

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Doing Experiments and Revising Rules with Natural Language and Probabilistic Reasoning

Permalink

<https://escholarship.org/uc/item/1b04k0xx>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Piriyakulkij, Top

Ellis, Kevin

Publication Date

2024

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Doing Experiments and Revising Rules with Natural Language and Probabilistic Reasoning

Top Piriyaakulkij

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.