Check for updates

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science in the Public Interest
2019, Vol. 20(1) 1–68
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1529100619832930
www.psychologicalscience.org/PSPI
SAGE

Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements

Lisa Feldman Barrett[1,2,3], Ralph Adolphs[4], Stacy Marsella[1,5,6], Aleix M. Martinez[7], and Seth D. Pollak[8]

[1]Department of Psychology, Northeastern University; [2]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts; [3]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, Massachusetts; [4]Division of Humanities and Social Sciences, California Institute of Technology; [5]College of Computer and Information Science, Northeastern University; [6]Institute of Neuroscience & Psychology, University of Glasgow; [7]Department of Electrical and Computer Engineering and Center for Cognitive and Brain Sciences, The Ohio State University; and [8]Department of Psychology, University of Wisconsin–Madison

Abstract

It is commonly assumed that a person's emotional state can be readily inferred from his or her facial movements, typically called *emotional expressions* or *facial expressions*. This assumption influences legal judgments, policy decisions, national security protocols, and educational practices; guides the diagnosis and treatment of psychiatric illness, as well as the development of commercial applications; and pervades everyday social interactions as well as research in other scientific fields such as artificial intelligence, neuroscience, and computer vision. In this article, we survey examples of this widespread assumption, which we refer to as the *common view*, and we then examine the scientific evidence that tests this view, focusing on the six most popular emotion categories used by consumers of emotion research: anger, disgust, fear, happiness, sadness, and surprise. The available scientific evidence suggests that people do sometimes smile when happy, frown when sad, scowl when angry, and so on, as proposed by the common view, more than what would be expected by chance. Yet how people communicate anger, disgust, fear, happiness, sadness, and surprise varies substantially across cultures, situations, and even across people within a single situation. Furthermore, similar configurations of facial movements variably express instances of more than one emotion category. In fact, a given configuration of facial movements, such as a scowl, often communicates something other than an emotional state. Scientists agree that facial movements convey a range of information and are important for social communication, emotional or otherwise. But our review suggests an urgent need for research that examines how people *actually* move their faces to express emotions and other social information in the variety of contexts that make up everyday life, as well as careful study of the mechanisms by which people perceive instances of emotion in one another. We make specific research recommendations that will yield a more valid picture of how people move their faces to express emotions and how they infer emotional meaning from facial movements in situations of everyday life. This research is crucial to provide consumers of emotion research with the translational information they require.

Keywords

emotion perception, emotional expression, emotion recognition

Faces are a ubiquitous part of everyday life for humans. People greet each other with smiles or nods. They have face-to-face conversations on a daily basis, whether in person or via computers. They capture faces with smartphones and tablets, exchanging photos of themselves and of each other on Instagram, Snapchat, and other social-media platforms. The ability to perceive faces is one of the first capacities to emerge after birth: An infant begins to perceive faces within the first few days

Corresponding Author:
Lisa Feldman Barrett, 125 Nightingale Hall, Northeastern University, Boston, MA 02115
E-mail: l.barrett@northeastern.edu

of life, equipped with a preference for face-like arrangements that allows the brain to wire itself, with experience, to become expert at perceiving faces (Arcaro, Schade, Vincent, Ponce, & Livingstone, 2017; Cassia, Turati, & Simion, 2004; Gandhi, Singh, Swami, Ganesh, & Sinhaet, 2017; Grossmann, 2015; L. B. Smith, Jayaraman, Clerkin, & Yu, 2018; Turati, 2004; but see Young and Burton, 2018, for a more qualified claim). Faces offer a rich, salient source of information for navigating the social world: They play a role in deciding whom to love, whom to trust, whom to help, and who is found guilty of a crime (Todorov, 2017; Zebrowitz, 1997, 2017; Zhang, Chen, & Yang, 2018). Beginning with the ancient Greeks (Aristotle, in the 4th century BCE) and Romans (Cicero), various cultures have viewed the human face as a window on the mind. But to what extent can a raised eyebrow, a curled lip, or a narrowed eye reveal what someone is thinking or feeling, allowing a perceiver's brain to guess what that someone will do next?[1] The answers to these questions have major consequences for human outcomes as they unfold in the living room, the classroom, the courtroom, and even on the battlefield. They also powerfully shape the direction of research in a broad array of scientific fields, from basic neuroscience to psychiatry.

Understanding what **facial movements** might reveal about a person's emotions is made more urgent by the fact that many people believe they already know. Specific **configurations of facial-muscle movements**[2] appear as if they summarily broadcast or **display** a person's emotions, which is why they are routinely referred to as **emotional expressions** and **facial expressions**. A simple Google search for the phrase "emotional facial expressions" (see Box 1 in the Supplemental Material available online) reveals the ubiquity with which, at least in certain parts of the world, people believe that certain emotion categories are reliably signaled or revealed by certain facial-muscle movement configurations—a set of beliefs we refer to as the **common view** (also called the *classical view*; L. F. Barrett, 2017b). Likewise, many cultural products testify to the common view. Here are several examples:

- Technology companies are investing tremendous resources to figure out how to objectively "read" emotions in people by detecting their presumed facial expressions, such as scowling faces, frowning faces, and smiling faces, in an automated fashion. Several companies claim to have already done it (e.g., Affectiva.com, 2018; Microsoft Azure, 2018). For example, Microsoft's Emotion API promises to take video images of a person's face to detect what that individual is feeling. Microsoft's website states that its software "integrates emotion recognition, returning the confidence across a set of emotions . . . such as anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. These emotions are understood to be cross-culturally and universally communicated with particular facial expressions" (screen 3).
- Countless electronic messages are annotated with emojis or emoticons that are schematized versions of the proposed facial expressions for various emotion categories (Emojipedia.org, 2019).
- Putative emotional expressions are taught to preschool children by displaying scowling faces, frowning faces, smiling faces, and so on, in posters (e.g., use "feeling chart for children" in a Google image search), games (e.g., Miniland emotion games; Miniland Group, 2019), books (e.g., Cain, 2000; T. Parr, 2005), and episodes of Sesame Street (among many examples, see Morenoff, 2014; Pliskin, 2015; Valentine & Lehmann, 2015).[3]
- Television shows (e.g., *Lie to Me*; Baum & Grazer, 2009), movies (e.g., *Inside Out*; Docter, Del Carmen, LeFauve, Cooley, and Lassetter, 2015), and documentaries (e.g., *The Human Face*, produced by the British Broadcasting Company; Cleese, Erskine, & Stewart, 2001) customarily depict certain **facial configurations** as **universal** expressions of emotions.
- Magazine and newspaper articles routinely feature stories in kind: Facial configurations depicting a scowl are referred to as "expressions of anger," facial configurations depicting a smile are referred to as "expressions of happiness," facial configurations depicting a frown are referred to as "expressions of sadness," and so on.
- Agents of the U.S. Federal Bureau of Investigation (FBI) and the Transportation Security Administration (TSA) were trained to detect emotions and other intentions using these facial configurations, with the goal of identifying and thwarting terrorists (R. Heilig, special agent with the FBI, personal communication, December 15, 2014; L. F. Barrett, 2017c).[4]
- The facial configurations that supposedly diagnose emotional states also figure prominently in the diagnosis and treatment of psychiatric disorders. One of the most widely used tasks in autism research, the Reading the Mind in the Eyes Test, asks test takers to match photos of the upper (eye) region of a posed facial configuration with specific mental state words, including emotion words (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). Treatment plans for people living with autism and other brain disorders often include learning to **recognize** these facial configurations

as emotional expressions (Baron-Cohen, Golan, Wheelwright, & Hill, 2004; Kouo & Egel, 2016). This training does not **generalize** well to real-world skills, however (Berggren et al., 2018; Kouo & Egel, 2016).

- "Reading" the emotions of a defendant—in the words of Supreme Court Justice Anthony Kennedy, to "know the heart and mind of the offender" (*Riggins v. Nevada*, 1992, p. 142)—is one pillar of a fair trial in the U.S. legal system and in many legal systems in the Western world. Legal actors such as jurors and judges routinely rely on facial movements to determine the guilt and remorse of a defendant (e.g., Bandes, 2014; Zebrowitz, 1997). For example, defendants who are perceived as untrustworthy receive harsher sentences than they otherwise would (J. P. Wilson & Rule, 2015, 2016), and such perceptions are more likely when a person appears to be angry (i.e., the person's facial structure looks similar to the hypothesized facial expression of anger, which is a scowl; Todorov, 2017). An incorrect inference about defendants' emotional state can cost them their children, their freedom, or even their lives (for recent examples, see L. F. Barrett, 2017b, beginning on page 183).

