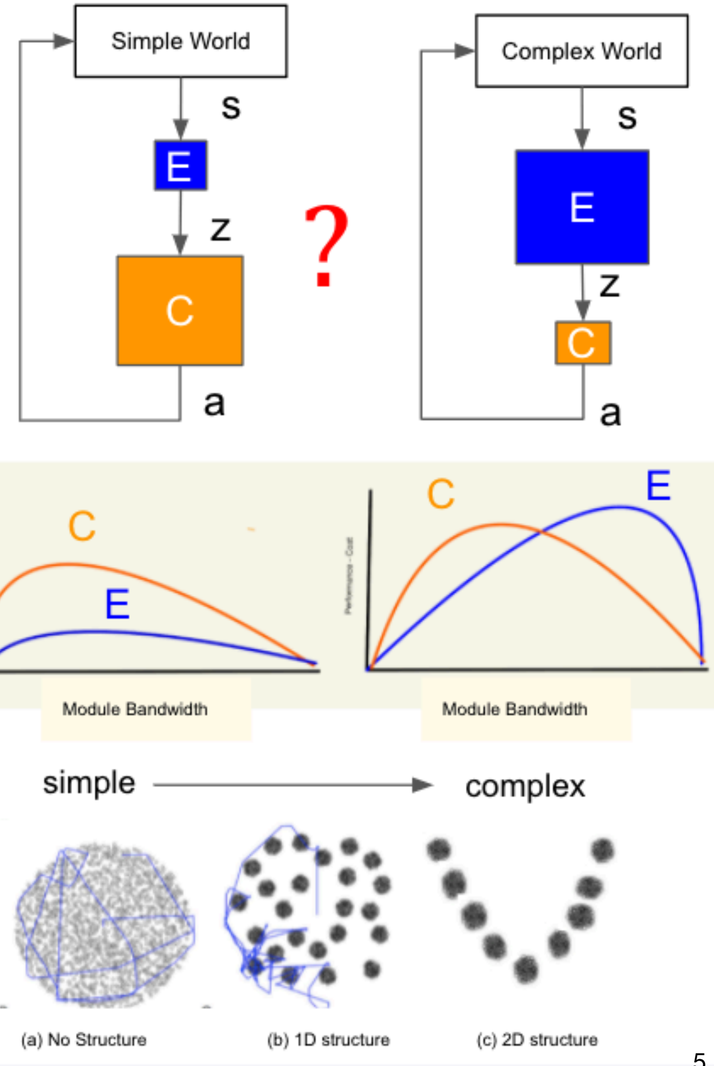
How do humans modulate the complexity of their representations and behavioral policies in response to the varying complexity of the world?

Approach

- Consider a simple learning model where an encoder E compresses input from the environment into a latent representation z, and a controller module C learns a policy mapping z→a
- An adaptive agent should be able to adapt the representational bandwidth of E and the capacity of C to the complexity of the environment
- What kind of adaptive learning system can facilitate optimal information exchange between the encoder and the controller, to minimize the costs of learning?

Scope

- Design and implement an online web-based experiment (HTML, JavaScript, PHP) to test the predictions of our model
- Work with a PhD student to learn about computational modeling using both evolutionary and classical RL frameworks



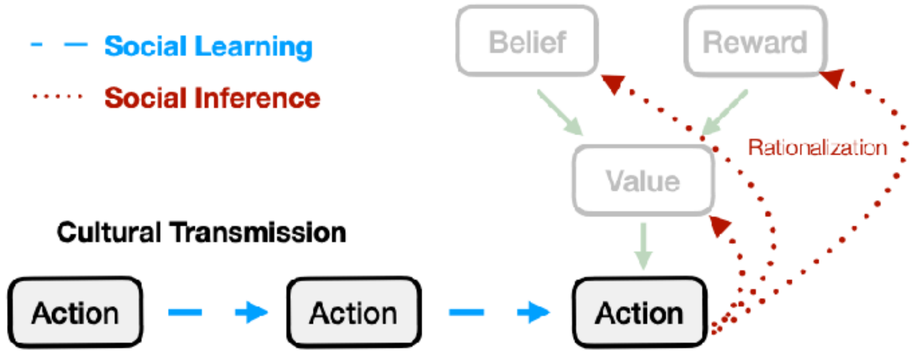
Project 4 : Learning from partial solutions in Virtual Reality

Humans are powerful social learners, and capable of learning even from failed or partial solutions (Wu et al., 2022).

