

# Spot the ball: Inferring Hidden Information from Human Behavioral Cues

**Neha Balamurugan**

Stanford University, Stanford, California, United States

**Sarah Wu**

Stanford University, Stanford, California, United States

**Cristobal Eyzaguirre**

Stanford University, Stanford, California, United States

**Adam Chun**

Stanford University, Stanford, California, United States

**Tobias Gerstenberg**

Stanford University, Stanford, California, United States

## Abstract

Humans often infer the state of the world by observing how others interact with it—when crossing a street, for instance, we may follow the movement of others without directly seeing the traffic. This ability to extract hidden information from human interactions with the environment is crucial for adaptive behavior. In this study, we explore how people make such inferences in Spot the Ball, a task where participants predict the location of a masked soccer ball in single-frame images. We created a large dataset by scraping YouTube videos, identifying compelling images using CLIP, and masking the soccer ball through inpainting. Our findings show that human participants rely heavily on pose and gaze cues to infer the ball’s location. While providing this information improves GPT-4o’s performance, it remains significantly below human accuracy. These results highlight the significance of intention inference, with potential applications in self-driving cars, assistive AI, and humanoid robotics.