But can a person's emotional state be reasonably inferred from that person's facial movements? In this article, we offer a systematic review of the evidence, testing the common view that instances of an emotion **category** are signaled with a distinctive configuration of facial movements that has enough **reliability** and **specificity** to serve as a diagnostic marker of those instances. We focus our review on evidence pertaining to six emotion categories that have received the lion's share of attention in scientific research—anger, disgust, fear, happiness, sadness, and surprise—and that, correspondingly, are the focus of the common view (as evidenced by our Google search, summarized in Box 1 in the Supplemental Material). Our conclusions apply, however, to all emotion categories that have thus far been scientifically studied. We open the article with a brief discussion of its scope, approach, and intended audience. We then summarize evidence on how people *actually* move their faces during episodes of emotion, referred to as studies of **expression production**, following which we examine evidence on which emotions are actually inferred from looking at facial movements, referred to as studies of **emotion perception.** We identify three key shortcomings in the scientific research that have contributed to a general misunderstanding about how emotions are expressed and perceived in facial movements and that limit the translation of this scientific evidence for other uses:

1. *Limited reliability* (i.e., instances of the same emotion category are neither reliably expressed through nor perceived from a common set of facial movements).
2. *Lack of specificity* (i.e., there is no unique mapping between a configuration of facial movements and instances of an emotion category).
3. *Limited generalizability* (i.e., the effects of context and culture have not been sufficiently documented and accounted for).

We then discuss our conclusions, followed by proposals for consumers on how they might use the existing scientific literature. We also provide recommendations for future research on emotion production and perception with consumers of that research in mind. We have included additional detail on some topics of import or interest in the Supplemental Material.

Scope, Approach, and Intended Audience of Article

The common view: reading an emotional state from a set of facial movements

In common English parlance, people refer to "an emotion" as if anger, happiness, or any emotion word referred to an event that is highly similar on most occurrences. But an emotion word refers to a category of instances that vary from one another in their **physical features** (e.g., facial movements and bodily changes) and **mental features** (e.g., pleasantness, arousal, experience of the surrounding situation as novel or threatening, awareness of these properties, and so on). Few scientists who study emotion, if any, take the view that every instance of an emotion category, such as anger, is identical to every other instance, sharing a set of necessary and sufficient features across situations, people, and cultures. For example, Keltner and Cordaro (2017) recently wrote that "there is no one-to-one correspondence between a specific set of facial muscle actions or vocal cues and any and every experience of emotion" (p. 62). Yet there is considerable scientific debate about the *extent* of the within-category variation, the specific *features* that vary, the *causes* of the within-category variation, and *implications* of this variation for the nature of emotion (see Fig. 1).

One popular scientific framework, referred to as the *basic-emotion approach*, hypothesizes that instances of an emotion category are expressed with facial movements that vary, to some degree, around a typical set of movements (referred to as a **prototype**; for examples, see Table 1). For example, it is hypothesized that in one situation or for one person, anger might be expressed

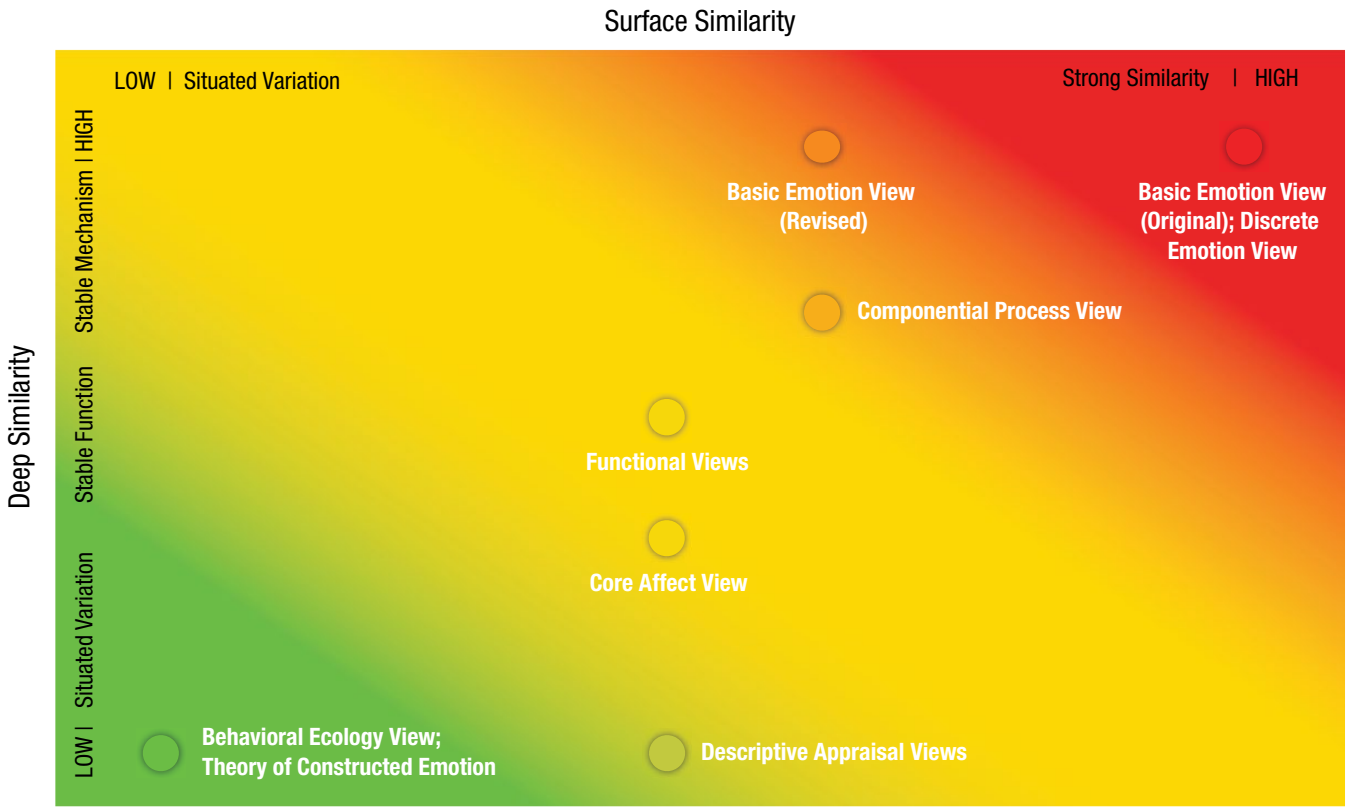


Fig. 1. Explanatory frameworks guiding the science of emotion: the nature of emotion categories and their concepts. The information in the figure is plotted along two dimensions. The horizontal dimension represents hypotheses about the similarities in surface features shared by instances of the same emotion category (e.g., the facial movements that express instances of the same emotion category). The vertical dimension represents hypotheses about the similarities in the mechanisms that cause instances of the same emotion category (e.g., the neural circuits or assemblies that cause instances of the same emotion category). The colors represent the type of emotion categories proposed in each theoretical framework. Approaches in the green area describe ad hoc, abstract categories; those in the yellow area describe prototype or theory-based categories, and those in the red area describe natural-kind categories.

with the facial prototype (e.g., brows furrowed, eyes wide, lips tightened) plus additional facial movements, such as a widened mouth, whereas on other occasions, one facial movement from the prototype might be missing (e.g., anger might be expressed with narrowed eyes or without movement in the eyebrow region; for a discussion, see Box 2 in the Supplemental Material). Nonetheless, the basic-emotion approach still assumes that there is a core facial configuration—the prototype—that can be used to diagnose a person's emotional state in much the same way that a fingerprint can be used to uniquely recognize a person. More substantial variation in expressions (e.g., smiling in anger, gasping with widened eyes in anger, and scowling not in anger but in confusion or concentration) is typically explained as the result of processes that are independent of an emotion itself and that modify its prototypic expression, such as display rules, emotion-regulation strategies (e.g., suppressing the expression), or culture-specific dialects (as proposed by various scientists, including Ekman & Cordaro, 2011; Elfenbein, 2013, 2017; Matsumoto, 1990; Matsumoto, Keltner, Shiota, O'Sullivan, & Frank, 2008; Tracy & Randles, 2011).

By contrast, other scientific frameworks propose that expressions of the same emotion category, such as anger, vary substantially across different people and situations. For example, when the goal of being angry is to overcome an obstacle, it may be more useful to scowl during some instances of anger, smile or laugh, or even stoically widen one's eyes, depending on the temporospatial context. This variation is thought to be a meaningful part of an emotional expression because facial movements are functionally tied to the immediate context, which includes a person's internal context (e.g., the person's metabolic condition, the past experiences that come to mind) and outward context (e.g., whether a person is at work, at school, or at home, who else is present and the broader cultural conditions), both of which vary in dynamic ways over time (see Box 2 in the Supplemental Material).

