EMPATH: A Neural Network that Categorizes Facial Expressions

Matthew N. Dailey[1], Garrison W. Cottrell[1], Curtis Padgett[2], and Ralph Adolphs[3]

Abstract

■ There are two competing theories of facial expression recognition. Some researchers have suggested that it is an example of ''categorical perception.'' In this view, expression categories are considered to be discrete entities with sharp boundaries, and discrimination of nearby pairs of expressive faces is enhanced near those boundaries. Other researchers, however, suggest that facial expression perception is more *graded* and that facial expressions are best thought of as points in a continuous, low-dimensional space, where, for instance, ''surprise'' expressions lie between ''happiness'' and ''fear'' expressions due to their perceptual similarity. In this article, we show that a simple yet biologically plausible neural network model, trained to classify facial expressions into six basic emotions, predicts data used to support both of these theories. Without any parameter tuning, the model matches a variety of psychological data on categorization, similarity, reaction times, discrimination, and recognition difficulty, both qualitatively and quantitatively. We thus explain many of the seemingly complex psychological phenomena related to facial expression perception as natural consequences of the tasks' implementations in the brain. ■

INTRODUCTION

How do we see emotions in facial expressions? Are they perceived as discrete entities, like islands jutting out of the sea, or are they more continuous, reflecting the structure beneath the surface? We believe that computational models of the process can shed light on these questions. Automatic facial expression analysis is an active area of computer vision research (Lien, Kanade, Cohn, & Li, 2000; Donato, Bartlett, Hager, Ekman, & Sejnowski, 1999; Lyons, Budynek, & Akamatsu, 1999; Rosenblum, Yacoob, & Davis, 1996). However, there has only been limited work in applying computational models to the understanding of human facial expression processing (Calder, Burton, Miller, Young, & Akamatsu, 2001; Lyons, Akamatsu, Kamachi, & Gyoba, 1998). In particular, the relationship between categorization and perception is controversial, and a computational model may help elucidate the connection between them.

Basic Emotions and Discrete Facial Expression Categories

Although the details of his theory have evolved substantially since the 1960s, Ekman remains the most vocal proponent of the idea that emotions are discrete entities. In a recent essay, he outlined his theory of basic emotions and their relationship with facial expressions (Ekman, 1999). ''Basic'' emotions are distinct families of affective states characterized by different signals, physiology, appraisal mechanisms, and antecedent events. Ekman cites early evidence suggesting that each emotion is accompanied by distinctive physiological changes that prepare an organism to respond appropriately. For instance, blood flow to the hands increases during anger, possibly in preparation for a fight. In addition to physiological changes, according to the theory, each basic emotion family is also accompanied by a fast appraisal mechanism that attends to relevant stimuli and a set of universal antecedent events (e.g., physical or psychological harm normally leads to a state of fear, and loss of a significant other normally leads to a state of sadness). Finally, and most importantly, Ekman believes that emotions evolved to ''inform conspecifics, without choice or consideration, about what is occurring: inside the person . . . , what most likely occurred . . . , and what is most likely to occur next'' (p. 47). Thus, every basic emotion family is necessarily accompanied by one (or perhaps a few for some families) distinctive prototypical signals, including a set of facial muscle movements and body movements (e.g., approach or withdrawal). The signals are not entirely automatic; they may be attenuated, masked, or faked in certain circumstances. Furthermore, within emotion families, individual differences and situational context allow for small variations on

[1]University of California, San Diego, [2]Jet Propulsion Laboratory, [3]University of Iowa

Note: EMPATH stands for ''EMotion PATtern recognition using Holons'', the name for a system developed by Cottrell & Metcalfe (1991).