Special Issue Article

Using automated computer vision and machine learning to code facial expressions of affect and arousal: Implications for emotion dysregulation research

Nathaniel Haines[1], Ziv Bell[1], Sheila Crowell[2,3], Hunter Hahn[1], Dana Kamara[1], Heather McDonough-Caplan[1],
Tiffany Shader[1] and Theodore P. Beauchaine[1]

[1]Department of Psychology, Ohio State University, Columbus, OH, USA; [2]Department of Psychology, University of Utah, Salt Lake City, UT, USA and [3]Department of Psychiatry, University of Utah, Salt Lake City, UT, USA

Abstract

As early as infancy, caregivers' facial expressions shape children's behaviors, help them regulate their emotions, and encourage or dissuade their interpersonal agency. In childhood and adolescence, proficiencies in producing and decoding facial expressions promote social competence, whereas deficiencies characterize several forms of psychopathology. To date, however, studying facial expressions has been hampered by the labor-intensive, time-consuming nature of human coding. We describe a partial solution: automated facial expression coding (AFEC), which combines computer vision and machine learning to code facial expressions in real time. Although AFEC cannot capture the full complexity of human emotion, it codes positive affect, negative affect, and arousal—core Research Domain Criteria constructs—as accurately as humans, and it characterizes emotion dysregulation with greater specificity than other objective measures such as autonomic responding. We provide an example in which we use AFEC to evaluate emotion dynamics in mother–daughter dyads engaged in conflict. Among other findings, AFEC (a) shows convergent validity with a validated human coding scheme, (b) distinguishes among risk groups, and (c) detects developmental increases in positive dyadic affect correspondence as teen daughters age. Although more research is needed to realize the full potential of AFEC, findings demonstrate its current utility in research on emotion dysregulation.

Keywords: arousal, emotion dysregulation, facial expression, negative valence system, positive valence system

(Received 16 November 2018; accepted 21 December 2018)

According to functionalist accounts, human emotions motivate evolutionarily adaptive behaviors, help us shape our environments to achieve desired ends, and serve important communicative purposes in social settings (see Beauchaine, 2015; Beauchaine & Haines, in press; Campos, Mumme, Kermoian, & Campos, 1994; Thompson, 1990). Displays of emotion, including facial expressions, gestures, and vocalizations, facilitate adaptive group cohesion by conveying our intentions, motivations, and subjective feeling states to others, which allows us to better understand and predict one another's behaviors (Darwin, 1872). Within social groups, expressions of emotion exhibit contagion, whereby individuals shift their attitudes, behaviors, and affective valence toward conformity with other group members and group norms (Smith & Mackie, 2015). Negative emotional contagion results in less cooperation and more conflict, whereas positive emotional contagion has the opposite effects (e.g., Barsade, 2002).

Given their importance to adaptive human function, it is of little surprise that facial expressions of emotion are used extensively to study socioemotional development and social interactions. Abnormalities in both production and decoding of facial expressions hinder socioemotional competence (Izard et al., 2001; Troster & Brambring, 1992). Moreover, environmental adversities including abuse and maltreatment confer biases in decoding facial expressions of anger and sadness, with adverse effects on social cognition and socioemotional function across development (Pollak, Cicchetti, Hornung, & Reed, 2000; Pollak & Sinha, 2002). Of note, treatments that improve children's facial expression recognition yield corresponding increases in social competence (Domitrovich, Cortes, & Greenberg, 2007; Golan et al., 2009).

Despite the important roles that production and decoding of facial expressions play in healthy socioemotional development, coding facial expressions is a time-consuming, labor-intensive process. Training a coding team takes months of regular meetings, between which hours are spent coding practice sets for reliability. Multiple coders must be trained in case one or more does not achieve reliability. After training, comprehensive coding of facial actions requires up to an hour to annotate a single minute of video, and a meaningful percentage of videos must be coded by two people to compute reliability (Bartlett, Hager, Ekman, & Sejnowski, 1999; Ekman & Friesen, 1978). In attempts to circumvent manual coding, researchers often rely on

less labor-intensive markers of emotional experience and expression, such as autonomic psychophysiology, but these methods provide information at a much narrower bandwidth, and are open to multiple interpretations (see, e.g., Beauchaine & Webb, 2017; Zisner & Beauchaine, 2016).

In efforts to address the labor intensity of manual coding, psychologists and computer scientists have devised automated methods of capturing facial expressions of emotion (e.g., Dys & Malti, 2016; Gadea, Aliño, Espert, & Salvador, 2015; Haines, Southward, Cheavens, Beauchaine, & Ahn, 2019; Sikka et al., 2015). In this paper, we describe innovations in computer-vision and machine learning, which offer a partial solution to impediments of manual coding. Automated facial expression coding (AFEC) involves two steps, including (a) using computer-vision to extract facial expression information from pictures or video streams, and (b) applying machine learning to map extracted facial expression information to prototype emotion ratings made by expert human judges (Cohn, 2010; Cohn & De la Torre, 2014). In their current state of development, AFEC models cannot capture the full complexity of human emotion, and therefore cannot be used to replace human coders in all studies. In other contexts, however, they code facial expressions with accuracy that is similar to highly skilled human coders. Although rarely used by developmental psychopathologists to date, AFEC models can evaluate and map development of specific emotional processes, including core Research Domain Criteria constructs of positive valence, negative valence, and arousal/regulation. Our objectives in writing this paper are to (a) describe AFEC models, including their historical validation vis-à-vis gold-standard human coding; (b) provide a critical analysis of what AFEC models can and cannot do in contemporary developmental psychopathology and emotion dysregulation research; and (c) present example analyses in which we compare AFEC with human ratings using a different coding scheme to evaluate emotion expression and emotion dysregulation in dyadic conflict discussions between mothers and their depressed, self-injuring, and typically developing daughters.

When AFEC models can be applied, they offer several advantages over human coding. Because they are mathematical, they always assign the same code(s) to an image or video, which eliminates an important source of error: intercoder disagreement. AFEC models can also leverage computer processing speed to code facial expressions in real time. Before describing AFEC models, we first review historical perspectives on emotion, including how facial expressions relate to socioemotional development.

Facial Expressions of Emotion: Historical Perspectives

In *Expression of Emotion in Man and Animals*, Darwin (1872) noted the importance of facial expressions for development and communication of emotions. In this foundational text, Darwin expanded Duchenne's (1990/1862) taxonomy of emotional expression by suggesting that facial expressions (a) map onto specific emotional states, which motivate adaptive behaviors; and (b) serve to regulate social interactions by allowing us to influence and understand others' behaviors and intentions. From this functionalist view, abnormalities in production of facial expressions result in ineffective conveyance of emotional information, whereas abnormalities in recognition of facial expressions result in misattribution of other's motivations, intentions, and emotional states. The abilities to accurately produce and recognize facial expressions of emotion are therefore tightly linked to healthy socioemotional development (Ekman & Oster, 1979).

Since Darwin's work, multiple theories of emotion have guided facial expression research. Broadly speaking, these can be divided into discrete versus dimensional views.[1] Discrete emotion theories postulate that distinct patterns of facial expressions represent separable emotional states that motivate adaptive responses to environmental cues. In support of this perspective, people across cultures display similar facial expressions in specific contexts and motivational states (Chevalier-Skolnikoff, 1973). Such findings and others led discrete emotion theorists to develop the concept of *basic emotions*. Cross-cultural studies consistently yield six universal (basic) facial expressions of emotion, including happiness, sadness, anger, fear, disgust, and surprise (Ekman, 1992, 1993; Ekman et al., 1987). Of note, these expressions are also observed in many nonhuman primate species (Preuschoft & van Hooff, 1995).

Although the existence of basic, universally recognized facial expressions is still a topic of debate (Jack, Garrod, Yu, Caldara, & Schyns, 2012), the idea that facial expressions represent discrete emotional states remains strong (Ekman & Cordaro, 2011). Many foundational studies on development of both facial expression recognition (e.g., Camras, Grow, & Ribordy, 1983; Pollak et al., 2000; Widen & Russell, 2003) and production (e.g., Buck, 1975; Field, Woodson, Greenberg, & Cohen, 1982; Odom & Lemond, 1972; Yarczower, Kilbride, & Hill, 1979) assume that emotions are discrete, and emphasis on conceptualizing emotions as discrete continues in the literature today.

In contrast to discrete emotion theory, dimensional perspectives assume that emotions, as well as facial expressions, are best represented as points along multiple latent continua, including dimensions such as valence and arousal. Dimensional theories gained traction in the mid-20th century, when Schlosberg (1941) showed that people categorize facial expressions along continua, analogous to the way we sort hues on a color wheel despite a continuous gradient of color in nature. Schlosberg (1952) showed that a large portion of variance in how people categorize facial expressions is captured by a circular rating space comprising two bipolar dimensions: pleasantness–unpleasantness and attention–rejection. Later, Schlosberg (1954) added a third dimension to capture arousal. Many other dimensional models of emotion have emerged since, including Russell's circumplex model of valence and arousal (Russell, 1980; Russell & Bullock, 1985) and Watson and Tellegen's (1985) variation of the circumplex, which orthogonalizes positive and negative affect intensity. Across studies, facial expressions map most reliably onto valence and arousal dimensions (Smith & Ellsworth, 1985). In contrast to discrete emotions theory, far fewer studies take a dimensional approach to investigate development of emotion (e.g., Luterek, Orsillo, & Marx, 2005; Yoshimura, Sato, Uono, & Toichi, 2014). In the next section, we summarize development of facial expressions and their recognition.

Facial Expressions of Emotion and Development of Social Communication

Production of facial expressions

Infants exhibit facial expressions of emotion at birth. Immediately thereafter, their facial expressions are modified through social learning and other reinforcement mechanisms. By age 1 month,

---

1. Constructionist theories (e.g., Barrett, 2009) offer a third perspective on facial expressions of emotion. However, because dimensional models capture key components of constructionist models (e.g., *core affective* processes such as valence), we do not consider them given space constraints.

infants produce emotionally congruent facial expressions to a variety of social and nonsocial incentives, such as smiling at others who smile at them, and smiling when presented with primary reinforcers. These expressions are readily and accurately interpreted by adults (Izard, Huebner, Risser, McGinnes, & Dougherty, 1980). Children who are blind at birth exhibit the same facial expressions as others, but become less expressive than typically developing peers as they age, indicating that development of emotion expression is determined in part by social mechanisms (Goodenough, 1932; Thompson, 1941). In very early infancy, infants do not have the capacity to regulate their spontaneous facial expressions (Izard & Malatesta, 1987).

By age 3–5 years, children can voluntarily produce facial expressions more accurately than they can recognize them, and the degree to which they accurately produce facial expressions predicts how willing their peers are to play with them (Field et al., 1982). Environmental stressors such as physical abuse during this age period influence how accurately children produce certain facial expressions (Camras et al., 1988). As children are exposed to increasingly complex social situations during the transition from preschool to elementary school, they learn to downregulate spontaneous facial expressions (Buck, 1977). Children who show more downregulation (better volitional control) enjoy higher social status than their peers (Zivin, 1977). Voluntary control over facial expressions continues to improve into early adolescence, when accuracy reaches a natural peak that can further improve with training (Ekman, Roper, & Hager, 1980). Adolescents who do not exhibit age-appropriate regulation of facial expressions have more internalizing and externalizing symptoms than their peers (Keltner, Moffitt, & Stouthamer-Loeber, 1995). Furthermore, male externalizing children and adolescents show poor correspondence between facial expressions of emotion and physiological reactivity to others' emotions (Marsh, Beauchaine, & Williams, 2008).

