Emotion attributions echo the structure of people's intuitive theory of psychology

Sean Houlihan
MIT, Cambridge, Massachusetts, United States

Max Kleiman-Weiner
Harvard, Cambridge, Massachusetts, United States

Josh Tenenbaum
MIT, Cambridge, Massachusetts, United States

Rebecca Saxe
MIT, Cambridge, Massachusetts, United States

Abstract

We present a generative model of how observers think about the emotions experienced by players in a socially-charged game: a public, high-stakes, one-shot Prisoner's Dilemma. The model extends inverse planning frameworks to capture observers' judgments about players' reactions to hypothetical events. Observers attribute different beliefs and values to players based on what decisions the players make. We model how observers' noisy inferences of players' mental contents bias emotion predictions. Incorporation of non-monetary features into forward planning enables us to model emotions that reflect complex social concerns (e.g. Embarrassment depends on how much players think others will infer that they tried to take advantage of their opponents). In addition to matching the intensities of twenty attributed emotions, the model reflects how observers' emotion judgments covary within single stimuli, indicating that the model captures important aspects of the generative process underlying humans' emotion attributions in this game.