These debates—regarding the source and magnitude of variation in the facial movements that express

Table 1. A Comparison of the Facial Configurations Listed as the Expressions of Selected Emotion Categories

	Proposed expressive configurations described using The Facial Action Coding System (FACS)					
	Matsumoto, Keltner, Shiota, O'Sullivan, & Frank (2008)		Cordaro et al. (2018)			
Emotion category	Darwin's (1872/1965) description	Observed in research	Reference configuration used	International core pattern	Keltner, Sauter, Tracy, & Cowen (2019)	Physical description
Amusement	Not listed	Not listed	6, 12, 26 or 27, 55 or 56, a "head bounce" (Shiota, Campos, & Keltner, 2003)	6, 7, 12, 16, 25, 26 or 27, 53	6 + 7 + 12 + 25 + 26 + 53	Head back, Duchenne smile (AU 6, 7, 12), lips separated, jaw dropped (AU 26, 27)
Anger	4 + 5 + 24 + 38	4 + 5 or 7 + 22 + 23 + 24	4 + 5 + 7 + 23 (Ekman, Levenson, & Friesen, 1983)	4, 7	4 + 5 + 17 + 23 + 24	Brows furrowed, eyes wide, lips tightened and pressed together
Awe	Not listed	Not listed	1, 5, 26 or 27, 57 and visible inhalation (Shiota et al., 2003)	1, 2, 5, 12, 25, 26 or 27, 53	Not listed	
Contempt	9 + 10 + 22 + 41 + 61 or 62	12 (unilateral) + 14 (unilateral)	12 + 14 (Ekman et al., 1983)	4, 14, 25	Not listed	
Disgust	10 + 16 + 22 + 25 or 26	9 or 10, 25 or 26	9 + 15 + 16 (Ekman et al., 1983)	4, 6, 7, 9, 10, 25, 26 or 27	7 + 9 + 19 + 25 + 26	Eyes narrowed, nose wrinkled, lips parted, jaw dropped, tongue show
Embarrassment	Not listed	Not listed	12, 24, 51, 54, 64 (Keltner & Buswell, 1997)	6, 7, 12, 25, 54, participant dampens smile with 23, 24, frown, etc.)	7 + 12 + 15 + 52 + 54 + 64	Eyelids narrowed, controlled smile, head turned and down, (not scored with FACS: hand touches face)
Fear	1 + 2 + 5 + 20	1 + 2 + 4 + 5 + 20, 25 or 26	1 + 2 + 4 + 5 + 7 + 20 + 26 (Ekman et al., 1983)	1, 2, 5, 7, 25, 26 or 27, participant suddenly shifts entire body backward in chair	1 + 2 + 4 + 5 + 7 + 20 + 25	Eyebrows raised and pulled together, upper eyelid raised, lower eyelid tense, lips parted and stretched
Happiness	6 + 12	6 + 12	6 + 12 (Ekman et al., 1983)	6, 7, 12, 16, 25, 26 or 27	6 + 7 + 12 + 25 + 26	Duchenne smile (6, 7, 12)
Pride	Not listed	Not listed	6, 12, 24, 53, a straightening of the back and pulling back of the shoulders to expose the chest (Shiota et al., 2003)	7, 12, 53, participant sits up straight	53 + 64	Head up, eyes down
Sadness	1 + 15	1 + 15, 4, 17	1 + 4 + 5 (Ekman et al., 1983)	4, 43, 54	1 + 4 + 6 + 15 + 17	Brows knitted, eyes slightly tightened, lip corners depressed, lower lip raised
Shame	Not listed	Not listed	54, 64 (Keltner & Buswell, 1997)	4, 17, 54	54 + 64	Head down, eyes down
Surprise	1 + 2 + 5 + 25 or 26	1 + 2 + 5 + 25 or 26	1 + 2 + 5 + 26 (Ekman et al., 1983)	1, 2, 5, 25, 26 or 27	1 + 2 + 5 + 25 + 26	Eyebrows raised, upper eyelid raised, lips parted, jaw dropped

Note: Descriptions attributed to Darwin are taken from Matsumoto et al. (2008), Table 13.1. Physical descriptions are taken from Keltner et al. (2019). International core patterns refer to expressions of 22 emotion categories that are thought to be conserved across cultures, taken from Cordaro et al. (2018), Tables 4 through 6. A plus sign means that action units would appear simultaneously. A comma means that action units are statistically the most probable to appear but do not necessarily happen simultaneously (D. Cordaro, personal communication, November 11, 2018).

instances of the same emotion category, as well as the magnitude and meaning of the similarity in the facial movements that express instances of different emotion categories—are useful to scientists. But these debates do not provide clear guidance for consumers of emotion research, who are focused on the practical issue of whether emotion categories are expressed with facial configurations of sufficient regularity and distinctiveness so that it is possible to read emotion in a person's face.

The common view of emotional expressions persists, too, because scientists' actions often do not follow their claims in a transparent, straightforward way. Many scientists continue to design experiments, use stimuli, and publish review articles that, ironically, leave readers with the impression that certain emotion categories have a unique, prototypic facial expression, even as those same scientists acknowledge that instances of every emotion category can be expressed with a variable set of facial movements. Published studies typically test the hypothesis that there are unique emotion-expression links (for examples, see the reference lists in Elfenbein & Ambady, 2002; Keltner, Sauter, Tracy, & Cowen, 2019; Matsumoto et al., 2008; also see most of the studies reviewed in this article, e.g., Cordaro et al., 2018). The exact facial configuration tested for each emotion category varies slightly from study to study (for examples, see Table 1), but a core, prototypic facial configuration for a given emotion category is still assumed within a single study. Review articles (again, perhaps unintentionally) reinforce the impression of unique face-emotion mappings by including tables and figures that display a single, unique facial configuration for each emotion category, referred to as the expression, signal or display for that emotion (Fig. 2 presents two recent examples).[5] This pattern of hypothesis testing and writing—that instances of one emotion category are expressed with a single prototypic facial configuration—reinforces (perhaps unintentionally) the common view that each emotion category is consistently and uniquely expressed with its own distinctive configuration of facial movements. Consumers of this research then assume that a distinctive configuration can be used to diagnose the presence of the corresponding emotion in everyday life (e.g., that a scowl indicates the presence of anger with high reliability and specificity).

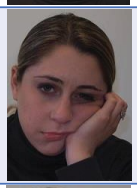
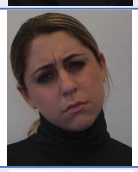
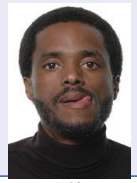
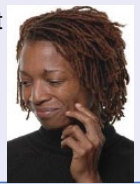
The common view of emotional expressions has also been imported into other scientific disciplines with an interest in understanding emotions, such as neuroscience and artificial intelligence (AI). For example, from a published article on AI:

> American psychologist Ekman noticed that some facial expressions corresponding to certain emotions are common for all the people independently of their gender, race, education, ethnicity, etc. He proposed the discrete emotional model using six universal emotions: *happiness, surprise, anger, disgust, sadness* and *fear*. (Brodny et al., 2016, p. 1; emphasis in original)

Similar examples come from our own articles. One series focused on the brain structures involved in perceiving emotions from facial configurations (Adolphs, 2002; Adolphs, Tranel, Damasio, & Damasio, 1994), and the other focused on early life experiences (Pollak, Cicchetti, Hornung, & Reed, 2000; Pollak & Kistler, 2002). These articles were framed in terms of "recognizing facial expressions of emotion" and exclusively presented participants with specific, posed photographs of scowling faces (the presumed facial expression for anger), wide-eyed, gasping faces (the presumed facial expression for fear), and other presumed prototypical expressions. Participants were shown faces of different individuals, and each person posed the same facial configuration for a given emotion category, ignoring the importance of individual and contextual variation. One reason for this flawed approach to investigating the perception of emotion from faces was that then—at the time these studies were conducted—as now, published experiments, review articles, and stimulus sets were dominated by the common view that certain emotion categories were signaled with an invariant set of facial configurations, referred to as "the facial expressions of basic emotions."

In our review of the scientific evidence, we test two hypotheses that arise from the common view of emotional expressions: that certain emotion categories are each routinely expressed by a unique facial configuration and, correspondingly, that people can reliably infer someone else's emotional state from a set of facial movements. Our discussion is written for consumers of emotion research, whether they be scientists in other fields or nonscientists, who need not have deep knowledge of the various theories, debates, and broad range of findings in the science of emotion, with sufficient pointers to those discussions if they are of interest (see Box 2 in the Supplemental Material).

In discussing what this article is about—the common view that a person's emotional state is revealed in facial movements—it bears mentioning what this article is *not* about: It is not a referendum on the "basic emotion" view that we mentioned briefly, earlier in this section, proposed by the psychologist Paul Ekman and his colleagues; nor is it a commentary on any other specific research program or individual psychologist's view. Ekman's theoretical approach has been highly influential in research on emotion for much of the past 50 years. We often cite studies inspired by the basic-emotion approach, and Ekman's work, for this reason.