As this discussion suggests, the ability to volitionally up- and downregulate facial expressions in response to contextual demands (termed *expressive flexibility*) is associated with a number of functional and social outcomes (Coifman & Almahmoud, 2017). Young adults who flexibly regulate their facial expressions show better adjustment to environmental stressors such as traumatic life events and sexual abuse, and experience fewer symptoms of psychopathology (Bonanno et al., 2007; Bonanno, Papa, Lalande, Westphal, & Coifman, 2004; Chen, Chen, & Bonanno, 2018; Westphal, Seivert, & Bonanno, 2010). In addition, college freshmen who show more positive facial expressions in response to peer rejection report less psychological distress than their peers (Coifman, Flynn, & Pinto, 2016). Given shifts from spontaneous to volitional control of facial expressions from infancy to adolescence, expressive flexibility may be a useful indicator of emotion regulation capacities, socioemotional competence, and resilience (Beckes, & Edwards, 2018; Bonanno & Burton, 2013).

Recognition of facial expressions

The ability to accurately recognize facial expressions of emotion develops at a slower pace than the ability to accurately produce them (Field & Walden, 1982). Nevertheless, infants can imitate adults' facial expressions at as young as age 12 days (Meltzoff & Moore, 1977). By age 5–7 months, infants begin to discriminate between happy, sad, fearful, and surprised expressions, yet they do so based on facial features (e.g., actions involving the eyes) rather than holistic facial expressions (Nelson, 1987). By age 1 year, infants extract meaning from facial expressions and modify their behaviors accordingly. For example, studies using the visual cliff paradigm show that infants reference their mother's facial expressions before crossing what appears to be a deep cliff. When mothers exhibit happy or fearful facial expressions, infants either cross or decide against crossing the apparent cliff, respectively (Sorce, Emde, Campos, & Klinnert, 1985).

As infants enter more complex social environments, they learn to associate specific facial expressions with emotion. For example, during intense emotional experiences such as crying and laughing, the orbicularis oculi (muscles surrounding the eyes, the *Duchenne* marker; Duchenne, 1990/1862) spontaneously contract to protect the eyes from increased pressure as blood rushes to the face. Because volitional control over these muscles is difficult even for adults, the Duchenne marker can be an involuntary indicator of positive or negative affect intensity (Messinger, Mattson, Mahoor, & Cohn, 2012). Young children learn to associate the Duchenne marker with more intense emotion by age 3–4 years (Song, Over, & Carpenter, 2016). This suggests that meaning attributed to the Duchenne marker is learned. In addition, preschoolers from low-socioeconomic status, high-crime neighborhoods can recognize facial expressions of fear more accurately than their nondisadvantaged peers (Smith & Walden, 1998). Thus, environments affect the rate at which children learn to associate specific facial expressions with discrete emotions.

Despite being able to categorize facial expressions of emotion, preschoolers often fail to account for context when interpreting facial expressions (Hoffner & Badzinski, 1989). However, as children begin to regulate their facial expressions in kindergarten into middle school, they learn to integrate facial expressions with contextual cues and prior experience, generating more accurate inferences about others' emotional states (Hoffner & Badzinski, 1989). Some research suggests that children who fail to learn accurate representations of facial expressions during this period show lower academic performance (Izard et al., 2001), consistent with accounts of shared neural representation of "cognition" and "emotion" in the brain (see Pessoa, 2008).

Of note, associations children learn between facial expressions and emotion are sensitive to environmental moderation. Children with histories of neglect or abuse are less accurate in recognizing basic emotions from facial expressions than their nonabused peers, and the degree of inaccuracy is related to their social competence (Camras et al., 1983, 1988; Pollak et al., 2000). Furthermore, physically abused children are biased toward interpreting ambiguous facial expressions as angry (Pollak et al., 2000), and require less perceptual information to accurately recognize anger from facial expressions (Pollak & Sinha, 2002).

During the transition from late childhood to adolescence, facial expression recognition abilities solidify and are slowly refined into adulthood (Thomas, De Bellis, Graham, & LaBar, 2007). Because there is little qualitative change in recognition abilities from childhood to adolescence, a history of impaired facial expression recognition can strongly affect adolescents' social and behavioral function. For example, adolescent boys with early onset conduct disorder show deficits in recognizing facial expressions of happiness, anger, and disgust, and boys with psychopathic traits show more impairment in recognizing fear, sadness, and surprise (Blair & Coles, 2000; Fairchild, Van Goozen, Calder, Stollery, & Goodyer, 2009). Adolescent girls with borderline personality disorder require more perceptual information than their typically developing peers to accurately judge facial expressions of happiness and anger (Robin et al., 2012). Among

typically developing children, facial expressions are used to reinforce and punish social behaviors; inaccurate facial expression recognition during the transition from childhood to adolescence can therefore lead to repetition of socially maladaptive behaviors that are not corrected over time (Blair, 1995).

As this abbreviated review shows, accurate decoding of and effective volitional control over facial expressions are integral to healthy socioemotional development. Studying facial expressions —including their dynamics in social interactions—has therefore long been of interest to developmental psychopathologists. In the next section, we discuss historical approaches to facial coding, before turning to advantages and disadvantages of AFEC models.

Historical Approaches to Facial Expression Coding

Manual coding

Traditional methods of facial coding fall into two broad categories (Cohn & Ekman, 2005), including (a) message judgment approaches (*message based*), and (b) measure of sign vehicles approaches (*sign based*). Message-based approaches ascribe meaning to facial expressions (e.g., "*is this an angry facial expression?*"), whereas sign-based approaches describe observable facial actions that comprise an expression (e.g., "*are they furrowing their brow?*"). Because we make inferences about the meaning of facial expressions based on facial actions, the distinction between message- and sign-based approaches is more blurry than sharp. Nevertheless, the approaches are unique in what they require observers to do (Cohn & Ekman, 2005). Whereas message-based coding requires observers to make holistic inferences about emotional states from observable facial expressions, sign-based coding involves only description, not interpretation.

Message-based approaches are used most often in psychology. They are more flexible and easier to implement than sign-based approaches and allow researchers to make inferences about specific emotional states. Furthermore, most research on development of facial expressions leverages message-based protocols to link facial expressions to emotion, so message-based protocols play a crucial role in our understanding of emotional development (see the next section for details). Nevertheless, message-based protocols have a major drawback. They are often created and implemented by individual labs, so there are no widely agreed-upon methods to objectively assign emotion ratings to facial expressions. For example, some labs code judgments of expressed valence intensity on Likert scales (e.g., Bonanno et al., 2004), whereas others code discrete emotions (e.g., Gross & Levenson, 1993). Without a standard approach, it is difficult to compare findings across sites. As noted by the National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria (2016), this is a major impediment to better understanding emotional processes. Even within labs, message-based protocols require human coders to make subjective judgments on specific emotions that facial expressions may represent, so there is no guarantee that different coding teams who are trained on the same protocol will use the same rules to make judgments.

In contrast to the lab-specific nature of message-based coding, there are a small number of highly standardized sign-based protocols (Cohn & Ekman, 2005; Ekman & Rosenberg, 2005). Among these, the Facial Action Coding System (FACS; Ekman & Friesen, 1978) is most widely used. FACS comprises approximately 33 anatomically based action units (AUs), which together can describe virtually any facial expression. For example, AU 12 indicates an oblique rotation of the lip corners (e.g., during a smile), and AU 6 denotes the Duchenne marker (i.e., crow's feet, contracting muscles around eyes). FACS was developed from a basic emotion perspective, so relations between AUs and basic facial expressions of emotion are well understood. For example, facial expressions of happiness typically involve AUs 6, 12, and sometimes 25 (parting of the lips), and people perceive smiles as more genuine when these AUs are activated simultaneously (Korb, With, Niedenthal, Kaiser, & Grandjean, 2014). Although FACS is standardized and therefore more generalizable than message-based protocols, FACS training takes upward of 100 hr, and just minutes of video require multiple hours to code (Cohn & Ekman, 2005). For these reasons, FACS coding can be cost-prohibitive, and is difficult to apply to dynamic systems, such as fast-moving dyadic exchanges during which participants influence one another's emotional expressions (cf. Crowell et al., 2017).

Automated coding

Automated facial expression coding (AFEC) offers a partial solution to the labor intensity of human coding (Cohn, 2010; Fasel & Luettin, 2003). AFEC combines image recognition with machine learning—an approach with solid foundations in other disciplines. Journals are now dedicated to human face and gesture recognition, and worldwide competitions to develop more accurate AFEC models are held each year (e.g., Valstar et al., 2016). Furthermore, there are many standardized data sets that AFEC researchers use to develop and test new emotion-detection models. This is consistent with broader efforts in the field to standardize tasks and measures used to assess emotion across labs (National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria, 2016), and has accelerated the pace of AFEC model development over the past few years (see Krumhuber, Skora, Küster, & Fou, 2016). Because the literature on AFEC is large and technically complex, we provide a conceptual overview of AFEC model development, and we selectively cite research to emphasize progress made on discrete and dimensional AFEC models over the past two decades.

To extract facial expression information from images and videos (i.e., time-series of images), AFEC models typically proceed by (a) detecting a face, (b) extracting pixel values from the face and translating them into features using a variety of different methods, and (c) determining—based on detected features—whether the face is making a particular movement or displaying a specific expression. It is important to note, however, that processes used to detect facial features vary widely across applications. For example, some studies use Gabor patch filtering methods, whereas others use deep machine learning, which automatically detects relevant features after extensive training. In this article, we use the latter approach. Full description of these processes, which are illustrated in Figure 1, is beyond the scope of this paper. Interested readers are referred to recent reviews (Cohn, 2010; Zhao & Zhang, 2016).

To generate valid ratings of emotion using machine learning applications of AFEC, researchers either (a) first have human coders annotate a data set using either message- or sign-based approaches, or (b) use a standardized reference data set. Next, a machine learning algorithm builds a model that predicts human coders' ratings given features extracted using the above listed steps. As with all machine learning approaches, AFEC is validated
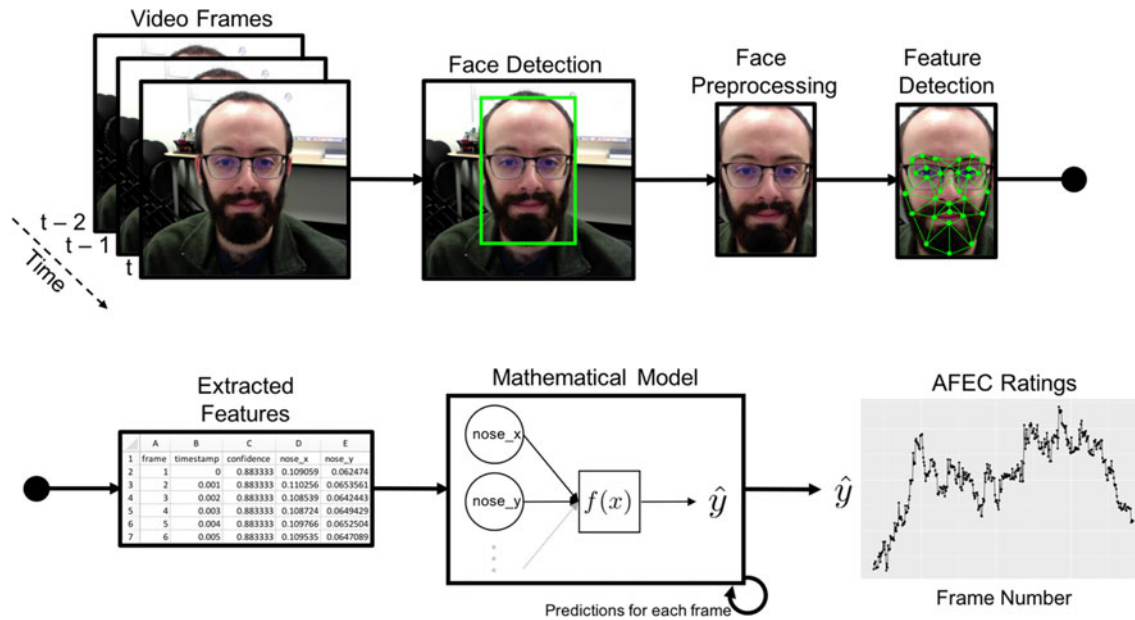
Figure 1. Steps in AFEC modeling. For each video frame (or image), AFEC uses computer-vision technology to detect a face, which is then preprocessed (cropped). Computer-vision then identifies a set of features on the preprocessed face (here represented by an *Active Appearance Model* that matches a statistical representation of a face to the detected face), which are tabulated and entered into a mathematical model. This model is developed to detect some aspect of emotion from a previously coded set of data (e.g., presence of a discrete emotion, valence intensity, FACS AUs, etc.), and outputs these emotion ratings based on extracted facial features of the new data. Ratings are output for each frame, resulting in a rich time series of emotional expressions. Of note, AFEC models can perform these steps in real time. In some cases, there can be a two-step mathematical model (e.g., FACS detection and then emotion detection given FACS AUs).

by checking how accurately it codes facial expressions on data that the model was not fit to. Therefore, training data used to develop AFEC models must be considered toward understanding how well the models generalize. For example, a model trained only on Caucasian participants may not generalize to all racial groups.

