

Unsupervised Learning Shapes Emotion Categories

Rista Plate

University of Wisconsin-Madison , Madison, Wisconsin, United States

Adrienne Wood

University of Wisconsin-Madison, Madison, Wisconsin, United States

Seth D Pollak

University of Wisconsin-Madison, Madison, Wisconsin, United States

Abstract

Humans perceive facial expressions categorically, though physical features of emotions vary continuously. How do categorical representations of facial expressions emerge or update? We explored how supervised and unsupervised learning influence emotion category boundaries. 91 children (6-8-years-old) and 105 adults categorized emotions varying along a neutral-angry continuum. Participants completed a supervised learning phase, which explicitly taught an emotion category boundary. Then, participants completed an unsupervised learning phase. Without feedback, participants categorized expressions sampled from statistical distributions that matched or did not match the distribution categorized during supervised learning. Participants learned the boundary via supervised learning, but responses rapidly shifted following the statistical distribution via unsupervised learning. Thus, participants quickly updated emotion categories, indicating boundaries are highly context-sensitive. Such flexibility allows individuals to adjust across situations and organize responses based on extant, versus explicitly taught, socio-emotional cues. Follow-up research explores how participants adjust category boundaries for multiple individuals varying in expressivity.