State	Example Photo	Action Units	Physical Description
Amusement		6 + 7 + 12 + 25 + 26 + 53	Head back, Duchenne smile, lips separated, jaw dropped
Anger		4 + 5 + 17 + 23 + 24	Brows furrowed, eyes wide, lips tightened and pressed together
Boredom		43 + 55	Eyelids drooping, head tilted (not scorable with FACS: slouched posture, head resting on hand)
Confusion		4 + 7 + 56	Brows furrowed, eyelids narrowed, head tilted
Contentment		12 + 43	Smile, eyelids drooping
Coyness		6 + 7 + 12 + 25 + 26 + 52 + 54 + 61	Duchenne smile, lips separated, head turned and down, eyes turned opposite to head turn
Desire		19 + 25 + 26 + 43	Tongue shown, lips parted, jaw dropped, eyelids drooping
Disgust		7 + 9 + 19 + 25 + 26	Eyes narrowed, nose wrinkled, lips parted, jaw dropped, tongue shown
Embarrassment		7 + 12 + 15 + 52 + 54 + 64	Eyelids narrowed, controlled smile, head turned and down (not scorable with FACS: hand touches face)

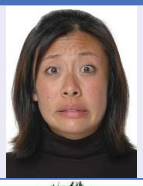
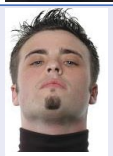
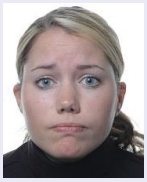
State	Example Photo	Action Units	Physical Description
Fear		1 + 2 + 4 + 5 + 7 + 20 + 25	Eyebrows raised and pulled together, upper eyelid raised, lower eyelid tense, lips parted and stretched
Happiness		6 + 7 + 12 + 25 + 26	Duchenne smile
Interest		1 + 2 + 12	Eyebrows raised, slight smile
Pain		4 + 6 + 7 + 9 + 17 + 18 + 23 + 24	Eyes tightly closed, nose wrinkled, brows furrowed, lips tight, pressed together, and slightly puckered
Pride		53 + 64	Head up, eyes down
Sadness		1 + 4 + 6 + 15 + 17	Brows knitted, eyes slightly tightened, lip corners depressed, lower lip raised
Shame		54 + 64	Head down, eyes down
Surprise		1 + 2 + 5 + 25 + 26	Eyebrows raised, upper eyelid raised, lips parted, jaw dropped
Sympathy		1 + 17 + 24 + 57	Inner eyebrow raised, lower lip raised, lips pressed together, head slightly forward

Fig. 2. *(continued on next page)*

EMOTION EXPRESSION	HYPOTHESIZED PHYSIOLOGICAL FUNCTION	HYPOTHESIZED COMMUNICATIVE FUNCTION	RELEVANT RESEARCH
Happiness	Research Needed	Communicates a Lack of Threat	Preuschoft & Van Hoof, 1997 Ramachandran, 1998
Sadness	Research Needed	Tears Handicap Vision to Signal Appeasement and Elicit Sympathy	Hasson (2009)
Anger	Research Needed	Alerts of Impending Threat, Communicates Dominance	Marsh, Ambady, & Kleck (2005) Wilkowski & Meier (2010)
Fear	Widened Eyes Increase Visual Field and Speed Up Eye Movements	Alerts of Possible Threat and Appeases Potential Aggressors	Marsh et al. (2005) Ohman & Mineka (2001) Susskind et al. (2008)
Surprise	Widened Eyes Increase Visual Field to See Unexpected Stimulus	Research Needed	Ekman (1989)
Disgust	Constricted Orifices Reduce Inhalation of Possible Contaminants	Warns About Aversive Foods, as Well as Distatesful Ideas and Behaviors	Rozin et al. (1994) Chapman, Kim, Susskind, & Anderson (2009)
Pride	Boots Testosterone and Increases Lung Capacity to Prepare for Agonistic Encounters	Communicates Heightened Social Status	Carney, Cuddy, & Yap (2010) Shariff & Tracy (2009) Tracy & Matsumoto (2008)
Shame	Recues/Hides Bodily Targets From Potential Attack	Communicates Lessened Social Status, Desire to Appease	Keltner & Harker (1998) Shariff & Tracy (2009) Tracy & Matsumoto (2008)
Embarrassment	Recues/Hides Bodily Targets From Potential Attack	Communicates Lessened Social Status, Desire to Appease	Keltner & Buswell (1997)

Fig. 2. Example figures from recently published articles that reinforce the common belief in prototypic facial expressions of emotion. The graphic in (a) was adapted from Table 2 in Keltner, D., Sauter, D., Tracy, J., and Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Non-Verbal Behavior*. Photos originally from in Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., and McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18*, 75–93, with permission from the American Psychological Association. Face photos copyright Dr. Lenny Kristal, used with permission. The graphic in (b) was adapted from Figure 2 in Shariff and Tracy (2011).

In addition, the common view of emotional expressions is most readily associated with a simplified version of the basic-emotion approach, as exemplified by the quotes above. Critiques of Ekman's basic-emotion view (and related views) are numerous (e.g., L. F. Barrett, 2006, 2011; L. F. Barrett, Lindquist et al., 2007; Russell, 1991, 1994, 1995; Ortony & Turner, 1990), as are rejoinders that defend it (e.g., Ekman, 1992, 1994; Izard, 2007). Our article steps back from these debates. We instead focus on the existing research on emotional expression and emotion perception in general and ask whether the scientific evidence is sufficiently strong and clear enough to justify the way it is increasingly being used by those who consume it.

A systematic approach for evaluating the scientific evidence

When you see someone smile and infer that the person is happy, you are making what is known as a **reverse inference**: You are assuming that the smile reveals something about the person's emotional state that you cannot access directly (see Fig. 3). Reverse inference requires calculating a **conditional probability**: the probability that a person is in a particular emotion episode (e.g., happiness) given the observation of a unique set of facial muscle movements (e.g., a smile). The conditional probability is written as

$$p(\text{emotion category}|\text{a unique facial configuration})$$

for example,

$$p(\text{happiness}|\text{a smiling facial configuration})$$

Reverse inferences about emotion are ubiquitous in everyday life—whenever you experience someone as emotional, your brain has performed a reverse inference, guessing at the cause of a facial movement when you have access only to the movement itself. Every time an app on a phone or computer measures someone's facial muscle movements, identifies a facial configuration such

