

Modeling Cognitive Strategies in Teaching: Integrating Theory of Mind and Heuristics

Sevan Harootonian

Princeton University, Princeton, New Jersey, United States

Yael Niv

Princeton University, Princeton, New Jersey, United States

Tom Griffiths

Princeton University, Princeton, New Jersey, United States

Mark Ho

Stevens Institute of Technology, Hoboken, New Jersey, United States

Abstract

Teaching plays a crucial role in human learning, from formal educational environments to mentorship scenarios, yet its cognitive underpinnings remain underexplored. We focus on the distinction between teaching by reasoning using Theory of Mind (i.e., explicitly inferring what a learner knows) and teaching using heuristics (i.e., relying on a simple rule). We use a graph-navigation task where a learner agent with limited knowledge attempts to navigate through the most rewarding trajectory, with guidance from a human teacher. Our findings reveal that teachers utilize a blend of learner-specific strategies and general heuristics. We model learner-specific strategies using Bayesian Theory of Mind (Baker, Saxe, & Tenenbaum, 2009) and demonstrate that the most effective teachers incorporate this strategy. Intriguingly, we show that teaching strategies can be altered without explicit feedback. This suggests that subtle changes in the environment may significantly alter teaching approaches, highlighting the importance of understanding the cognitive processes behind teaching.