After validation, AFEC models can be shared across sites to code new data sets in real time. By detecting objective facial movements (AUs or other features), which are translated to affect ratings by mathematical models, AFEC combines characteristics of both message-based and sign-based approaches to facial coding. Next, we review progress made on AFEC models that detect discrete and dimensional features of facial expressions, and we describe commercial and open-source software that may be useful to interested researchers.

Discrete perspective

Most early research on AFEC focused on classifying facial expressions using six basic emotions, including happiness, sadness, surprise, anger, fear, and disgust (neutral was sometimes added). Given adequate image quality, AFEC models have long been able to classify basic emotions with up to 90% accuracy (see Fasel & Luettin, 2003).[2] More contemporary models classify basic emotions with up to 97% accuracy (Lopes, de Aguiar, De Souza, & Oliveira-Santos, 2017). With six basic emotions, such accuracy rates far exceed chance, and often exceed accuracy levels of human coders (Isaacowitz et al., 2007).

Although less work addresses AFEC coding of complex emotions (e.g., shame, boredom, interest, pride, etc.), available research indicates far lower accuracy. Humans can interpret an inordinately large number of facial expressions (Parrott, 2000). Some validated data sets used in emotion research contain over 400 distinct facial expressions and related mental states (e.g., doubt, hunger, reasoning, etc.; see Golan, Baron-Cohen, & Hill, 2006). However, in their current state of development, AFEC models cannot accommodate this level of expressive complexity (Adams & Robinson, 2015). AFEC models show at best modest success in concurrently classifying up to 18 distinct emotions from different combinations of facial expressions (47% accuracy; chance = 6%; Adams & Robinson, 2015).

Many evaluations of AFEC models compare their performance to human codings of FACS AUs (Ekman & Friesen, 1978). Because basic emotion coding relies in part on detection of specific combinations of FACS AUs, AFEC models have progressed in conjunction with FACS studies of basic emotions. Early AFEC models could detect 16–17 AUs with high accuracy (93%–95%; Kotsia & Pitas, 2007; Tian, Kanade, & Cohn, 2001), but data sets used to create and test these models were relatively small and therefore of limited generalizability.

In more recent literature, AFEC developers have stopped reporting simple accuracy rates, which do not correct for different base rates of emotions, and instead report true positives, false positives, true negatives, false negatives, and combinations of these metrics (see Footnote 1). The most widely used metric for comparing FACS AU detection across different AFEC models is the F1 score, which is a single number that takes into account precision and recall.[3] Like Cohen's kappa, F1 is useful when certain

---

2. Accuracy is defined as $\frac{TP+TN}{N}$, where TP is the number of true positives, TN is the number of true negatives, and N is the total number of images or videos coded. Note that this collapses across all emotions. AFEC researchers typically do not report kappa (κ) statistics, which are commonly used to assess agreement for categorical coding in developmental psychopathology research. We refer interested readers to confusion matrices within cited papers for complete information on human–computer agreement.

3. $F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$; $precision = \frac{true\ positives}{true\ positives + false\ positives}$; $recall = \frac{true\ positives}{true\ positives + false\ negatives}$. An F1 score of 1.0 indicates perfect precision and recall, whereas a score of 0 indicates no true positives.

AUs are expressed more often than others. For example, a person might respond to a particular video stimulus with AU 12 (associated with happiness) 90% of the time and AU 9 (associated with disgust) 5% of the time. If an AFEC model correctly classifies 95% of AU 12 expressions but only 50% of AU 9 expressions, overall accuracy will be high because the base rate of AU 12 is higher than the base rate of AU 9.

This change in accuracy reporting makes it challenging to quantify progress in automated FACS coding. Nevertheless, clear advances have been made. Validation studies of AFEC software (details below) indicate accuracy rates for contemporary models ranging from 71% to 97%, with F1 scores ranging from .27 to .90, depending on specific AUs (Lewinski, den Uyl, & Butler, 2014). Large ranges in F1 scores for AU detection within and across studies suggests that researchers should exercise caution when interpreting single AFEC-detected AUs, and always report either F1 or kappa. As might be surmised from our previous discussion, (a) higher F1 values are obtained when fewer AUs are used to detect basic as opposed to complex emotions, and (b) much more research is needed before AFEC models can replace human FACS coders.

Dimensional perspective

Dimensional AFEC models, which we provide an example of below, have also gained attention over the past decade (Gunes, Schuller, Pantic, & Cowie, 2011). Dimensional coding requires AFEC models to estimate *intensities* of facial expression dimensions (e.g., positive affect intensity on a scale of 0 to 10). Because dimensional AFEC models output continuous ratings, their accuracy is measured by correlations between their intensity ratings and those averaged across multiple human coders who manually annotate valence, arousal, and other affective dimensions. Early dimensional AFEC models showed reasonable correspondence with human coders, yielding correlations ranging from .57 to .58 for valence and .58 to .62 for arousal (Kanluan, Grimm, & Kroschel, 2008). More recent models yield correlations ranging from .60 to .71 for valence, .89 for positive affect intensity, and .76 for negative affect intensity (Haines et al., 2018; Mollahosseini, Hasani, & Mahoor, 2017; Nicolaou, Gunes, & Pantic, 2011).

Of note, some studies are validated on video recordings of seated laboratory participants who react to emotionally evocative images (Haines et al., 2018), whereas others are validated on images collected "in the wild" (e.g., video found on the internet; Mollahosseini et al., 2017). In the wild recordings present challenges given limited quality control over facial position, image resolution, image refresh rates, and so on. A related issue concerns quality of images collected during social interactions, where participants are free to move about. We are currently developing the capacity to code facial expressions in real time using head-mounted cameras in the lab. Although slightly intrusive, these cameras can collect precise, high-definition video in contexts that are much more ecologically valid than many other popular methods, such as EEG and magnetic resonance imaging, allow.

The upper bounds of correlations listed above may represent ceilings for AFEC models of arousal based on facial expressions (e.g., Mollahosseini et al., 2017; Nicolaou et al., 2011). However, further increments in accuracy may be possible by including vocal features (e.g., pitch or tone) in future AFEC models (Gunes & Pantic, 2010). Such findings have prompted AFEC researchers to include a variety of "response channels" in dimensional models of emotion, including vocalizations and various psychophysiological metrics (Gunes & Pantic, 2010; Ringeval et al., 2015). At present, more research is needed before models leveraging multiple response channels are used routinely by developmentalists.

Applications of AFEC in Psychology

The literature reviewed in preceding sections indicates that AFEC models cannot—at least in their current state of development—capture the full complexity of human emotion, as evaluated by highly trained FACS coders. In contrast, AFEC models are reasonably efficient at coding basic emotions, affective valence, and affect intensity. Accordingly, studies that evaluate developmental processes and socioemotional correlates of basic emotions, including valance and intensity, might benefit considerably from using AFEC, in terms of both time saved and the ability to model dynamic social interactions in real time.

To date, however, AFEC has enjoyed limited use in developmental research, although this has begun to change. Its first application in developmental psychology was a study showing that subtle changes in affective synchrony (simultaneously expressed emotions) among mother–infant dyads can be captured by AFEC (Messinger, Mahoor, Chow, & Cohn, 2009). Follow-up studies used AFEC to detect limited numbers of FACS AUs among infants. In one such study, AU 6 (the Duchenne marker) was associated with both increased positive and negative affect intensity (Mattson, Cohn, Mahoor, Gangi, & Messinger, 2013; Messinger et al., 2012). AFEC has also been used to determine how infants as young as age 13 months communicate emotion through facial expressions (Hammal, Cohn, Heike, & Speltz, 2015). Other applications include detecting changes in children's emotional expressions while lying (Gadea et al., 2015), while engaging in moral transgressions (Dys & Malti, 2016), and while experiencing pain (Sikka et al., 2015). Researchers have also used AFEC to develop interactive emotion–learning environments for children with autism (e.g., Gordon, Pierce, Bartlett, & Tanaka, 2014). Among adults, AFEC can differentiate between active versus remitted depression (Dibeklioğlu, Hammal, Yang, & Cohn, 2015; Girard & Cohn, 2015), genuine versus feigned pain (Bartlett, Littlewort, Frank, & Lee, 2014), and facial expression dynamics between those with versus without schizophrenia (Hamm, Kohler, Gur, & Verma, 2011).

AFEC software available to psychologists

In the past decade, several commercial and open-source AFEC software products have emerged. However, not all are validated, so we limit discussion to those described in published validation studies. Both FACET and AFFDEX are modules of the iMotions platform (iMotions, 2018). This interface accommodates data from multiple modalities (facial expressions, EMG, ECG, eye tracking). FACET is a commercial successor to the computer expression recognition toolbox (Littlewort, Whitehill, Wu, & Fasel, 2011). AFFDEX is based on the Affectiva Inc. (https://www.affectiva.com/) AFFDEX algorithm (e.g., McDuff, El Kaliouby, Kassam, & Picard, 2010). FaceReader is a module of Noldus (Noldus, 2018), which offers similar services to iMotions for integrating across modalities. Automated coding tools including FACET and FaceReader have been used in prior studies to automate AU coding in children as young as age 5 years (Dys & Malti, 2016; Gadea et al., 2015; Sikka et al., 2015).

FACET, AFFDEX, and FaceReader are commercial software systems designed for easy use with minimal technical expertise. All can process facial expressions in real time from a webcam, or from stored video data. All are capable of detecting basic emotions as well as 20 FACS AUs. FaceReader is the only commercial software that can detect nonbasic emotions and AU intensity, as opposed to AU presence. OpenFace is the only free, open-source AFEC software available (Baltrusaitis, Robinson, & Morency, 2016). Unlike commercial products, OpenFace only detects AUs, and its graphical interface cannot be used to create experiments or integrate facial expression analysis with other data modalities. All of these products allow users to export data into standard file formats (e.g., .csv), which can then be loaded into any statistical program. We refer interested readers to validation studies for further details (Baltrusaitis et al., 2016; Lewinski et al., 2014; Stöckli, Schulte-Mecklenbeck, Borer, & Samson, 2017). In the next section, we present a purposefully straightforward application of AFEC that shows its promise in research on emotion dysregulation. More complex analyses and technical descriptions can be found in recent articles (e.g., Haines et al., 2019).

An AFEC Example: Emotion Dysregulation During Mother–Daughter Conflict

To demonstrate the potential utility of AFEC for emotion dysregulation research, we apply an AFEC model to data collected as part of a previous study (Crowell et al., 2012, 2014, 2017). This study compared emotional valence and intensity in mother–daughter dyads during conflict discussions. Dyads included mothers with typically developing adolescent girls, mothers with self-injuring adolescent girls, and mothers with depressed but non-self-injuring adolescent girls. These data are well suited for AFEC analysis given (a) previous annotation with a different coding scheme focused on basic rather complex emotional processes (to assess convergent validity), (b) a broad developmental age range (to assess maturational effects), (c) dynamic dyadic expressions of emotion (to evaluate social-emotional dynamics), (d) the ability to compare such dynamics and their development across typically developing and clinical groups, and (e) a laboratory setting that included facial expressions with little occlusion. Before presenting AFEC results, we describe the sample, and summarize the original coding scheme and previously reported results. Of note, these data have not been coded previously with AFEC.

