# Doing Experiments and Revising Rules with Natural Language and Probabilistic Reasoning

**Top Piriyakulkij**
Cornell University, Ithaca, New York, United States

**Kevin Ellis Ellis**
Cornell, Ithaca, New York, United States

**Abstract**

We build a computational model of how humans actively infer hidden rules by doing experiments. The basic principles behind the model is that, even if the rule is deterministic, the learner considers a broader space of fuzzy probabilistic rules, which it represents in natural language, and updates its hypotheses online after each experiment according to approximately Bayesian principles. In the same framework we also model experiment design according to information-theoretic criteria. We find that the combination of these three principles – explicit hypotheses, probabilistic rules, and online updates – can explain human performance on a Zendo-style task, and that removing any of these components leaves the model unable to account for the data.



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.* 