Learning Causal Overhypotheses through Exploration in Children and Computational Models

Eliza Kosoy
UC Berkeley , Berkeley, California, United States

Adrian Liu
UC Berkeley, Berkeley, California, United States

Jasmine Collins
UC Berkeley, Berkeley, California, United States

David Chan
University of California, Berkeley, Berkeley, California, United States

Jessica Hamrick
DeepMind, London, United Kingdom

Sandy Huang
DeepMind, London, United Kingdom

Nan Rosemary Ke
Deepmind, Mila, Montreal, Quebec, Canada

Bryanna Kaufmann
UC Berkeley, Berkeley, California, United States

Alison Gopnik
University of California at Berkeley, Berkeley, California, United States

Abstract

Human children are proficient explorers, using causal information to great benefit. In contrast, typical AI agents do not consider underlying causal structures during exploration. To improve our understanding of the differences between children and agents—and ultimately to improve AI agents' performance—we designed a virtual Blicket experiment to test childrens' ability to leverage causal information while exploring a novel environment. This experiment doubles as an RL environment with a controllable causal structure, allowing us to evaluate exploration strategies used by both agents and children. Our results demonstrate that there are significant differences between information-gain optimal RL exploration and the exploration of children: in particular, children appear to consider a wide range of creative overhypotheses, including stochasticity, total weight, object ordering, and more. We leverage this new insight to lay the groundwork for future research into efficient exploration and disambiguation of causal structures for RL algorithms.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*