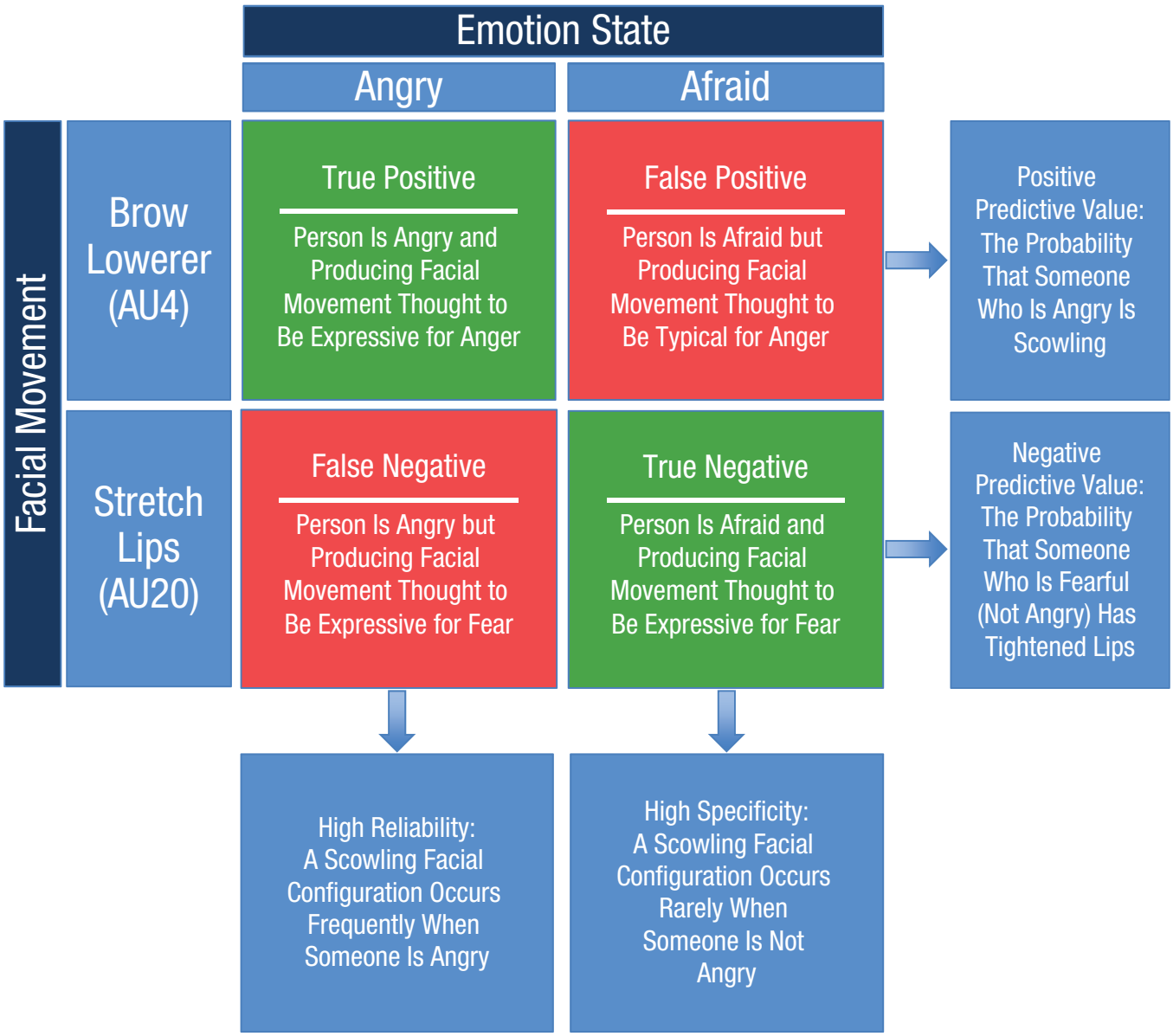


Fig. 3. Defining reliability and specificity. Anger and fear are used as the example categories.

as a frowning facial configuration, and proclaims that the target person is sad, that app has engaged in reverse inference, such as

$$p(\text{sadness}|\text{a frowning facial configuration})$$

Whenever a security agent infers anger from a scowl, the agent has assumed a strong likelihood for

$$p(\text{anger}|\text{a scowling facial configuration})$$

Four criteria must be met to justify a reverse inference that a particular facial configuration expresses and therefore reveals a specific emotional state: reliability, specificity, **generalizability**, and **validity** (explained in Table 2 and Fig. 3). These criteria are commonly encountered in the field of psychological measurement, and over the past several decades, there has been an ongoing dialogue about thresholds for these criteria as they apply in production and perception studies, with some consensus emerging for the first three criteria (see Haidt & Keltner, 1999). Only when a pattern of facial muscle movements strongly satisfies these four criteria can we justify calling it an "emotional expression." If any of these criteria are not met, then we should instead use neutral, descriptive terms to refer to a facial configuration without making unwarranted inferences, simply calling it a smile (rather than an expression of happiness), a frown

Table 2. Criteria Used to Evaluate the Empirical Evidence

Criterion	Expression production	Emotion perception
Cutoffs	Reliability and specificity rates between 70% and 90% provide strong evidence for the common view, rates between 40% and 69% provide moderate support for the common view, and rates between 20% and 39% provide weak support (Ekman, 1994; Haidt & Keltner, 1999; Russell, 1994).	
Reliability	When a person is sad, the proposed expression (a frowning facial configuration) should be observed more frequently than would be expected by chance. This likewise must be true for every other emotion category that is subject to a common belief. Reliability is related to a **forward inference**: Given that someone is happy, what is the likelihood of observing a smile, *p*[set of facial muscle movements | emotion category].	When a person makes a scowling facial configuration, perceivers should consistently infer that the person is angry. This likewise must be true for every facial configuration that has been proposed as *the* expression of a specific emotion category. That is, perceivers must consistently make a reverse inferenc**e**: Given that someone is scowling, what is the likelihood that he or she is angry, *p*[emotion category | set of facial muscle movements].
	Chance means that facial configurations occur randomly with no predictable relationship to a given emotional state. This would mean that the facial configuration in question carries no information about the presence or absence of an emotion category. For example, in an experiment that observes the facial configurations associated with instances of happiness and anger, chance levels of scowling or smiling would be 50%.	Chance means that emotional states occur randomly with no predictable relationship to a given facial configuration. This would mean that the presence or absence of an emotion category cannot be inferred from the presence or absence of the facial configuration. For example, in an experiment that observes how people perceive 51 different facial configurations, chance levels for correctly labeling a scowling face as anger would be 2%.
	Reliability also depends on the base rate: how frequently people make a particular facial configuration. For example, if a person frequently makes a scowling facial configuration during an experiment examining the expressions of anger, sadness and fear, he or she will seem to be consistently scowling in anger when in fact the scowling may be indiscriminate.	Reliability also depends on the base rate: how frequently people use a particular emotion label or make a particular emotional inference. For example, if a person frequently labels facial configurations as "angry" during an experiment examining scowling, smiling, and frowning faces, she will seem to be consistently perceiving anger when in fact she is labeling indiscriminately.
Specificity	If a facial configuration is diagnostic of a specific emotion category, then the facial configuration should express instances of one and only one emotion category better than chance; it should not consistently express instances of any other mental event (emotion or otherwise) at better than chance levels. For example, to be considered *the* expression of anger, a scowling facial configuration must not express sadness, confusion, indigestion, an attempt to socially influence, etc., at better than chance levels. Estimates of specificity, like reliability, depend on base rates and on how chance levels are defined.	If a frowning facial configuration is perceived as *the* diagnostic expression of sadness, then a frowning facial configuration should be labeled only as sadness (or sadness should be inferred only from a frowning facial configuration) at above chance levels. And it should not be consistently perceived as an expression of any mental state other than sadness at better than chance levels. Estimates of specificity, like reliability, depend on base rates and on how chance levels are defined.
Generalizability	Patterns of reliability and specificity should replicate across studies, particularly when different populations are sampled (e.g., infants, congenitally blind individuals, and individuals sampled from diverse cultural contexts, including small-scale, remote cultures). High reliability and specificity across different circumstances ensures that scientific findings are generalizable.	Patterns of reliability and specificity should replicate across studies, particularly when different populations are sampled (e.g., infants, congenitally deaf individuals, and individuals sampled from diverse cultural contexts, including small-scale, remote cultures). High reliability and specificity across different circumstances ensures that scientific findings are generalizable.
Validity	Even if a facial configuration is consistently and uniquely observed in relation to a specific emotion category across many studies (strong generalizability), it is necessary to demonstrate that the person in question is really in the expected emotional state. This is the only way that a given facial configuration leads to **accurate** inferences about a person's emotional state. A facial configuration is valid as a display or a signal for emotion if and only if it is strongly associated with other measures of emotion, preferably those that are objective and do not rely on anyone's subjective report (i.e., a facial configuration should be strongly and consistently related to perceiver-independent evidence about the emotional state of the expresser).	Even if a facial configuration is consistently and uniquely labeled with a specific emotion word across many studies (strong generalizability), it is necessary to demonstrate that the person making the facial configuration is really in the expected emotional state. This is the only way that a given perception or inference of emotion is accurate. A perceiver can be said to be recognizing an emotional expression if and only if the person being perceived is verifiably in the expected emotional state.

Note: Reliability is also related to sensitivity, **consistency**, informational value, and the true positive rate (for further description, see Fig. 3). Specificity is related to uniqueness, discreteness, the true negative rate and referential specificity. In principle, we can also ask more parametrically whether there is a link between the intensity of an emotional instance and the intensity of facial muscle contractions, but scientists rarely do.

(rather than an expression of sadness), a scowl (rather than an expression of anger), and so on.[6]

The null hypothesis and the role of context

Tests of reliability, specificity, generalizability, and validity are almost always compared with what would be expected by sheer chance, if facial configurations (in studies of expression production) and inferences about facial configurations (in studies of emotion perception) occurred randomly with no relation to particular emotional states. In most studies, chance levels constitute the **null hypothesis**. An example of the null hypothesis for reliability is that people do not scowl when angry more frequently than would be expected by chance.[7] If people are observed to scowl more frequently when angry than they would by chance, then the null hypothesis is rejected on the basis of the reliability of the findings. We can also test the null hypothesis for specificity: If people scowl more frequently than they would by chance not only when angry but also when fearful, sad, confused, hungry, and so forth, then the null hypothesis for specificity is retained.[8]

Tests of generalizability are becoming more common in the research literature, again using the null hypothesis. Questions about generalizability test whether a finding in one experiment is reproduced in other experiments in different contexts, using different experimental methods or sampling people from different populations. There are two crucial questions about generalizability when it comes to the production and perception of emotional expressions: Do the findings from a laboratory experiment generalize to observations in the real world? And, do the findings from studies that sample participants from Westernized, educated, industrialized, rich, and democratic (WEIRD; Henrich, Heine, & Norenzayan, 2010) populations generalize to people who live in small-scale remote communities?

