

Probing Nonhuman Primate Errors on False Belief Tasks to Explore the Evolutionary Roots of Theory of Mind

Amanda Royka

Yale University, New Haven, Connecticut, United States

Daniel Horschler

Yale University, New Haven, Connecticut, United States

Walker Bargmann

Yale University, New Haven, Connecticut, United States

Laurie Santos

Yale University, New Haven, Connecticut, United States

Abstract

Theory of Mind (ToM) is central to human social cognition, yet the roots of this capacity remain poorly understood. Both infants and nonhuman primates perform inconsistently on false belief tasks, limiting our understanding of the representations that characterize their ToM. Here, we seek to better understand this often-contradictory literature by dissecting these failures. Specifically, we focus on primates' characteristic null performance on false belief tasks. Across three studies, we find that—despite succeeding on a closely-matched control—rhesus monkeys fail to predict how agents with false beliefs will behave even when the agents perform highly unexpected, unlikely actions. We interpret this pattern of performance as evidence that monkeys have no representation of another agent's past awareness once the scene changes outside of that agent's view. This work moves beyond the success/failure dichotomy typically used to assess ToM, and instead gives a more precise characterization of primates' signature limits in ToM.