In total, 76 adolescent girls, ages 13–18 years, participated with their mothers. Among adolescents, 26 were recruited based on significant self-injury, 24 were recruited for depression without self-injury, and 24 were typically developing controls. Given space constraints, we refer readers to previous publications for more specific details about the sample (Crowell et al., 2012, 2014, 2017). For purposes of the current analyses, videos of conflict discussions between adolescents and their mothers were available for 20 self-injuring, 15 depressed, and 13 control participants ($N = 48$). Videos were unavailable for a sizable subset of participants because several DVDs became corrupted over time, and because video quality was insufficient to subject all DVDs to AFEC analysis. Given the current state of AFEC, we coded intensity of both positive and negative affective valence, and correspondence between mother and daughter expressions of affect and intensity. To evaluate convergent validity, we compare AFEC ratings to previously coded expressions of positive and negative affect using a well-validated coding scheme: the Family and Peer Process Code (Stubbs, Crosby, Forgatch, & Capaldi, 1998).

Of note, it is well known that emotional lability characterizes dyadic interactions in families of girls who self-harm (e.g., Crowell et al., 2013, 2014, 2017; Di Pierro, Sarno, Perego, Gallucci, & Madeddu, 2012; Hipwell et al., 2008; Muehlenkamp, Kerr, Bradley, & Adams Larsen, 2010), and that facial expressions play an important role in regulating dyadic social interactions (see extended discussion above).

Procedure

Mothers and adolescent girls engaged in a 10-min conflict discussion while being monitored with audio and video equipment. Before the discussion, mothers and daughters completed the Issues Checklist (Prinz, Foster, Kent, & O'Leary, 1979), a 44-item questionnaire that contains a variety of common conflict topics (e.g., privacy). All are rated for frequency (1 = never to 5 = very often) and intensity (0 = calm to 40 = intense). A trained research assistant chose a topic that best represented disagreement between mothers and daughters on frequency and intensity ratings of dyad members, but did not exceed an intensity rating of 20. After a discussion topic was selected, the research assistant read the following script:

> For the next 10 minutes, I would like you to discuss a topic that you both rated as a frequent area of disagreement. I will tell you the topic and then I will exit the room to restart the recording equipment. I will knock on the wall when it is time for you to start the discussion, and I will knock again in 10 minutes when the discussion is over. It is important for you to keep the conversation going for the full 10 minutes. Your discussion topic is [e.g., keeping the bedroom clean].

Coding

Initially, videos were coded manually by highly trained research assistants. For the present analyses, we applied an AFEC model. Both are described below. Of note, our objective is not to compare the coding systems directly, but rather to evaluate convergent validity. Some description of each is necessary to illustrate efficiency advantages of AFEC models.

Manual coding

Expressions of emotion for both dyad members were coded manually using the Family and Peer Process Code (FPPC; Stubbs et al., 1998). Full procedures used for FPPC coding are described in detail elsewhere (Crowell et al., 2012, 2013, 2017). Briefly, FPPC is a microanalytic coding scheme in which ratings are assigned to all verbal utterances made by dyad members. FPPC codes capture content (e.g., negative interpersonal), affect (e.g., distress and neutral), and direction (mother to daughter vs. daughter to mother) of utterances throughout the conversation. New codes are assigned whenever there is a transition in speaker, listener, verbal content, or affect. FPPC comprises 25 content codes for verbal behavior and 6 affect codes (3 negative, 1 neutral, and 2 positive), resulting in 75 combinations. Facial expressions of emotion are integral to the affect codes.

Painstaking and meticulous training is required for FPPC coders. Two research assistants coded the videos. Both received 15 hr of training per week for 3 months (180 hr total). Training included (a) learning the manual and codes, (b) coding and
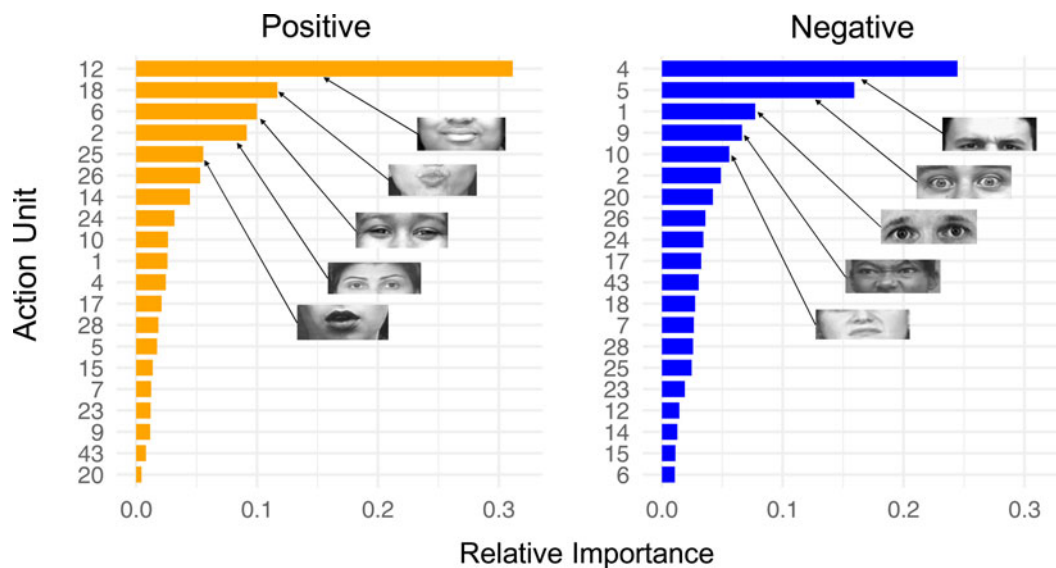
Figure 2. Relative importance weights for the 20 AUs used by the Random Forest model to generate ratings of positive and negative affect intensity. Relative importance is computed based on the change in standard deviation attributable to each given AU while integrating over all other AUs, which indicates the AUs that are most predictive of changes in positive/negative affect intensity. Images representing the five most important AUs are superimposed on each subplot. Image adapted with permission from Haines et al. (2019).

discussing practice tapes, and (c) coding criterion tapes without discussion to assess reliability. Coders ultimately achieved good reliability for both content ($\kappa = .76$) and affect ($\kappa = .69$). In addition, coders were required to achieve 10-key typing speed of 8,000 keystrokes per hour at 95% accuracy.

Following annotation, FPPC codes across the entire discussion were reduced to a single number for each dyad member from 0 (*highly positive utterances*) to 9 (*coercive or attacking utterances*). These scores were subcategorized as follows: high-level aversives (ratings 7–9), intermediate-level aversives (ratings 5–6), and low-level aversives (ratings 3–5). Next, transition probabilities were computed between mother and daughter aversive escalations as described by Crowell et al. (2013). These reflect probabilities of a dyad member escalating aversive behaviors (e.g., mother intermediate-level aversives followed by daughter high-level aversives = daughter escalation).

Automated coding

We applied an AFEC model developed in our previous work to code for positive and negative affect intensity on a scale from 1 (*no affect*) to 7 (*extreme affect*; Haines et al., 2018). This model detects positive and negative affect intensity separately rather than on a single valence continuum. It therefore accommodates seemingly paradoxical facial expressions that are both positive and negative in valence (e.g., Du, Tao, & Martinez, 2014; Watson & Tellegen, 1985), which are sometimes interpreted as disingenuous (e.g., Porter & ten Brinke, 2010). The age range of participants used to train the model spanned late adolescence to early adulthood, overlapping partly with ages of participants in the current study. However, FACET reliably codes facial expressions of emotion valence and intensity in children as young as age 5 years (e.g., Sikka et al., 2015). The model was trained on videos collected in a controlled laboratory setting that was quite similar to the setting in which mother–daughter dyads were recorded for the current study. Of note, our AFEC model does not use baseline or individual-subject corrections to achieve accurate

emotion ratings, which mitigates human error associated with determining individual-level neutral expressions.

Our AFEC application uses FACET to detect 20 different AUs.[4] In assigning positive and negative affect intensity ratings, these AUs are not weighted equally. Rather, they are weighted empirically using a Random Forest machine learning model that was validated on an independent sample (Haines et al., 2018). This empirically derived weighting procedure is depicted in Figure 2, which shows specific AUs that contribute more (and less) to positive and negative facial expressions. Videos were imported into iMotions (www.imotions.com), where they were analyzed using the FACET post hoc video processing feature. FACET outputs a time series of "evidence" ratings for presence of each of 20 AUs at a rate of 30 Hz. Evidence ratings reflect levels of confidence that a given AU is expressed. Ratings are rendered along a scale of roughly –16 to 16, where lower and higher ratings indicate lower and higher probabilities that an AU is present. To transform AU evidence ratings into positive and negative affect ratings, we split each individual's 10-min video into nonoverlapping 10-s segments and computed the normalized area under the curve (AUC) separately for each AU within each segment. AUCs were normalized by dividing by the total length of time participants' faces were detected throughout each 10-s segment. This ensures that resulting AUC values are not biased by varying face detection accuracies between segments (e.g., so segments with 90% face detection accuracy are not assigned a higher AUC than those with 70% accuracy). Normalized AUC scores for each AU are then entered as predictors into two pretrained Random Forest models, which generate positive and negative valence ratings. This process was iterated for each 10-s segment, resulting in a time series of 60 intensity ratings for positive and negative affect for each dyad member across the 10-min discussion.

---

4. FACET detects the following FACS AUs: 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 18, 20, 23, 24, 25, 26, 28, and 43.

In our validation study with an independent sample, the model showed average correlations with human ratings of .89 for positive affect intensity and .76 for negative affect intensity. Coding all 48 dyad members' facial expressions, which translates into 16 hr of raw video data, for 20 FACS AUs and subsequent positive and negative affect ratings, was executed overnight on a standard desktop PC.

Mother–daughter correspondence analyses

We assessed positive and negative affect correspondence by constructing multilevel models (MLM) in R, *lme4* (Bates, Maechler, Bolker, & Walker, 2015). MLM accommodates nested data structure (in this case, ratings nested within individuals). Because previous studies show that mothers play a primary role in "driving" adolescent emotion during conflict (Crowell et al., 2017), we specified MLMs as follows, which enabled us to evaluate between-groups differences in both (1) mothers' positive and negative affect, and (2) mother–daughter affect correspondence:

$$\text{Level 1}: \text{ mother affect}_{ij} = \pi_{0j} + \pi_{1j} \times (\text{daughter affect}_{ij}) + e_{jj}$$
$$\text{Level 2}: \pi_{0j} = \beta_{00} + \beta_{01} \times (C1) + \beta_{02} \times (C2) + r_{0j}$$
$$\pi_{1j} = \beta_{10} + \beta_{11} \times (C1) + \beta_{12} \times (C2) + r_{1j}$$

We included nested orthogonal contrasts to compare control dyads to both depressed and self-harm dyads (C1) and self-harm dyads to depressed dyads (C2). Mother and daughter affect indicate AFEC ratings, and i and j indicate time points and dyads, respectively. We fit separate models for positive and negative affect. Before running multilevel correspondence analyses, we smoothed the AFEC-coded positive and negative affect ratings using a simple moving average with a window size of 10. Moving averages act as low pass-filters (e.g., Smith, 2003) to dampen high-frequency noise (e.g., moment-to-moment variations in AFEC codes attributable to speaking). We conducted sensitivity analyses using varying widths (8–16) for the moving window; results were consistent across variations.

