

# A Neurosymbolic Model of Human Reasoning on the Abstraction and Reasoning Corpus

**Solim LeGris**

NYU, New York, New York, United States

**Brenden Lake**

NYU, New York, New York, United States

**Todd Gureckis**

New York University, New York, New York, United States

## Abstract

The Abstraction and Reasoning Corpus (ARC) is a visual program synthesis benchmark designed to test out-of-distribution generalization in machines. Although recent advancements have led to human-level performance on ARC, it is still unclear how people solve ARC tasks. Other work has demonstrated that people often make errors when reasoning about these problems, and sometimes fail to infer the true underlying program. In this work, we explore the hypothesis that the cognitive mechanisms supporting reasoning are more approximate, graded and resource-limited compared to those suggested by purely discrete, program-induction models. Taking inspiration from previous work on program sketches and partial programs, we model human reasoning in the ARC domain using a neurosymbolic, meta-learned model that interleaves symbolic operations with approximate, statistical pattern completion. We then evaluate the model against human errors from H-ARC, a recent, comprehensive behavioral dataset on the training and evaluation sets of ARC.