# Naturalistic Transmission of Causal Knowledge between Machines and Humans

**Cédric Colas**
MIT, Cambridge, Massachusetts, United States

**Tracey Mills**
MIT, Cambridge, Massachusetts, United States

**Ben Prystawski**
Stanford University, Stanford, California, United States

**Michael Tessler**
Google DeepMind, London, United Kingdom

**Noah Goodman**
Stanford University, Stanford, California, United States

**Jacob Andreas**
MIT, Cambridge, Massachusetts, United States

**Josh Tenenbaum**
MIT, Cambridge, Massachusetts, United States

**Abstract**

Human ecological success stems from our ability to absorb and build upon cultural knowledge, a process we aim to model computationally by integrating individual and cultural learning from language — one of the main vehicles of cultural transmission (e.g. instructions, explanations, stories). In simple video games, our model infers game rules from both interaction data (individual learning) and partial causal models extracted from game descriptions (cultural learning). Given exhaustive descriptions (either hand-written or generated by a model given access to oracle data), models leveraging the two learning sources induce more accurate game rules from limited data than both the individual- and cultural-only controls. Interestingly, descriptions from human game players do not consistently yield better rule induction. We hypothesize that players may preferentially communicate information that will be essential to the others' future decision-making and we aim to investigate cultural transmission by integrating individual and cultural learning with both causal understanding and decision-making.



In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* 