Results

Associations between FPPC and AFEC

Table 1 shows correlations between AFEC-coded positive and negative affect for each dyad member and the FPPC-derived values for (a) total aversiveness; (b) high, intermediate, and low aversives; and (c) probabilities of escalating. Among other findings, in dyads where mothers expressed more positive affect, daughters showed fewer escalations. Moreover, daughters who expressed more positive affect exhibited fewer high-level aversives and escalations and less overall aversiveness. Many of these correlations fall in the medium effect size range according to Cohen's (1992) standards. They are nevertheless notable given that (a) different coding schemes were used (FPPC vs. AFEC), (b) different "informants" coded behaviors (humans vs. computers), and (c) the algorithm used by AFEC was derived from an independent data set. Our findings therefore suggest at least some degree of convergence across methods. Specific relations between FPPC and AFEC codes also support validity of the AFEC model (e.g., mother positive affect was associated positively with daughter positive affect and negatively with daughter escalations).

Table 1. Correlations between AFEC and FPPC ratings

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Mother (+)	—												
2. Mother (−)	−.16	—											
3. Daughter (+)	.44**	−.12	—										
4. Daughter (−)	−.11	−.14	.05	—									
5. Mother HI	−.02	.08	−.03	.02	—								
6. Mother IN	.38**	−.25†	.27†	.03	−.16	—							
7. Mother LW	−.25†	.11	−.16	−.04	−.73**	−.56**	—						
8. Mother Tot. Ave.	.01	.05	−.01	.15	.93**	−.06	−.74**	—					
9. Mother Esc.	.15	.13	.25†	.01	−.15	−.24	.29*	−.26†	—				
10. Daughter HI	−.17	−.02	−.35*	−.04	.52**	.00	−.44**	.57**	−.82**	—			
11. Daughter IN	−.15	−.13	.20	.12	−.19	.03	.14	−.17	.20	−.48**	—		
12. Daughter LW	.22	.06	.33*	.01	−.52**	−.01	.45**	−.58**	.85**	−.97**	.26†	—	
13. Daughter Tot. Ave.	−.21	−.05	−.32*	.02	.53**	.00	−.44**	.59**	−.84**	.98**	−.35*	−.99**	—
14. Daughter Esc.	−.35*	.26†	−.31*	.03	−.22	−.65**	.64**	−.21	.08	.09	−.07	−.08	.09

Note: (+) represents AFEC positive affect intensity. (−) represents AFEC negative affect intensity. HI, high-level aversives. IN, intermediate-level aversives. LW, low-level aversives. Tot. Ave., total aversives. Esc., escalation. †$p \le .1$. *$p \le .05$. **$p \le .01$.
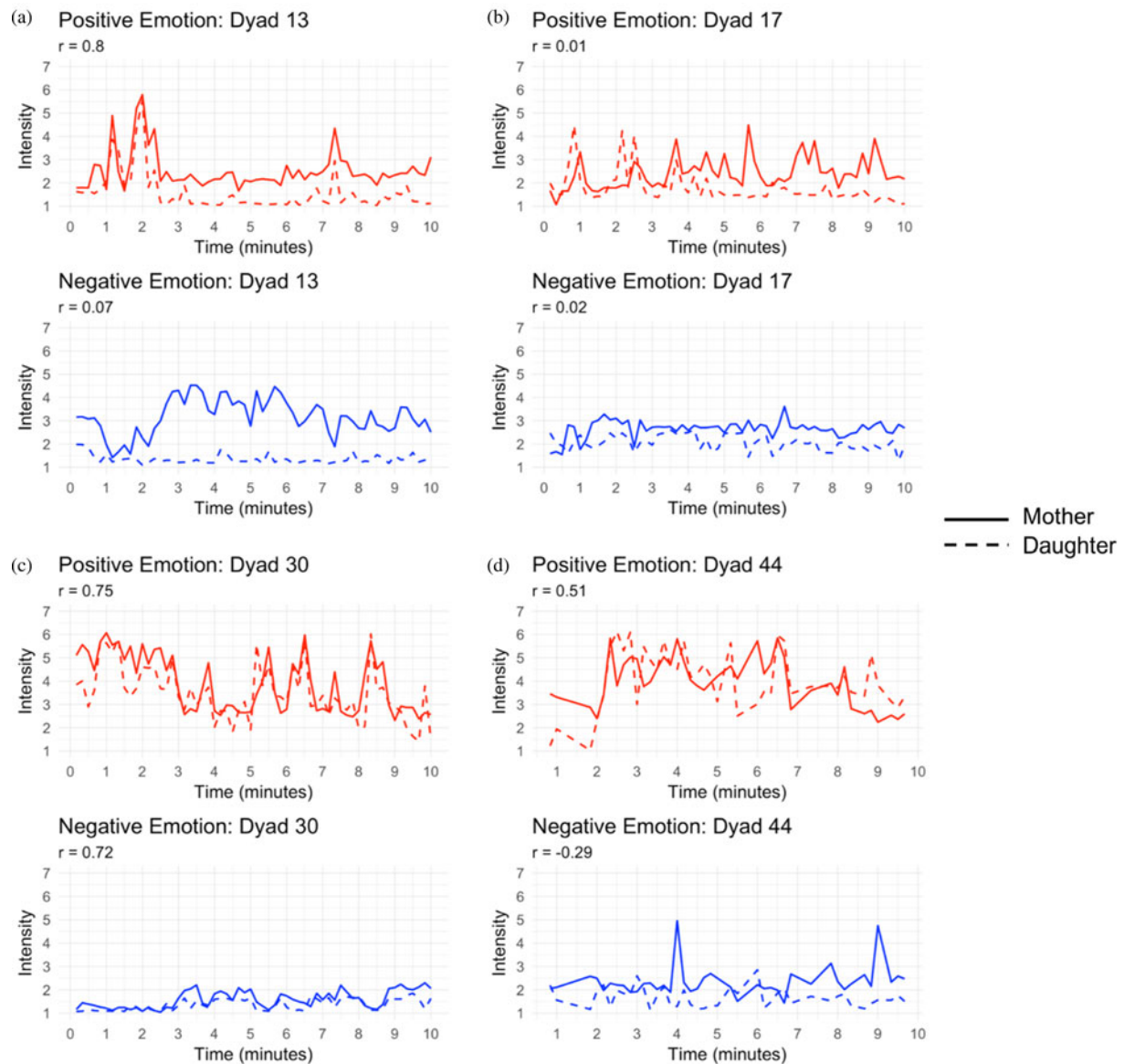
Figure 3. Heterogeneity of raw facial expression correspondence (correlations) among dyads. (a) Some dyads showed high correspondence of only positive affect, (b) some showed no correspondence of either positive or negative affect, (c) some showed high correspondence of both positive and negative affect, and (d) some showed positive correspondence of positive affect but negative correspondence of negative affect. Intensity ratings range from 1 (*no emotion*) to 7 (*extreme emotion*) within each 10-s interval across the 10-min discussion.

Average affect intensity

Next, we tested for pairwise differences in average positive and negative affect within and between mother–daughter dyads. Sample-wide, average positive and negative affect intensities for mothers over the 10-min discussion were 3.16 ($SD = 0.65$) and 2.56 ($SD = 0.73$), respectively. The difference between positive and negative affect was significant, $t(47) = 3.92$, $p < .001$, indicating that mothers expressed more intense positive than negative affect. Daughters also displayed higher positive affect ($M = 2.99$, $SD = 0.94$) than negative affect ($M = 1.73$, $SD = 0.51$), $t(47) = 8.35$, $p < .001$. Within dyads, mothers and daughters showed similar levels of positive affect intensity, $t(47) = -1.29$, $p = .20$, but daughters showed less negative affect than mothers, $t(47) = -6.06$, $p < .001$. Differences in average positive and negative affect between groups are reported in the correspondence analyses section below.

Dyadic correspondences of facial expressions

Before conducting correspondence analyses, we removed data from dyads for whom AFEC could code fewer than 15 co-occurring 10-s segments. This was necessary to estimate reliable correlations and resulted in exclusion of one dyad from the depressed group with only 3 segments of usable data.

Dyads exhibited considerable heterogeneity in raw facial expression dynamics across the discussion (see Figure 3). Some dyads showed high correspondence of positive affect but low correspondence of negative affect (Figure 3a), some showed low correspondence of both positive and negative affect (Figure 3b), some showed high correspondence of both positive and negative affect (Figure 3c), and others showed positive correspondence for positive affect but negative correspondence for negative affect (Figure 3d). Such differences in correspondence within dyads support the use of MLM to account for dyad-level variation.

Table 2. Summary of fixed slope and centered intercept effects from multilevel models

Variable	Coefficient	Standard error	t	p
Predicting mother positive affect				
For intercept				
Intercept	0.081	0.139	0.578	.566
C1	0.682	0.308	2.215	.032
C2	−0.019	0.327	−0.058	.954
For slope				
Intercept	0.609	0.086	7.037	<.001
C1	0.346	0.191	1.812	.077
C2	0.197	0.203	0.970	.338
Predicting mother negative affect				
For intercept				
Intercept	−0.042	0.127	−0.332	.742
C1	−0.165	0.278	−0.592	.557
C2	−0.370	0.299	−1.237	.223
For slope				
Intercept	−0.076	0.094	−0.813	.422
C1	−0.181	0.206	−0.879	.385
C2	0.505	0.222	2.274	.029

Note: *For Intercept* coefficients indicate differences in average mother positive affect, where the intercept is averaged across groups (mean-centered). C1 and C2 indicate contrasts of (a) control versus depressed and self-harm groups and (b) self-harm versus depressed groups, respectively. *For Slope* coefficients indicate differences in mother–daughter affect correspondence, where the intercept is averaged across groups (mean-centered). C1 and C2 indicate contrasts of (a) control versus depressed and self-harm groups and (b) self-harm versus depressed groups, respectively. Mother and daughter affect intensity ratings were smoothed (see Automated Coding section) and z-scored before modeling analyses.

Table 2 shows results from MLMs for both positive and negative affect correspondence. All three groups showed strong mother–daughter correspondence for positive affect intensity, yet mothers in control dyads showed higher average positive affect than mothers in both the depressed and self-harm dyads. No group differences in positive affect correspondence were observed. In contrast, only self-harm dyads showed negative affect correspondence, despite no group differences in average levels of mother negative affect.

Daughter age effects on correspondence
Finally, although the sample was not longitudinal, we examined daughter age effects on mother–daughter correspondence for both positive and negative affect. To do so, we separately regressed within-dyad positive and negative affect correspondence on daughters' ages. Although there was no age effect for negative affect correspondence, $r = .03$, $p = .83$, positive affect correspondence increased with daughters' age, $r = .41$, $p = .004$ (see Figure 4). Thus, for older girls, mother–daughter dyads showed greater synchrony of positive affect. Follow-up analyses indicated that this age effect was not attributable to group status, although power to detect any group difference was limited.

Interim discussion of findings

AFEC allowed us to code 16 continuous hours of facial expression data for positive and negative affect overnight. Even for a straight-forward AFEC application, this a small fraction of the hundreds of hours required to train a human coding team, achieve reliability, and code even modest segments of data. Furthermore, AFEC detected differences in temporal dynamics of emotion expression within and between dyads at a fine-grained level of analysis (10-s intervals; see Figure 3), and showed modest convergent validity with a different coding scheme. AFEC was also sensitive to dyadic correspondences in expressions of affect, and despite restricted power, to group differences in correspondence (see Table 2). Finally, AFEC detected age-related changes in positive affect correspondence. Mothers from control dyads exhibited more overall positive facial expressions throughout the conflict discussion compared to mothers from depressed and self-harm dyads, consistent with previous research. Nevertheless mother–daughter positive affect correspondence was strong across all groups.

In contrast to positive affect findings, mothers showed no group difference in overall negative facial expressions throughout the conflict discussion, yet correspondence of negative affect was greater for self-harm dyads than both typically developing and depressed dyads. Our finding of strong correspondence of negative facial expressions among self-harm dyads is consistent with previous analyses of these data using human coders (Crowell et al., 2013). Taken together, these findings indicate that AFEC is sensitive to emotion dynamics in mother–daughter interactions, and can be used to objectively code facial expressions in settings where human coding might be cost prohibitive. It is worth reemphasizing, however, that our findings are specific to coding of emotional valence and intensity, and do not suggest that AFEC can replace more complex coding schemes.

