Counterfactual error-monitoring in human planning

Doris Yu
New York University , New York, New York, United States

Frederick Callaway
New York University, New York, New York, United States

Marcelo Mattar
New York University, New York, New York, United States

Abstract

Human decision-making involves evaluating choices and learning from outcomes, yet how individuals process information about paths previously or no longer available options remains unclear. While traditional reinforcement learning models suggest that evaluating unchosen options optimizes future decisions, online planning algorithms typically ignore unavailable paths that can not inform current choices. This raises a critical question: do humans actively monitor unchosen alternatives during online planning? Using a two-stage decision task with eye-tracking and reaction-time analyses, we show that participants systematically monitor alternative paths, especially when their selected path proves suboptimal. Attention to unavailable options scales with their potential value, revealing an adaptive metacognitive process absent in online planning algorithms. These findings indicate that humans maintain and update representations of unavailable alternative choices rather than discarding them after selection. This work provides novel insights into the interplay between planning, metacognition, and adaptive behavior by demonstrating counterfactual evaluation during decision-making.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*