Machine Learning Optimizes Assessment: New Insights for the Development of Numerosity Estimation

Sang Ho Lee
The Ohio State University, Columbus, Ohio, United States

Dan Kim
The Ohio State University, Columbus, Ohio, United States

John Opfer
The Ohio State University, Columbus, Ohio, United States

Mark Pitt
Ohio State University, Columbus, Ohio, United States

Jay Myung
The Ohio State University, Columbus, Ohio, United States

Abstract

In a conventional number-line task, a given number that varies every trial is estimated on a line flanked with 0 and an upper-bound number. An upper-bound number is often arbitrarily selected, although this design variable has been shown to affect non-linearity in estimates. Examining estimates of varying given numbers (design variable 1) with varying upper-bound numbers (design variable 2) can be costly because adding another design variable into the task drastically increases the number of trials required to examine the numerical representation. In the present study, a novel Bayesian machine learning algorithm, dubbed Gaussian Process Active Learning (GPAL), was used to make this costly paradigm feasible by presenting only the most informative combinations of the design variables every trial. We found that children were more logarithmic than adults across upper bounds, replicating log-to-linear shifts in development. More importantly, children and even educated adults became more logarithmic as the upper bound increased, indicating the persistent use of log representation across age groups.