General Discussion and Implications for Emotion Dysregulation Research

Production of and accurate interpretation of facial expressions are critical for effective social communication and healthy socioemotional development. From infancy through adolescence, these skills improve as individuals attain increasing levels of social competence and self-regulation. Accordingly, development of facial expressions, correspondences between facial expressions and physiological reactivity, and dyadic concordance in facial expressions have been core interests of developmental psychopathologists for years. Nevertheless, our understanding of developing facial expressions and their associations with both typical and atypical emotional development have been hampered by the labor, cost, and time intensity of human coding. Although AFEC cannot yet capture the full complexity of human emotion, it can be used in its current state to code affective valence and intensity with greater efficiency than human coders. Moreover, some AFEC models (including the application illustrated here) are validated and embedded in a strong empirical research base, which enables researchers to standardize coding across studies and sites. Progression of affective science depends on establishing shared methods of assessment, data collection, and data reduction (National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria, 2016).

As our demonstration shows, existing applications of AFEC are sensitive to changes in affective valance and intensity dynamics, including individual differences, dyad-level differences, and
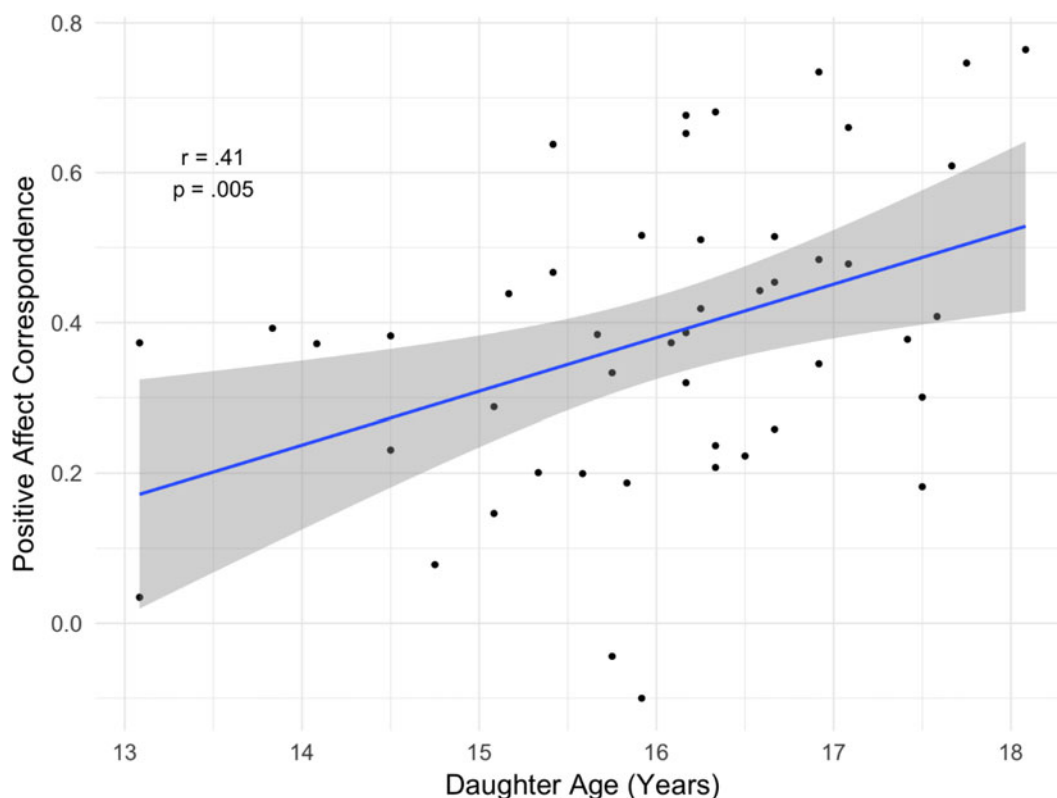
Figure 4. Sample-wide increase in dyadic correspondence of facial expressions of positive affect from ages 13 to 18 years (daughters).

group-level differences. One direction for future research concerns how emotion dynamics, including emotion dysregulation, develop in families, beginning at younger ages than we assessed. As noted above, AFEC accurately captures valence and intensity information from facial expressions of emotion in children as young as age 5 years, but how facial expressions shape emotional development within families is not fully understood. Future research can apply AFEC to assess maturation of parent–child emotional dynamics for typically developing children and children with various forms of psychopathology for whom extreme and lingering expressions of negative emotion are debilitating (see Beauchaine, 2015; Ramsook, Cole, & Fields-Olivieri, 2019). Of note, AFEC can code at higher temporal resolution than many human coding schemes, which often average ratings across epochs to reduce complexity. Moreover, computational advantages of AFEC hold potential to make more complicated analyses, such as dynamic evaluation of triadic interaction quality (cf. Bodner, Kuppens, Allen, Sheeber, & Ceulemans, 2018), far less cumbersome.

Given the labor intensity and cost of human coding, most studies of affective dynamics within families have been cross sectional. These studies reveal important group differences between families of typically developing children and adolescents versus families of adolescents with conduct problems, depression, and self-injury, among other conditions for which emotion dysregulation is a prominent feature (e.g., Bodner et al., 2018; Crowell et al., 2013, 2014; Snyder, Schrepferman, & St Peter, 1997). Evaluating family dynamics of emotion expression longitudinally is much more difficult because doing so requires research groups to train new coders for different time points in follow-up studies. Thus, labor intensity and cost are magnified over time. AFEC

can circumvent this problem by applying a well-validated model to all time points. This opens avenues for a wide range of applications, including cross-lag panel designs that evaluate directions of association between affect dynamics and various adjustment outcomes, both healthy and adverse.

Longitudinal assessment can also be applied in studies of within-individual and within-dyad correspondences between expressed emotion and physiological reactivity. For example, according to operant perspectives, parental negative reinforcement of emotional lability and physiological reactivity promote development of persistent emotion dysregulation among children, whereas validation of affective experiences and expression promotes emotional competence (e.g., Beauchaine & Zalewski, 2016; Crowell et al., 2013; Snyder et al., 1997). To date, however, evidence for this perspective derives primarily from cross-sectional studies. Future research that evaluates time-linked correspondences between emotional and physiological reactivity both within individuals and within dyads across successive longitudinal assessments could be invaluable in identifying (a) key developmental periods during which such patterns emerge, (b) shapes of trajectories in responding over time, and (c) intervention response, among other questions. Furthermore, AFEC holds promise in allowing researchers to explore concurrent and longitudinal relations between facial expressions and peripheral nervous system responding (e.g., cardiac and electrodermal), central nervous system responding (e.g., EEG and functional magnetic resonance imaging), and trajectories in neurohormonal responding (e.g., cortisol and salivary alpha amylase).

By combining AFEC with other technologies such as ecological momentary assessment (EMA), more naturalistic, outside-the-lab applications may also be possible. To date, smartphone-based

EMA has been used to assess parent–child interactions in children as young as 6 years (Li & Lansford, 2018), and computer games have been developed that use AFEC to train facial expression production abilities in children of the same age (Gordon et al., 2014). One possibility is to combine EMA and computer-based AFEC games to measure children's interpretation and use of facial expressions across different contexts over time. Such data-driven approaches may facilitate our search for prospective predictors of emotion dysregulation, a key target for prevention efforts in developmental psychopathology (Beauchaine, 2015).

A particularly appealing aspect of AFEC is its ability to code in real time from live streams of facial expressions. Real-time coding offers a number of advantages over human coding. In one real-time application, Gordon et al. (2014) leveraged real-time AFEC feedback to teach children with autism spectrum disorder to better produce facial expressions and recognize emotion from others' facial expressions. In the future, therapeutic applications might include using AFEC to evaluate changes in interpretations and reactions to others' facial expressions among individuals with abuse histories. Attributional biases among victims of abuse persist across development, with adverse implications for long-term function (Pollak et al., 2000; Pollak, & Sinha, 2002). Other potential applications include real-time AFEC coding during cooperative social exchange games for which performance is influenced heavily by others' facial expressions (Reed, DeScioli, & Pinker, 2014). Skills required to play such games have clear implications for development of social competence (Gallup, O'Brien, & Wilson, 2010).

Because AFEC is relatively new—at least to psychologists—there are currently no "best practices" for analyzing facial expression data after they are collected. However, psychologists have analyzed data from dyadic interactions for decades, and many of the same techniques apply to analyzing raw AFEC data (see, e.g., Bodner et al., 2018). We expect the field will develop best practices as AFEC becomes more popular. We are currently collaborating with another lab to develop a toolbox for analyzing facial expression data collected from various software products mentioned above. This will be analogous to toolboxes currently available for analyzing neuroimaging data. We hope this endeavor makes AFEC models more accessible to developmental and other behavioral scientists.

Finally, it is worth noting that not all AFEC models perform equally, and choices of software can lead to different conclusions from seemingly equivalent analyses. However, we do not think such differences should discourage researchers from leveraging potentially valuable features of AFEC. Common software packages for analyzing neuroimaging data (e.g., SPM and FSL) yield varying results when analyzing the same data, and even within a single software package one's choice of preprocessing steps affects estimates of blood oxygen level dependent responding (Carp, 2012). Despite these issues, functional magnetic resonance imaging studies continue to provide critical insights into human behavior and its development. Moving forward, it is important that researchers interested in AFEC technologies use validated models and provide details on preprocessing steps used.

References

Adams, A., & Robinson, P. (2015). *Automated recognition of complex categorical emotions from facial expressions and head motions*. Paper presented at the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, China.

Baltrusaitis, T., Robinson, P., & Morency, L.-P. (2016). *OpenFace: An open source facial behavior analysis toolkit*. Paper presented at the 2016 IEEE Winter Conference on Applications of Computer Vision, Lake Placid, NY.

Barrett, L. F. (2009). The future of psychology: Connecting mind to brain. *Perspectives on Psychological Science*, 4, 326–339. doi:10.1111/j.1745-6924.2009.01134.x

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, 47, 644–675. doi:10.2307/3094912

Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology*, 36, 253–263. doi:10.1017/S0048577299971664

Bartlett, M. S., Littlewort, G. C., Frank, M. G., & Lee, K. (2014). Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology*, 24, 738–743. doi:10.1016/j.cub.2014.02.009

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01

Beauchaine, T. P. (2015). Future directions in emotion dysregulation and youth psychopathology. *Journal of Clinical Child and Adolescent Psychology*, 44, 875–896. doi:10.1080/15374416.2015.1038827

Beauchaine, T. P., & Haines, N. (in press). Functionalist and constructionist perspectives on emotion. In T. P. Beauchaine & S. E. Crowell (Eds.), *The Oxford handbook of emotion dysregulation*. New York: Oxford University Press.

Beauchaine, T. P., & Webb, S. J. (2017). Developmental processes and psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (4th ed., pp. 495–510). New York: Cambridge University Press.

Beauchaine, T. P., & Zalewski, M. (2016). Physiological and developmental mechanisms of emotional lability in coercive relationships. In T. J. Dishion & J. J. Snyder (Eds.), *The Oxford handbook of coercive relationship dynamics* (pp. 39–52). New York: OXford University Press.

Beckes, L., & Edwards, W. L. (2018). Emotions as regulators of social behavior. In T. P. Beauchaine & S. E. Crowell (Eds.), *The Oxford handbook of emotion dysregulation*. New York: Oxford University Press.