Questions of validity are almost never addressed in production and perception studies. Even if **reliable** and **specific** facial movements are observed across generalizable circumstances, whether these facial movements can justify an inference about a person's emotional state is a difficult and unresolved question. (We have more to say about this later.) Consequently, in this article, we evaluate the common view by reviewing evidence pertaining to the reliability, specificity, and generalizability of research findings from production and perception studies.

When observations allow scientists to reject the null hypothesis for reliability, defined as observations that could be expected by chance alone, such evidence provides necessary but not sufficient support for the common view of emotional expressions. A slightly above chance co-occurrence of a facial configuration and instances of an emotion category, such as scowling in anger—for example, a correlation coefficient (r) of about .20 to .39 (adapted from Haidt & Keltner, 1999)—suggests that a person *sometimes* scowls in anger, but not most or even much of the time. Weak evidence for reliability suggests that other factors not measured in the experiment are likely causing people to scowl during an instance of anger. It also suggests that people may express anger with facial configurations other than a scowl, possibly in reliable and predictable ways. Following common usage, we refer to these unmeasured factors collectively as **context**. A similar situation can be described for studies of emotion perception: When participants label a scowling facial configuration as "anger" in a weakly reliable way (between 20% and 39% of the time; Haidt & Keltner, 1999), then this suggests the possibility of unmeasured context effects.

In principle, context effects make it possible to test the common view by comparing it directly with an **alternative hypothesis**—that a person's brain will be influenced by other causal factors—as opposed to comparing the findings with those expected by random chance. It is possible, for example, that a state of anger is expressed differently depending on various factors that can be studied, including the situational context (e.g., whether a person is at work, at school, or at home), social factors (e.g., who else is present in the situation and the relationship between the expresser and the perceiver), a person's internal physical context (e.g., how much sleep they had, how hungry they are), a person's internal mental context (e.g., the past experiences that come to mind or the evaluations they make), the temporal context (what occurred just a moment ago), differences between people (e.g., whether someone is male or female, warm or distant), and the cultural context, such as whether the expression is occurring in a culture that values the rights of individuals (compared with group cohesion) and is open and allows for a variety of behaviors in a situation (compared with closed, having more rigid rules of conduct). Other theoretical approaches offer some of these specific alternative hypotheses (see Box 2 in the Supplemental Material). In practice, however, experiments almost always test the common view against the null hypothesis and rarely test specific alternative hypotheses. When context is acknowledged and studied, it is usually examined as a factor that might moderate a common and universal emotional expression, preserving the core assumptions of the common view (e.g., Cordaro et al., 2018; for more discussion, see Box 3 in the Supplemental Material).

A focus on six emotion categories: anger, disgust, fear, happiness, sadness, and surprise

Our critical examination of the research literature in this article focuses primarily on testing the common view of facial expressions for six emotion categories—anger, disgust, fear, happiness, sadness, and surprise. We do not discuss every emotion category ever studied in the science of emotion. We do not discuss the many emotion categories that exist in non-English-speaking cultures, such as *gigil* (the irresistible urge to pinch or squeeze something cute) or *liget* (exuberant, collective aggression; for discussion of non-English emotion categories, see Mesquita & Frijda, 1992; Pavlenko, 2014; Russell, 1991). We do not discuss the various emotion categories that have been documented throughout history (e.g., T. W. Smith, 2016). Nor do we discuss every English emotion category for which a prototypical facial expression has been suggested. For example, recent studies motivated primarily by the basic-emotion approach have suggested that there are "more than six distinct facial expressions . . . in fact, upwards of 20 multimodal expressions" (Keltner et al., 2019, Introduction, para. 6), meaning that scientists have proposed a distinct, prototypic facial configuration as the facial expression for each of 20 or so emotion categories, including confusion, embarrassment, pride, sympathy, awe, and others.

We focus on six emotion categories for two reasons. First, as we already noted, these categories anchor common beliefs about emotions and their expressions and therefore represent the clearest, strongest test of the common view. They can be traced to Charles Darwin, who stipulated (rather than discovered) that certain facial configurations are expressions of certain emotion categories, inspired by photographs taken by Duchenne (1862/1990) and drawings made by the Scottish anatomist Charles Bell (Darwin, 1872/1965). The proposed expressive facial configurations for each emotion category are presented in Figure 4, and the origin of these facial configurations is discussed in Box 4 in the Supplemental Material.

Second, these six emotion categories have been the primary focus of systematic research for almost a century and therefore provide the largest corpus of scientific evidence that can be evaluated. Unfortunately, the same cannot be said for any of the other emotion categories in question. This is a particularly important point when considering the more than 20 emotion categories that are now the focus of research attention. A PsycInfo search for the term "facial expression" combined with "anger, disgust, fear, happiness, sadness, surprise" produced over 700 entries, but a similar search including "love, shame, contempt, hate, interest, distress, guilt" returned fewer than 70 entries (Duran & Fernández-Dols, 2018). Almost all cross-cultural studies of emotion perception have focused on anger, disgust, fear, happiness, sadness, and surprise (plus or minus a few), and experiments that measure how people spontaneously move their faces to express instances of emotion categories rarely include categories beyond these six. In particular, too few studies measure spontaneous facial movements during episodes of other emotion categories (i.e., production studies) to conclude anything about reliability and specificity, and there are too few studies of how these additional emotion categories are perceived in small-scale, remote cultures to conclude anything about generalizability. In an era where the generalizability and robustness of psychological findings are under close scrutiny, it seemed prudent to focus on the emotion categories for which there are, by a factor of 10, the largest number of published experiments. Nonetheless, our review of the empirical evidence for expressions of emotion categories beyond anger, disgust, fear, happiness, sadness, and surprise did not reveal any new information that weakens the conclusions we discuss in this article. As a consequence, our discussion here, which is based on a sample of six emotion categories, generalizes to those other emotion categories that have been studied.[9]

Producing Facial Expressions of Emotion: A Review of the Scientific Evidence

In this section, we first review the design of a typical experiment in which emotions are induced and facial movements are measured. We highlight several observations to keep in mind as we review the reliability, specificity, and generalizability for expressions of anger, disgust, fear, happiness, sadness, and surprise in a variety of populations, including adults in urban or small-scale remote cultures, infants and children, and **congenitally blind** individuals. Our review is the most comprehensive to date and allows us to comment on whether the scientific findings generalize across different populations of individuals. The value of doing so becomes apparent when we observe how similar conclusions emerge from these research domains.

The anatomy of a typical experiment designed to observe people's facial movements during episodes of emotion

In the typical expression-production experiment, scientists expose participants to objects, images, or events that they (the scientists) believe will evoke an **instance of**

