Reasoning Across Minds and Machines

Hanbo Xie[1] **Jian-Qiao Zhu**[2] **Hua-Dong Xiong**[1] **Robert C. Wilson**[1] **Thomas L. Griffiths**[2,3]
[1]School of Psychology, Georgia Institute of Technology, Atlanta, USA
[2]Department of Computer Science, Princeton University, Princeton, USA
[3]Department of Psychology, Princeton University, Princeton, USA
hanboxie1997@gatech.edu, jz5205@princeton.edu, hdx@gatech.edu, rwilson337@gatech.edu, tomg@princeton.edu

Keywords: Reasoning, Language, Large Language Models, Planning, Learning

Introduction

Reasoning is one of the hallmarks of both natural and artificial intelligence. Understanding how reasoning operates in the human mind is crucial in cognitive science. Despite a long history of research on human reasoning in cognitive science—ranging from heuristics (Tversky & Kahneman, 1974) to mental models (Johnson-Laird, 1983), and from Bayesian modeling (Oaksford & Chater, 2007; Griffiths, Chater, & Tenenbaum, 2024) to neuroscience (Goel & Dolan, 2003)—little is known about how humans reason so flexibly in real life and how reasoning contributes to high-level cognitive functions including planning, social interaction, complex problem-solving, and open-ended mental exploration. Previous studies face challenges that hinder a deeper understanding of reasoning, including the difficulty of designing well-balanced experimental paradigms that maintain both control and ecological validity, efficient data collection and analysis beyond pure behavioral measures (e.g., Think-Aloud text data (Simon & Ericsson, 1984)), and understanding complex interactions between reasoning and other high-level cognitive functions, such as memory, theory of mind, and language.

On the other hand, with the rapid development of Artificial Intelligence (AI), particularly Large Language Models (LLMs), we have the first artificial systems that have been claimed to show human-like reasoning capabilities (Binz & Schulz, 2023; Webb, Holyoak, & Lu, 2023; Yax, Anlló, & Palminteri, 2024). AI researchers train these models on vast linguistic datasets, and improve their performance using techniques such as Chain-of-Thought prompting (Wei et al., 2022) and Tree-of-Thought reasoning (Yao et al., 2024). Recent reasoning models have reached or even surpassed human performance on numerous benchmarks, showcasing the potential for highly capable reasoning systems, such as OpenAI-O1 (Jaech et al., 2024) and DeepSeek-R1 (Guo et al., 2025). However, a major challenge in AI research remains limited understanding of the underlying reasoning processes used by these models.

To bridge the gap between human and machine reasoning—and to facilitate both a better understanding of human cognition and improved AI reasoning capabilities—this workshop aims to bring together leading scholars from cognitive science, psychology, neuroscience, and computer science. Our goal is to discuss key challenges in the field, share cutting-edge methodologies, and exchange perspectives on studying reasoning in intelligent systems.

Goals and Scope

This workshop aims to create a collaborative platform where researchers from psychology, cognitive science, neuroscience, and artificial intelligence can explore how reasoning processes manifest in both humans and machines. We aim to address questions surrounding reasoning from conceptual, methodological, algorithmic, and computational perspectives, fostering an interdisciplinary exchange of ideas and approaches. This exploration will deepen our understanding of human cognition, and such insights could potentially inform advancements in AI.

Topics we plan to cover include, but are not limited to:

- Human reasoning strategies and heuristics.
- Methods for studying human reasoning (e.g., introspection, experimental tasks).
- How language reflects or informs reasoning.
- Instilling reasoning skills in machines (e.g., chain-of-thought, tree-of-thought, O1 model, RL training).
- Neural substrates underlying reasoning.

Target Audience

This workshop will be of interest to researchers working in any field related to reasoning, as well as those with a general interest in intelligence. Potential participants include psychologists, cognitive scientists, neuroscientists, and computer scientists, among others.

Organizers and Presenters

Hanbo Xie (Organizer) is a Ph.D student in the School of Psychology at Georgia Tech. His research focuses on understanding what makes humans so intelligent and how can humans and artificial intelligence benefit each other.