Blair, R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57, 1–29. doi:10.1016/0010-0277(95)00676-P

Blair, R. J. R., & Coles, M. (2000). Expression recognition and behavioural problems in early adolescence. *Cognitive Development*, 15, 421–434. doi:10.1016/S0885-2014(01)00039-9

Bodner, N., Kuppens, P., Allen, N. B., Sheeber, L. B., & Ceulemans, E. (2018). Affective family interactions and their associations with adolescent depression: A dynamic network approach. *Development and Psychopathology*, 30, 1459–1473. doi:10.1017/S0954579417001699

Bonanno, G. A., & Burton, C. L. (2013). Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on Psychological Science*, 8, 591–612. doi:10.1177/1745691613504116

Bonanno, G. A., Colak, D. M., Keltner, D., Shiota, M. N., Papa, A., Noll, J., … Trickett, P. K. (2007). Context matters: The benefits and costs of expressing positive emotion among survivors of childhood sexual abuse. *Emotion*, 7, 824–837. doi:10.1037/1528-3542.7.4.824

Bonanno, G. A., Papa, A., Lalande, K., Westphal, M., & Coifman, K. (2004). The importance of being flexible: The ability to both enhance and suppress emotional expression predicts long-term adjustment. *Psychological Science*, 15, 482–487. doi:10.1111/j.0956-7976.2004.00705.x

Buck, R. (1975). Nonverbal communication of affect in children. *Journal of Personality and Social Psychology*, 31, 644–653. doi:10.1037/h0077071

Buck, R. (1977). Nonverbal communication of affect in preschool children: Relationships with personality and skin conductance. *Journal of Personality and Social Psychology*, 35, 225–236. doi:10.1037/0022-3514.35.4.225

Campos, J. J., Mumme, D. L., Kermoian, R., & Campos, R. G. (1994). A functionalist perspective on the nature of emotion. *Monographs of the Society for*

Research in Child Development, 59, 284–303. doi:10.1111=j.1540-5834.1994.tb01289.x

Camras, L. A., Grow, J. G., & Ribordy, S. C. (1983). Recognition of emotional expression by abused children. *Journal of Clinical Child Psychology*, 12, 325–328. doi:10.1207/s15374424jccp1203_16

Camras, L. A., Ribordy, S., Hill, J., Martino, S., Spaccarelli, S., & Stefani, R. (1988). Recognition and posing of emotional expressions by abused children and their mothers. *Developmental Psychology*, 24, 776–781. doi:10.1037//0012-1649.24.6.776

Carp, J. (2012). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fMRI experiments. *Frontiers in Neuroscience*, 6, 149. doi:10.3389/fnins.2012.00149

Chen, S., Chen, T., & Bonanno, G. A. (2018). Expressive flexibility: Enhancement and suppression abilities differentially predict life satisfaction and psychopathology symptoms. *Personality and Individual Differences*, 126, 78–84. doi:10.1016/j.paid.2018.01.010

Chevalier-Skolnikoff, S. (1973). Facial expression of emotion in nonhuman primates. In P. Ekman (Ed.), *Darwin and facial expression* (pp. 11–83). New York: Oxford University Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. doi:10.1037/0033-2909.112.1.155

Cohn, J. F. (2010). Advances in behavioral science using automated facial image analysis and synthesis [social sciences]. *IEEE Signal Processing Magazine*, 27, 128–133. doi:10.1109/MSP.2010.938102

Cohn, J. F., & De la Torre, F. (2014). Automated face analysis for affective. In R. A. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing* (pp. 131–150). New York: Oxford University Press.

Cohn, J. F., & Ekman, P. (2005). Measuring facial action. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of nonverbal behavior research methods in the affective sciences* (pp. 9–64). New York: Oxford University Press.

Coifman, K., & Almahmoud, S. (2017). Emotion flexibility and psychological risk and resilience. In U. Kumar (Ed.), *The Routledge international handbook of psychosocial resilience* (pp. 1–14). London: Routledge.

Coifman, K. G., Flynn, J. J., & Pinto, L. A. (2016). When context matters: Negative emotions predict psychological health and adjustment. *Motivation and Emotion*, 40, 602–624. doi:10.1007/s11031-016-9553-y

Crowell, S. E., Baucom, B. R., McCauley, E., Potapova, N. V., Fitelson, M., Barth, H., … Beauchaine, T. P. (2013). Mechanisms of contextual risk for adolescent self-injury: Invalidation and conflict escalation in mother–child interactions. *Journal of Clinical Child and Adolescent Psychology*, 42, 467–480. doi:10.1080/15374416.2013.785360

Crowell, S. E., Baucom, B. R., Yaptangco, M., Bride, D., Hsiao, R., McCauley, E., & Beauchaine, T. P. (2014). Emotion dysregulation and dyadic conflict in depressed and typical adolescents: Evaluating concordance across psychophysiological and observational measures. *Biological Psychology*, 98, 50–58. doi:10.1016/j.biopsycho.2014.02.009

Crowell, S. E., Beauchaine, T. P., Hsiao, R. C., Vasilev, C. A., Yaptangco, M., Linehan, M. M., & McCauley, E. (2012). Differentiating adolescent self-injury from adolescent depression: Possible implications for borderline personality development. *Journal of Abnormal Child Psychology*, 40, 45–57. doi:10.1007/s10802-011-9578-3

Crowell, S. E., Butner, J., Wiltshire, T. J., Munion, A. K., Yaptangco, M., & Beauchaine, T. P. (2017). Evaluating emotional and biological sensitivity to maternal behavior among depressed and self-injuring adolescent girls using nonlinear dynamics. *Clinical Psychological Science*, 5, 272–285. doi:10.1177/2167702617692861

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: Murray.

Dibeklioğlu, H., Hammal, Z., Yang, Y., & Cohn, J. F. (2015). *Multimodal detection of depression in clinical interviews*. Paper presented at the 2015 ACM on International Conference on Multimodal Interaction (pp. 307–310), Seattle, WA.

Di Pierro, R., Sarno, I., Perego, S., Gallucci, M., & Madeddu, F. (2012). Adolescent nonsuicidal self-injury: The effects of personality traits, family relationships and maltreatment on the presence and severity of behaviours.

European Child and Adolescent Psychiatry, 21, 511–520. doi:10.1007/s00787-012-0289-2

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *Journal of Primary Prevention*, 28, 67–91. doi:10.1007/s10935-007-0081-0

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences, 111*, E1454–E1462. doi:10.1073/pnas.1322355111

Duchenne, G. (1990). *The mechanism of human facial expression*. Translated by R. A. Cuthbertson. New York: Cambridge University Press. (Originally published as *Mecanisme de la Physionomie Humaine* in 1862.)

Dys, S. P., & Malti, T. (2016). It's a two-way street: Automatic and controlled processes in children's emotional responses to moral transgressions. *Journal of Experimental Child Psychology*, 152, 31–40. doi:10.1016/j.jecp.2016.06.011

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169–200. doi:10.1080/02699939208411068

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, 48, 384–392. doi:10.1037/0003-066X.48.4.384

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, 3, 364–370. doi:10.1177/1754073911410740

Ekman, P., & Friesen, W. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., … Ricci-Bitti, P. E. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53, 712–717. doi:10.1037/0022-3514.53.4.712

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology*, 30, 527–554. doi:0.1146/annurev.ps.30.020179.002523

Ekman, P., Roper, G., & Hager, J. C. (1980). Deliberate facial movement. *Child Development*, 51, 886–891. doi:10.2307/1129478

Ekman, P., & Rosenberg, E. L. (Eds.). (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system* (2nd ed.). New York: Oxford University Press.

Fairchild, G., Van Goozen, S. H. M., Calder, A. J., Stollery, S. J., & Goodyer, I. M. (2009). Deficits in facial expression recognition in male adolescents with early-onset or adolescence-onset conduct disorder. *Journal of Child Psychology and Psychiatry*, 50, 627–636. doi:10.1111/j.1469-7610.2008.02020.x

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition*, 36, 259–275. doi:10.1016/S0031-3203(02)00052-3

Field, T. M., & Walden, T. A. (1982). Production and discrimination of facial expressions by preschool children. *Child Development*, 53, 1299–1311. doi:10.2307/1129020

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expression by neonates. *Science*, 218, 179–181. doi:10.1126/science.7123230

Gadea, M., Aliño, M., Espert, R., & Salvador, A. (2015). Deceit and facial expression in children: The enabling role of the "poker face" child and the dependent personality of the detector. *Frontiers in Psychology*, 6, 6. doi:10.3389/fpsyg.2015.01089

Gallup, A. C., O'Brien, D. T., & Wilson, D. S. (2010). The relationship between adolescent peer aggression and responses to a sequential Prisoner's Dilemma game during college: An explorative study. *Journal of Social, Evolutionary, and Cultural Psychology*, 4, 277–289. doi:10.1037/h0099283

Girard, J. M., & Cohn, J. F. (2015). Automated audiovisual depression analysis. *Current Opinion in Psychology*, 4, 75–79. doi:10.1016/j.copsyc.2014.12.010

Golan, O., Ashwin, E., Granader, Y., McClintock, S., Day, K., Leggett, V., & Baron-Cohen, S. (2009). Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces. *Journal of Autism and Developmental Disorders*, 40, 269–279. doi:10.1007/s10803-009-0862-9

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders*, 36, 169–183. doi:10.1007/s10803-005-0057-y

Goodenough, F. L. (1932). Expression of the emotions in a blind-deaf child. *Journal of Abnormal and Social Psychology*, 27, 328–333. doi:10.1037/h0076099

Gordon, I., Pierce, M. D., Bartlett, M. S., & Tanaka, J. W. (2014). Training facial expression production in children on the autism spectrum. *Journal of Autism and Developmental Disorders*, 44, 2486–2498. doi:10.1007/s10803-014-2118-6

Gross, J. J., & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology*, 64, 970–986. doi:10.1037/0022-3514.64.6.970

Gunes, H., & Pantic, M. (2010). Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions*, 1. doi:10.4018/jse.2010101605

Gunes, H., Schuller, B., Pantic, M., & Cowie, R. (2011). Emotion representation, analysis and synthesis in continuous space: A survey. *Proceedings of the Automatic Face and Gesture Recognition Workshop*, 2011 IEEE International Conference (pp. 827–834), Santa Barbara, CA.

Haines, N., Southward, M. W., Cheavens, J. S., Beauchaine, T. P., & Ahn, W.-Y. (2019). Using computer-vision and machine learning to automate facial coding of positive and negative affect intensity. *PLOS ONE, 14*, e0211735. doi:10.1371/journal.pone.0211735

Hamm, J., Kohler, C. G., Gur, R. C., & Verma, R. (2011). Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of Neuroscience Methods*, 200, 237–256. doi:10.1016/j.jneumeth.2011.06.023

Hammal, Z., Cohn, J. F., Heike, C., & Speltz, M. L. (2015). Automatic measurement of head and facial movement for analysis and detection of infants' positive and negative affect. *Frontiers in ICT, 2*, 21. doi:10.3389/fict.2015.00021

Hipwell, A., Keenan, K., Kasza, K., Loeber, R., Stouthamer-Loeber, M., & Bean, T. (2008). Reciprocal influences between girls' conduct problems and depression, and parental punishment and warmth: A six-year prospective analysis. *Journal of Abnormal Child Psychology*, 36, 663–677. doi:10.1007/s10802-007-9206-4

Hoffner, C., & Badzinski, D. M. (1989). Children's integration of facial and situational cues to emotion. *Child Development*, 60, 411–422. doi:10.2307/1130986

iMotions. (2018). *Facial expression analysis: The definitive guide.* Retrieved from https://imotions.com/facialexpression-guide-ebook/ on 3/17/2018

Isaacowitz, D. M., Löckenhoff, C. E., Lane, R. D., Wright, R., Sechrest, L., Riedel, R., & Costa, P. T. (2007). Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychology and Aging, 22*, 147–159. doi:10.1037/0882-7974.22.1.147

Izard, C. E., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science, 12*, 18–23. doi:10.1111/1467-9280.00304

Izard, C. E., Huebner, R. R., Risser, D., McGinnes, G., & Dougherty, L. (1980). The young infant's ability to produce discrete emotion expressions. *Developmental Psychology*, 16, 132–140. doi:10.1037/0012-1649.16.2.132

Izard, C. E., & Malatesta, C. Z. (1987). Perspectives on emotional development: I. Differential emotions theory of early emotional development. In J. D. Osofsky (Ed.), *Wiley series on personality processes: Handbook of infant development* (pp. 494–554). Oxford: Wiley.

