

# Empathy in Explanation

**Katherine Collins**

MIT, Cambridge, Massachusetts, United States

**Kartik Chandra**

MIT, Cambridge, Massachusetts, United States

**Jonathan Ragan-Kelley**

MIT, Cambridge, Massachusetts, United States

**Adrian Weller**

University of Cambridge, Cambridge, United Kingdom

**Joshua Tenenbaum**

MIT, Cambridge, Massachusetts, United States

## Abstract

Why do we give the explanations we do? Recent work has suggested that we should think of explanation as a kind of cooperative social interaction, between a why-question-asker and an explainer. Here, we apply this perspective to consider the role emotion plays in this social interaction. We develop a computational framework for modeling explainers who consider the emotional impact an explanation might have on a listener. We test our framework by modeling human intuitions about how a doctor should explain to a patient why they have a disease, depending on the patient's propensity for regret. Our model predicts human intuitions well, better than ablations suggestive that people do indeed reason about emotion when giving explanations. See <https://sites.google.com/view/empathy-in-explanation> for further details and pre-print.