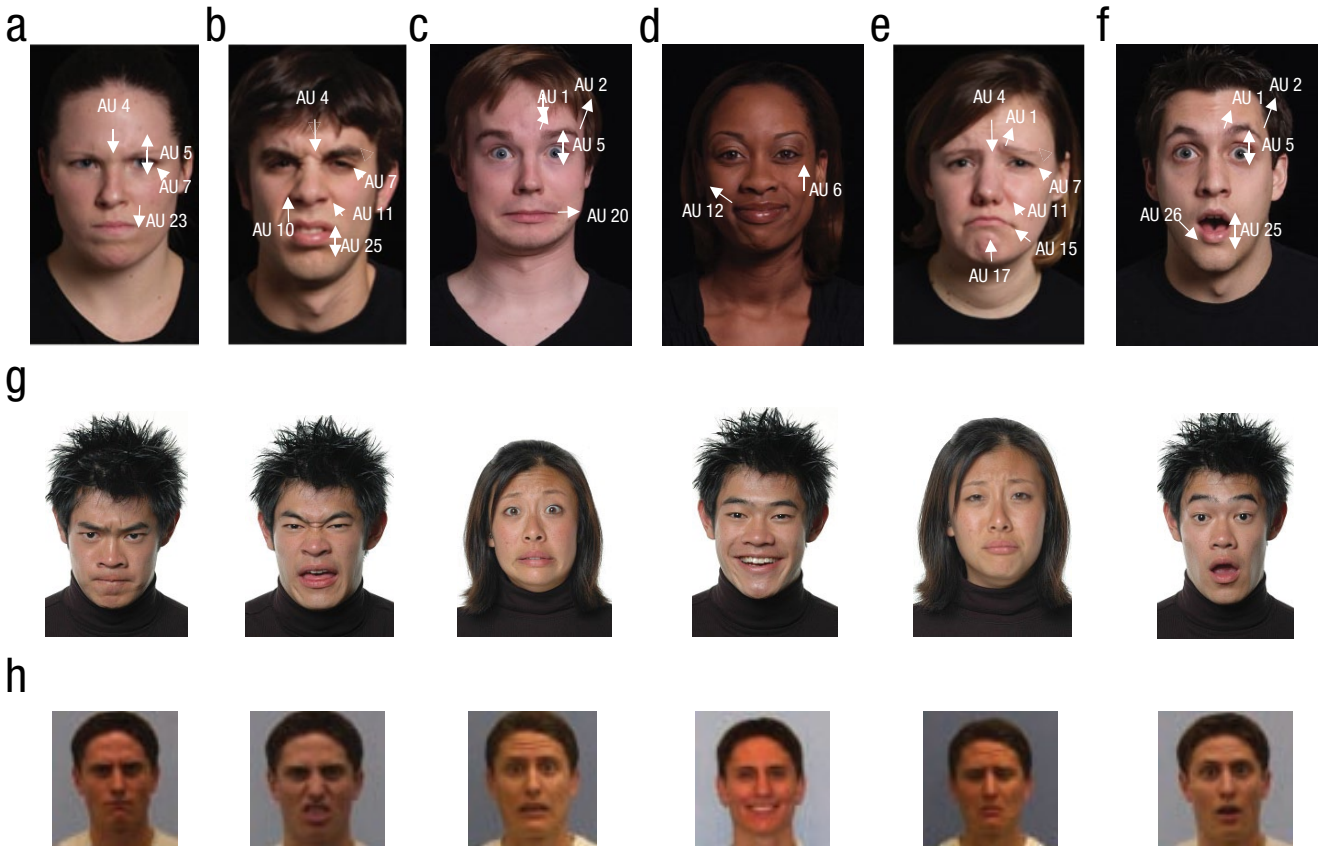


Fig. 4. Facial action ensembles for common-view facial expressions. Facial action coding system (FACS) codes can be used to describe the proposed facial configuration in adults. The proposed expression for anger (a) corresponds to a prescribed emotion FACS (EMFACS) code for anger (described as AUs 4, 5, 7, and 23). The proposed expression for disgust (b) corresponds to a prescribed EMFACS code for disgust (described as AU 10). The proposed expression for fear (c) corresponds to a prescribed EMFACS code for fear (AUs 1, 2, and 5 or 5 and 20). The proposed expression for happiness (d) corresponds to a prescribed EMFACS code for the so-called Duchenne smile (AUs 6 and 12). The proposed expression for sadness (e) corresponds to a prescribed EMFACS code for sadness (AUs 1, 4, 11, and 15 or 1, 4, 15, and 17). The proposed expression for surprise (f) corresponds to a prescribed EMFACS code for surprise (AUs 1, 2, 5, and 26). It was originally proposed that infants express emotions with the same facial configurations as adults. Later research revealed morphological differences between the proposed expressive configurations for adults and infants. Of a possible 19 proposed configurations for negative emotions from the infant coding scheme, only 3 were the same as the configurations proposed for adults (Oster, Hegley, & Nagel, 1992). The proposed expressive prototypes in (g) are adapted from Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., and McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18*, 75–93, with permission from the American Psychological Association. Face photos copyright Dr. Lenny Kristal. The proposed expressive prototypes in (h) are adapted from Figure 2 in Shariff and Tracy (2011).

emotion. It is possible, in principle, to evoke a wide variety of instances for a given emotion category (e.g., Wilson-Mendenhall, Barrett, & Barsalou, 2015); in practice, however, published studies evoke what scientists believe are the most typical instances of each category, usually elicited with a stimulus that is presented without context (e.g., a photograph, a short movie clip separated from the rest of the film or a simplified description of an event, such as "your cousin has just died, and you feel very sad"; Cordaro et al., 2018). Scientists usually include some measure to verify that participants are in the expected emotional state (e.g., asking participants to describe how they feel by rating their experience against a set of emotion adjectives). They then observe participants' facial movements during the **emotional episode** and quantify how well the measure of emotion predicts the observed facial movements. When done properly, this yields estimates of reliability and specificity and, in principle, provides data to assess generalizability. There are limitations to assessing the validity of a facial configuration as an expression of emotion, as we explain below.

Measuring facial movements. Healthy humans have a common set of 34 muscle groups, 17 on each side of the face, that contract and relax in patterns.[10] To create facial movements that are visible to the naked eye, facial

muscles contract, changing the distance between facial features (Neth & Martinez, 2009) and shaping skin into folds and wrinkles on an underlying skeletal structure. Even when facial movements look the same to the naked eye, there may be differences in their execution under the skin. There are individual differences in the mechanics of making a facial movement, including variation in the anatomical details (e.g., muscle configuration and relative size vary, and some people lack certain muscle components), in the neural control of those muscles (Cattaneo & Pavesi, 2014; Hutto & Vattoth, 2015; Müri, 2015), and in the underlying skeletal structure of the face (discussed in Box 5 in the Supplemental Material).

There are three common procedures for measuring facial movements in a scientific experiment. The most sensitive, objective measure of facial movements, called *facial electromyography* (EMG), detects the electrical activity from actual muscular contractions (again, see Box 5 in the Supplemental Material). This is a **perceiver-independent** way of assessing facial movements that detects muscle contractions that are not necessarily visible to the naked eye (Tassinary & Cacioppo, 1992). The utility of facial EMG is unfortunately offset by its impracticality: It requires placing electrodes on a participant's face in a particular configuration. In addition, a person can typically tolerate only a few electrodes on the face at one time. At the writing of this article, relatively few published articles (we identified 123) reported the use of facial EMG, the overwhelming majority of which sparsely sampled the face, measuring the electrical signals for only a small number of muscles (between one and six); none of the studies measured naturalistic facial movements as they occur outside the lab, in everyday life. Consequently, we focus our discussion on two other measurement methods: a **perceiver-dependent** method that describes visible facial movements, called facial actions. Human coders indicate the presence or absence of a facial action while viewing video recordings of participants. Automated methods also exist for detecting facial actions from photographs or videos.

Measuring facial movements with human coders. The **Facial Action Coding System**, or FACS (Ekman, Friesen, & Hager, 2002), is a systematic approach to describe what a face looks like when facial muscle movements have occurred. FACS codes describe the presence and intensity of facial movements. FACS is purely descriptive and is therefore agnostic about whether those movements might express emotions or any other mental event.[11] Human coders train for many weeks to reliably identify specific movements called *action units* (AUs). Each AU is hypothesized to correspond to the contraction of a distinct facial muscle or a distinct grouping of muscles that is visible as a specific facial movement. For example, the raising of the inner corners of the eyebrows (contracting the *frontalis* muscle *pars medialis*) corresponds to AU 1. Lowering of the inner corners of the brows (activation of the corrugator supercilii, depressor glabellae, and depressor supercilii) corresponds to AU 4. AUs are scored and analyzed as independent elements, but the underlying anatomy of many facial muscles constrains them so that they cannot move independently of one another, which generates dependencies between AUs (e.g., see Hao, Wang, Peng, & Ji, 2018). A list of facial AUs and their corresponding facial muscles can be found in Figure 5. Expert FACS coders approach interrater reliabilities of .80 for individual AUs (Jeni, Cohn, & De la Torre, 2013). The first version of FACS (Ekman & Friesen, 1978) was based largely on the work of Swedish anatomist Carl-Herman Hjortsjö, who catalogued the facial configurations described by Duchenne (Hjortsjö, 1969). In addition to the updated versions of FACS (Ekman et al., 2002), other facial coding systems have been devised for human infants (Izard et al., 1995; Oster, 2007), chimpanzees (Vick, Waller, Parr, Smith Pasqualini, & Bard, 2007) and macaque monkeys (L. A. Parr, Waller, Burrows, Gothard, & Vick, 2010; see also L. F. Barrett, 2017a). Figure 4 displays the common FACS codes for the configurations of the facial movements that have been proposed as the prototypic expressions of anger, disgust, fear, happiness, sadness, and surprise, respectively.

Measuring facial movements with automated algorithms. Human coders require time-consuming, intensive training and practice before they can reliably assign AU codes. After training, coding photographs or videos frame by frame is a slow process, which makes human FACS coding impractical to use on facial movements as they occur in everyday life. Large inventories of naturalistic photographs and videos—which have been curated only fairly recently (Benitez-Quiroz, Srinivasan, & Martinez, 2016)—would require decades to manually code. This problem is addressed by automated FACS coding systems using computer-vision algorithms (Martinez, 2017; Martinez & Du, 2012; Valstar, Zafeiriou, & Pantic, 2017).[12] Recently developed computer vision systems have automated the coding of some (but not all) facial AUs (e.g., Benitez-Quiroz, Srinivasan, & Martinez, 2018; Benitez-Quiroz, Wang, & Martinez, 2017; Chu, De la Torre, & Cohn, 2017; Corneanu, Simon, Cohn, & Guerrero, 2016; Essa & Pentland, 1997; Martinez, 2017a; Martinez & Du, 2012; Valstar et al., 2017; see Box 6 in the Supplemental Material), making it more feasible to observe facial movements as they occur in everyday life, at least in principle (see Box 7 in the Supplemental Material).

