Modeling Training Judgments of Ambiguous Face Emotions to treat Psychopathology

Joel Stoddard\textsuperscript{1}, Simone P. Haller\textsuperscript{2}, Vincent Costa\textsuperscript{3}, Melissa A. Brotman\textsuperscript{2}, Matt Jones\textsuperscript{4}

\textsuperscript{1}Emotion and Development Branch, National Institute of Mental Health, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD, 20892
\textsuperscript{2}Pediatric Mental Health Institute, Department of Psychiatry, School of Medicine, University of Colorado
\textsuperscript{3}Department of Behavioral Neuroscience, Oregon Health and Science University
\textsuperscript{4}Department of Psychology and Neuroscience, University of Colorado Boulder

Abstract

Some psychopathologies are associated with biases towards judging ambiguous social stimuli as reflecting hostile intent. Interventions targeting this bias are amenable to computational modeling to describe their associative learning mechanisms. Here we apply reinforcement learning and generalization models to describe associative learning during interpretation bias training and describe its initial validation in a group of youth with affective psychopathology. The model describes trialwise learning dynamics and predicts future behavior. Theoretical implications for interpretation training to treat psychopathology are discussed. This work has implications for current the development of interpretation bias training and the analysis of current clinical trials.