UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Thinking about thinking through inverse reasoning

Permalink
https://escholarship.org/uc/item/77v2c6pm

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Berke, Marlene
Jara-Ettinger, Julian

Publication Date
2021

Peer reviewed

Thinking about thinking through inverse reasoning

Marlene Berke

Yale University, New Milford, Connecticut, United States

Julian Jara-Ettinger

Yale University, New Milford, Connecticut, United States

Abstract

Human Theory of Mind enables us to attribute mental states like beliefs and desires based on how other people act. However, in many social interactions (particularly ones that lack observable action), people also directly think about other people's thinking. Here we present a computational framework, Bayesian inverse reasoning, for thinking about other people's thoughts. Our framework formalizes inferences about thinking by inferring a generative model of reasoning decisions and computational processes, structured around a principle of rational mental effort – the idea that people expect other agents to allocate thinking rationally. We show that this model quantitatively predicts human judgements in a task where participants must infer the mental causes behind an agent's pauses as they navigate and solve a maze. Our results contribute to our understanding of the richness of the human ability to think about other minds, and to even think about thinking itself.