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109, 7241–7244. doi:10.1073/pnas.1200155109

Kanluan, I., Grimm, M., & Kroschel, K. (2008). *Audio-visual emotion recognition using an emotion space concept.* Paper presented at the 16th European Signal Processing Conference, Lausanne, Switzerland.

Keltner, D., Moffitt, T. E., & Stouthamer-Loeber, M. (1995). Facial expressions of emotion and psychopathology in adolescent boys. *Journal of Abnormal Psychology, 104*, 644–652. doi:10.1037/0021-843X.104.4.644

Korb, S., With, S., Niedenthal, P., Kaiser, S., & Grandjean, D. (2014). The perception and mimicry of facial movements predict judgments of smile authenticity. *PLOS ONE, 9*, e99194. doi:10.1371/journal.pone.0099194

Kotsia, I., & Pitas, I. (2007). Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Transactions on Image Processing*, 16, 172–187. doi:10.1109/TIP.2006.884954

Krumhuber, E. G., Skora, L., Küster, D., & Fou, L. (2016). A review of dynamic datasets for facial expression research. *Emotion Review*, 9, 280–292. doi:10.1177/1754073916670022

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, 7, 227–236. doi:10.1037/npe0000028

Li, J. J., & Lansford, J. E. (2018). A smartphone-based ecological momentary assessment of parental behavioral consistency: Associations with parental stress and child ADHD symptoms. *Developmental Psychology*, 54, 1086–1098. doi:10.1037/dev0000516

Littlewort, G., Whitehill, J., Wu, T., & Fasel, I. (2011). *The computer expression recognition toolbox (CERT).* Paper presented at the IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG 2011), Santa Barbara, CA.

Lopes, A. T., de Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order. *Pattern Recognition*, 61, 610–628. doi:10.1016/j.patcog.2016.07.026

Luterek, J. A., Orsillo, S. M., & Marx, B. P. (2005). An experimental examination of emotional experience, expression, and disclosure in women reporting a history of childhood sexual abuse. *Journal of Traumatic Stress*, 18, 237–244. doi:10.1002/jts.20027

Marsh, P., Beauchaine, T. P., & Williams, B. (2008). Dissociation of sad facial expressions and autonomic nervous system responding in boys with disruptive behavior disorders. *Psychophysiology*, 45, 100–110. doi:10.1111/j.1469-8986.2007.00603.x

Mattson, W. I., Cohn, J. F., Mahoor, M. H., Gangi, D. N., & Messinger, D. S. (2013). Darwin's Duchenne: Eye constriction during infant joy and distress. *PLOS ONE*, 8, e80161. doi:10.1371/journal.pone.0080161

McDuff, D., El Kaliouby, R., Kassam, K., & Picard, R. (2010). *Affect valence inference from facial action unit spectrograms.* Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75–78. doi:10.1126/science.198.4312.75

Messinger, D. S., Mahoor, M., Chow, S.-M., & Cohn, J. (2009). Automated measurement of facial expression in infant-mother interaction: A pilot study. *Infancy*, 14, 285–305. doi:10.1080/15250000902839963

Messinger, D. S., Mattson, W. I., Mahoor, M. H., & Cohn, J. F. (2012). The eyes have it: Making positive expressions more positive and negative expressions more negative. *Emotion*, 12, 430–436. doi:10.1037/a0026498

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 1. doi:10.1109/TAFFC.2017.2740923

Muehlenkamp, J. J., Kerr, P. L., Bradley, A. R., & Adams Larsen, M. (2010). Abuse subtypes and nonsuicidal self-injury: Preliminary evidence of complex emotion regulation patterns. *Journal of Nervous and Mental Disease*, 198, 258–263. doi:10.1097/NMD.0b013e3181d612ab

National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria. (2016). *Behavioral assessment methods for RDoC constructs.* Bethesda, MD: National Institute of Mental Health.

Nelson, C. A. (1987). The recognition of facial expressions in the first two years of life: Mechanisms of development. *Child Development*, 58, 889–909. doi:10.2307/1130530

Nicolaou, M. A., Gunes, H., & Pantic, M. (2011). Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Transactions on Affective Computing*, 2, 92–105. doi:10.1109/T-AFFC.2011.9

Noldus. (2018). FaceReader: Tool for automatic analysis of facial expression: Version 6.0 [Software]. Wageningen, Netherlands: Noldus Information Technology.

Odom, R. D., & Lemond, C. M. (1972). Developmental differences in the perception and production of facial expressions. *Child Development*, 43, 359–369. doi:10.2307/1127541

Parrott, W. G. (2000). *Emotions in social psychology: Key readings*. Philadelphia, PA: Psychology Press.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, 9, 148–158. doi:10.1038/nrn2317

Pollak, S. D., Cicchetti, D., Hornung, K., & Reed, A. (2000). Recognizing emotion in faces: Developmental effects of child abuse and neglect. *Developmental Psychology*, 36, 679–688. doi:10.1037/0012-1649.36.5.679

Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology*, 38, 784–791. doi:10.1037/0012-1649.38.5.784

Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology*, 15, 57–75. doi:10.1348/135532509X433151

Preuschoft, S., & van Hooff, J. A. (1995). Homologizing primate facial displays: A critical review of methods. *Folia Primatologica*, 65, 121–137. doi:10.1159/000156878

Prinz, R. J., Foster, S., Kent, R. N., & O'Leary, K. D. (1979). Multivariate assessment of conflict in distressed and nondistressed mother-adolescent dyads. *Journal of Applied Behavior Analysis*, 12, 691–700. doi:10.1901/jaba.1979.12-691

Ramsook, K. A., Cole, P. M., & Fields-Olivieri, M. A. (2019). What emotion dysregulation looks like: Inferences from behavioral observations. In T. P. Beauchaine & S. E. Crowell (Eds.), *The Oxford handbook of emotion dysregulation*. New York: Oxford University Press.

Reed, L. I., DeScioli, P., & Pinker, S. A. (2014). The commitment function of angry facial expressions. *Psychological Science*, 25, 1511–1517. doi:10.1177/0956797614531027

Ringeval, F., Schuller, B., Valstar, M., Jaiswal, S., Marchi, E., Lalanne, D., … Pantic, M. (2015). *AV+EC 2015: The first affect recognition challenge bridging across audio, video, and physiological data*. Paper presented at the 5th International Workshop on Audio/Visual Emotion Challenge, Brisbane, Australia.

Robin, M., Pham-Scottez, A., Curt, F., Bigre, C. D.-L., Speranza, M., Sapinho, D., … Kedia, B. S. (2012). Decreased sensitivity to facial emotions in adolescents with borderline personality disorder. *Psychiatry Research*, 200, 417–421. doi:10.1016/j.psychres.2012.03.032

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178. doi:10.1037/h0077714

Russell, J. A., & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. *Journal of Personality and Social Psychology*, 48, 1290–1298. doi:10.1037//0022-3514.48.5.1290

Schlosberg, H. (1941). A scale for the judgment of facial expressions. *Journal of Experimental Psychology*, 29, 497–510. doi:10.1037/h0061489

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology*, 44, 229–237. doi:10.1037/h0055778

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review*, 61, 81–88. doi:10.1037/h0054570

Sikka, K., Ahmed, A. A., Diaz, D., Goodwin, M. S., Craig, K. D., Bartlett, M. S., & Huang, J. S. (2015). Automated assessment of children's postoperative pain using computer vision. *Pediatrics*, 136, e124–e131. doi:10.1542/peds.2015-0029

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813–838. doi:10.1037/0022-3514.48.4.813

Smith, E. R., & Mackie, D. M. (2015). Dynamics of group-based emotions: Insights from intergroup emotions theory. *Emotion Review*, 7, 349–354. doi:10.1177/1754073915590614

Smith, M., & Walden, T. (1998). Developmental trends in emotion understanding among a diverse sample of African-American preschool children. *Journal of Applied Developmental Psychology*, 19, 177–197. doi:10.1016/S0193-3973(99)80035-5

Smith, S. (2003). *Digital signal processing: A practical guide for engineers and scientists*. Burlington, MA: Elsevier Science.

Snyder, J., Schrepferman, L., & St Peter, C. (1997). Origins of antisocial behavior: Negative reinforcement and affect dysregulation of behavior as socialization mechanisms in family interaction. *Behavior Modification*, 21, 187–215. doi:10.1177/01454455970212004

Song, R., Over, H., & Carpenter, M. (2016). Young children discriminate genuine from fake smiles and expect people displaying genuine smiles to be more prosocial. *Evolution and Human Behavior*, 37, 490–501. doi:10.1016/j.evolhumbehav.2016.05.002

Sorce, J. F., Emde, R. N., Campos, J., & Klinnert, M. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, 21, 195–200. doi:10.1037/0012-1649.21.1.195

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2017). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, 26, 1–15. doi:10.3758/s13428-017-0996-1

Stubbs, J., Crosby, L., Forgatch, M. S., & Capaldi, D. M. (1998). *Family and peer process code: A synthesis of three Oregon Social Learning Center behavior codes* [Training manual]. Eugene, OR: Oregon Social Learning Center.

Thomas, L. A., De Bellis, M. D., Graham, R., & LaBar, K. S. (2007). Development of emotional facial recognition in late childhood and adolescence. *Developmental Science*, 10, 547–558. doi:10.1111/j.1467-7687.2007.00614.x

Thompson, J. (1941). Development of facial expression of emotion in blind and seeing children. *Archives of Psychology* (Columbia University), 264, 47.

Thompson, R. A. (1990). Emotion and self-regulation. In R. A. Thompson (Ed.), *Nebraska Symposium on Motivation: Vol. 36. Socioemotional development* (pp. 383–483). Lincoln, NE: University of Nebraska Press.

Tian, Y. I., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 97–115. doi:10.1109/34.908962

Troster, H., & Brambring, M. (1992). Early social-emotional development in blind infants. *Child: Care, Health and Development*, 18, 207–227. doi:10.1111/j.1365-2214.1992.tb00355.x

Valstar, M., Gratch, J., Schuller, B., Ringeval, F., Lalanne, D., Torres, M. T., … Pantic, M. (2016). *AVEC 2016: Depression, mood, and emotion recognition workshop and challenge*. Paper presented at the Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, Amsterdam, Netherlands.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219–235. doi:10.1037/0033-2909.98.2.219

Westphal, M., Seivert, N. H., & Bonanno, G. A. (2010). Expressive flexibility. *Emotion*, 10, 92–100. doi:10.1037/a0018420

Widen, S. C., & Russell, J. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, 39, 114–128. doi:10.1037/0012-1649.39.1.114

Yarczower, M., Kilbride, J. E., & Hill, L. (1979). Imitation and inhibition of facial expression. *Developmental Psychology*, 15, 453–454. doi:10.1037/0012-1649.15.4.453

Yoshimura, S., Sato, W., Uono, S., & Toichi, M. (2014). Impaired overt facial mimicry in response to dynamic facial expressions in high-functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 45, 1318–1328. doi:10.1007/s10803-014-2291-7

Zhao, X., & Zhang, S. (2016). A review on facial expression recognition: Feature extraction and classification. *IETE Technical Review*, 33, 505–517. doi:10.1080/02564602.2015.1117403

Zisner, A., & Beauchaine, T. P. (2016). Psychophysiological methods and developmental psychopathology. In D. Cicchetti (Ed.), *Developmental psychopathology: Vol. 2. Developmental neuroscience* (3rd ed., pp. 832–884). Hoboken, NJ: Wiley.

Zivin, G. (1977). On becoming subtle: Age and social rank changes in the use of a facial gesture. *Child Development*, 48, 1314–1321. doi:10.2307/1128489