

A generative model of people's intuitive theory of emotions: inverse planning in rich social games

Sean Houlihan

MIT, Cambridge, Massachusetts, United States

Max Kleiman-Weiner

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

Josh Tenenbaum

MIT, Cambridge, Massachusetts, United States

Rebecca Saxe

MIT, Cambridge, Massachusetts, United States

Abstract

We propose a formal model of humans' intuitive theories of others' emotions. From a single choice in a social interaction (e.g. the choice to cooperate in a Prisoner's Dilemma game), human observers can infer a player's complex values, such as prosocial preferences and reputational concerns. When the player then experiences a new situation (the game's outcome), observers infer the player's reaction to the event based on the mental state likely to have produced the player's action. Here we capture this process by inverting a richly structured generative model of social gameplay, including social equity and reputational dimensions, and translate players' subjective motivations, expectations, and prediction errors into forward predictions of the emotional experiences of the players. Our model infers players' values and expectations, generates patterns of play that match observers' intuitions, and supports formally generated emotion predictions with substantially extended breadth and nuance.