What are the computational mechanisms underlying the transmission of social information that allow us to reconstruct the causal structure of observed behavior?

Scope

- Design and implement a multi-participant experiment in a new VR lab
- Work with Unity and other VR frameworks to record data and create interactive environments for studying behavior
- Use gaze and eye-tracking (HTC Vive Pro Eye) to analyze social attention



Project 5: Propose your own project!

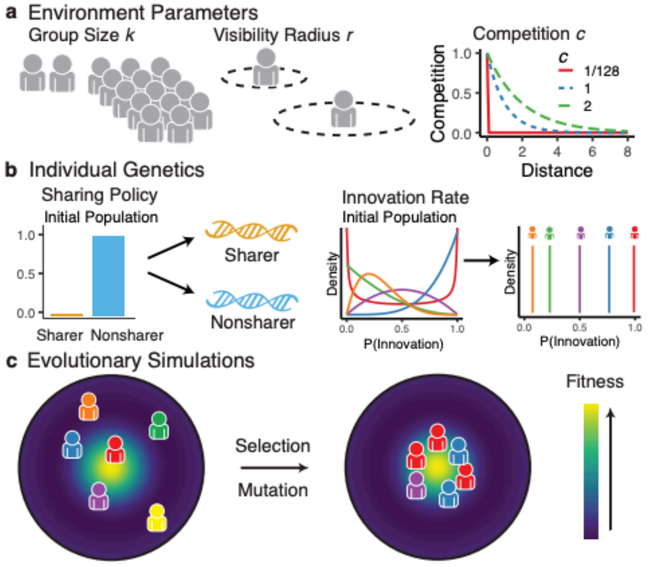
- Take the reigns and propose your own research project! To make things feasible within the rotation period or for a thesis, here are some suggestions of projects with existing data/code that could be built upon:

- **How does cooperation arise in competitive environments?** Through a series of agent-based and evolutionary simulations, we found that unconditional sharing of information can be beneficial, even in the absence of traditional reciprocity or reputation-based mechanisms. Many open questions, new environments, and learning mechanisms that can be tested

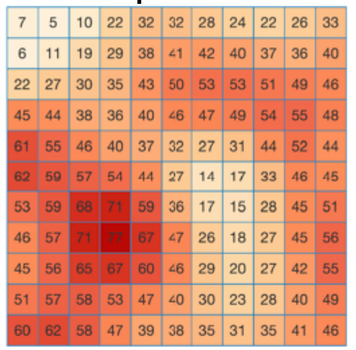
- **Why do people systematically under-generalize? Why are people systematically biased towards performing local search?** These are unexplained questions from a series of previous papers studying the search for rewards in spatially structured (Wu et al., 2018) and conceptually structured (Wu et al., 2020), and graph-structured environments (Wu et al., 2021). All the code and data are publicly available (1, 2, 3)

- **Note**: proposing your own project requires a high level of independent thinking and ability to craft an interesting and obtainable research question

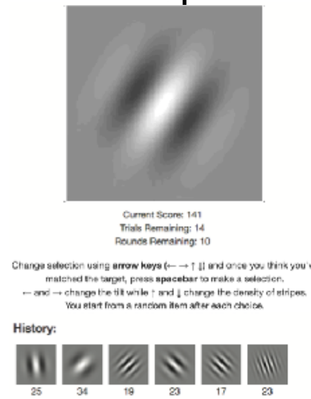
Evolutionary simulations



Spatial



Conceptual



Graph-structured