Jian-Qiao Zhu (Organizer) is a postdoctoral fellow in the Department of Computer Science at Princeton University. He works in Bayesian and neural network models of cognition.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

Huadong Xiong (Organizer) is a Ph.D. student in the School of Psychology at Georgia Tech. He is interested in the normative explanation of intelligence.

Robert C. Wilson (Organizer) is an Associate Professor in the School of Psychology at Georgia Tech. His research is to build mathematical theories of the mind and brain that explain behavior in health, mental illness, and cognitive decline.

Thomas L. Griffiths (Organizer) is a Professor in the Departments of Psychology and Computer Science at Princeton University. His research aims to develop mathematical models of human cognition, with a particular focus on the aspects of human minds that are meaningfully different from modern AI systems.

Ulrike Hahn (Presenter) is a Professor in the School of Psychological Sciences at Birbeck, University of London. She is interested in human rationality, particularly how individuals and groups evaluate evidence, reason, and make decisions, with a focus on argumentation, testimony, and the influence of social networks on belief accuracy.

Mark Ho (Presenter) is an Assistant Professor in the Department of Psychology at New York University. He is interested in how people's goals, values, and motivations structure their thoughts, decisions, and interactions with others.

Anna Ivanova (Presenter) is an Assistant Professor in the School of Psychology at Georgia Tech. She is interested in studying the relationship between language and other aspects of human cognition.

David Lagnado (Presenter) is a Professor in the Department of Psychology and Language Science at University College London. He is interested in how people use causal models to learn about the world, to draw inferences, and to assign blame and praise.

Andrew Lampinen (Presenter) is a Staff Research Scientist at Google DeepMind. He is interested in how the complex behaviors and representations of models, agents, or humans emerge from their learning experiences or data.

Marcelo Mattar (Presenter) is an Assistant Professor in the Department of Psychology at New York University. His research circles how the brain simulates the past and future to guide our decisions.

The panel will be moderated by Robert Wilson with all the presenters.

Workshop Structure

This workshop will be a half-day event. Each presenter will have 25 minutes for their presentation, followed by a 5-minute Q&A session.

The panel discussion will last 30 minutes, with the first 20 minutes dedicated to discussing prepared questions that address philosophical, conceptual, and methodological aspects of human and machine reasoning research. Topics may include questions such as, "Is understanding human reasoning necessary for developing better reasoning models?" The remaining 15 minutes will be allocated for audience questions.

After the panel discussion, attendees and presenters are welcome to stay for further informal discussions or leave as they wish.

Presenter	**Topics Covered**
Ulrike Hahn	Human reasoning strategies and heuristics How does language reflect/inform reasoning?
Mark Ho	Human reasoning strategies and heuristics Methods for studying human reasoning
Anna Ivanova	How does language reflect/inform reasoning?
David Lagnado	Human reasoning strategies and heuristics Causal Reasoning
Andrew Lampinen	Human reasoning strategies and heuristics Instilling reasoning skills in machines
Marcelo Mattar	Human reasoning strategies and heuristics Methods for studying human reasoning
Panel Discussion: Robert Wilson, Ulrike Hahn, Mark Ho, Anna Ivanova, David Lagnado, Andrew Lampine, Marcelo Mattar	

Table 1: Workshop Presenters and Their Topics

References

Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand gpt-3. *Proceedings of the National Academy of Sciences*, *120*(6), e2218523120.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, *87*(1), B11–B22.

Griffiths, T. L., Chater, N., & Tenenbaum, J. B. (2024). *Bayesian models of cognition: reverse engineering the mind*. MIT Press.

Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., ... others (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Jaech, A., Kalai, A., Lerer, A., Richardson, A., El-Kishky, A., Low, A., ... others (2024). Openai o1 system card. *arXiv preprint arXiv:2412.16720*.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press, USA.

Simon, H. A., & Ericsson, K. A. (1984). Protocol analysis: Verbal reports as data.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), 1124–1131.

Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, *7*(9), 1526–1541.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... others (2022). Chain-of-thought prompting elicits

reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824–24837.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2024). Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, *36*.

Yax, N., Anlló, H., & Palminteri, S. (2024). Studying and improving reasoning in humans and machines. *Communications Psychology*, *2*(1), 51.