Automated FACS coding is **accurate** (> 90%) compared with coding from expert human coders, provided that the images were captured under ideal laboratory conditions, where faces are viewed from the front, are

AU	Description	Facial Muscles (Type of Activation)
1	Inner Brow Raiser	*Frontalis (pars medialis)*
2	Outer Brow Raiser	*Frontalis (pars lateralis)*
4	Brow Lowerer	*Corrugator supercilii, depressor supercilii*
5	Upper-Lid Raiser	*Levator palpebrae superioris*
6	Cheek Raiser	*Orbicularis oculi (pars orbitalis)*
7	Lid Tightener	*Orbicularis oculi (pars palpebralis)*
9	Nose Wrinkle	*Levatorlabii superioris alaquaenasi*
10	Upper-Lip Raiser	*Levatorlabii superioris*
11	Nasolabial Deepener	*Zygomaticus minor*
12	Lip-Corner Puller	*Zygomaticus major*
13	Cheeks Puffer	*Levatoranguli oris*
14	Dimpler	*Buccinator*
15	Lip-Corner depressor	*Depressor anguli oris*
16	Lower-Lip depressor	*Depressor labii inferioris*
17	Chin Raiser	*Mentalis*

AU	Description	Facial Muscles (Type of Activation)
18	Lip Puckerer	*Incisiviilabii superioris* and *incisiviilabii inferioris*
20	Lip Stretcher	*Risorius* with *platysma*
22	Lip Funneler	*Orbicularis oris*
23	Lip Tightener	*Orbicularis oris*
24	Lip Pressor	*Orbicularis oris*
25	Lips Part	*Depressor labii inferioris or* relaxatio of *mentalis,* or *orbicularis oris*
26	Jaw Drop	*Masseter, relaxed temporalis and internal pterygoid*
27	Mouth Stretch	*Pterygoids, digastric*
28	Lip Suck	*Orbicularis oris*
41	Lid Droop	
42	Slit	
43	Eyes Closed	
44	Squint	
45	Blink	
46	Wink	

Fig. 5. Facial Action Coding System (FACS; Ekman & Friesen, 1978) codes for adults. AU = action unit.

well illuminated, are not occluded, and are posed in a controlled way (Benitez-Quiroz et al., 2016). (It is important to note, however, that "**accuracy**" here is defined as the FACS coding produced by human judges—which may well have errors.) Under ideal conditions, accuracy is highest (~99%) when algorithms are tested and trained on images from the same database (Benitez-Quiroz et al., 2016). The best of these algorithms works quite well when trained and tested on images from different databases (~90%), as long as the images are

all taken in ideal conditions (Benitez-Quiroz et al., 2016). Accuracy (compared with human FACS coding) decreases substantially when coding facial actions in still images or in video frames taken in everyday life, in which conditions are unconstrained and facial configurations are not stereotypical (e.g., Yitzhak et al., 2017).[13] For example, 38 automated FACS coding algorithms were recently trained on 1 million images (the 2017 EmotioNet Challenge; Benitez-Quiroz, Srinivasan, Feng, Wang, & Martinez, 2017) and evaluated against separate test images that were FACS coded by experts.[14] In these less constrained conditions, accuracy dropped below 83%, and a combined measure of precision and recall (a measure called F_1, ranging from zero to one) was below .65 (Benitez-Quiroz, Srinivasan, et al., 2017).[15] These results indicate that current algorithms are not accurate enough in their detection of facial AUs to fully substitute for expert coders when describing facial movements in everyday life. Nonetheless, these algorithms offer a distinct practical advantage because they can be used in conjunction with human coders to speed up the study of facial configurations in millions of images **in the wild**. It is likely that automated methods will continue to improve as better and more robust algorithms are developed and as more diverse face images become available.

Measuring an emotional state. Once an approach has been chosen for measuring facial movements, a clear test of the common view of emotional expressions depends on having valid measures that reliably and specifically characterize, in a generalizable way, the instances of each emotion category to which the measurements of facial muscle movements can be compared. The methods that scientists use to assess people's emotional states vary in their dependence on **human inference**, however, which raises questions about the validity of the measures.

Relatively objective measures of an emotional instance. The more objective end of the measurement spectrum includes assessing emotions with dynamic changes in the autonomic nervous system (ANS), such as cardiovascular, respiratory, or perspiration changes (measured as variations in skin conductance), and dynamic changes in the central nervous system, such as changes in blood flow or electrical activity in the brain. These measures are thought to be more objective because the measurements themselves (assigning the numbers) do not require a human judgment (i.e., the measurements are **perceiver-independent**). Only the interpretation of the measurements (their psychological meaning) requires human inference. For example, a human observer does not judge whether skin conductance or neural activity increases or decreases; human judgment comes into play when the measurements are interpreted for the emotional meaning.

Currently, there are no objective measures, either singly or as a pattern, that reliably, uniquely, and replicably identify an instance of one emotion category compared with an instance of another. Statistical summaries of hundreds of experiments (i.e., **meta-analyses**) show, for example, that currently there is no reliable relationship between an emotion category, such as anger, and a specific set of physical changes in the ANS that accompany the instances of that category, even probabilistically (the most comprehensive study published to date is Siegel et al., 2018, but for earlier studies, see Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000; Stemmler, 2004; also see Box 8 in the Supplemental Material). In anger, for example, skin conductance can go up, go down, or stay the same (i.e., changes in skin conductance are not consistently associated with anger). And a rise in skin conductance is not unique to instances of anger; it also can occur during a range of other emotional episodes (i.e., changes in skin conductance do not specifically occur in anger and only in anger).[16]

Individual studies often report patterns of ANS measures that distinguish an instance of one emotion category from another, but those patterns are not **replicable** across studies and instead vary across studies, even when studies (a) use the same methods and stimuli and (b) sample from the same population of participants (e.g., compare findings from Kragel & LaBar, 2013, with those from Stephens, Christie, & Friedman, 2010). Similar within-category variation is routinely observed for changes in neural activity measured with brain imaging (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012) and single-neuron recordings (Guillory & Bujarski, 2014). For example, pattern-classification studies discover multivariate patterns of activity across the brain for emotion categories such as anger, sadness, fear, and so on, but these patterns are not replicable from study to study (e.g., compare Kragel & LaBar, 2015; Saarimäki et al., 2016; Wager et al., 2015; for a discussion, see Clark-Polner, Johnson, & Barrett, 2017). This observed variation does not imply that biological variability during emotional episodes is random; rather, it may be context-dependent (e.g., the yellow and green zones of Fig. 1). It may also be the case that current biological measures are simply insufficiently sensitive or comprehensive to capture situated variation in a precise way. If this is so, then such variation should be considered *unexplained* rather than random.

It is worth pointing out the difficult circularity built into these studies that we encounter again a few paragraphs down: Scientists must use some criterion for identifying when instances of an emotion category are present in the first place (so as to draw conclusions about whether emotion categories can be distinguished by different patterns of physical measurements).[17] In most studies that attempt to find bodily or neural

"signatures" of emotions, the criterion is subjective—it is either reported by the participants or provided by the scientist—which introduces problems of its own, as we discuss in the next section.

Subjective measures of an emotional instance. Without objective measures to identify the emotional state of a participant, scientists typically rely on the relatively more subjective measures that anchor the other end of the measurement spectrum. The subjective judgments can come from the participants (who complete self-report measures), from other observers (who infer emotion in the participants), or from the scientists themselves (who use a variety of criteria, including common sense, to infer the presence of an emotional episode). These are all examples of **perceiver-dependent** measurements because the measurements themselves, as well as their interpretation, rely directly on human inference.

Scientists often rely on their own judgments and intuitions (as Charles Darwin did) to stipulate when an emotion is present or absent in participants. For example, snakes and spiders are said to evoke fear. So are situations that involve escaping from a predator. Sometimes scientists stipulate that certain actions indicate the presence of fear, such as freezing or fleeing or even attacking in defense. The validity of the conclusions that scientists draw about emotions depends on the validity of their initial assumptions.[18]

Inferences about emotional episodes can also come from other people—for example, independent samples of study participants, who categorize the situations in which facial movements are observed. Scientists can also ask observers to infer when participants are emotional by having them judge subjects' behavior or tone of voice (e.g., see our later discussion of Camras et al., 2007, in the section on infants and children).

Another common strategy for identifying the emotional state of participants is simply to ask them what they are experiencing. Their self-reports of emotional experience then become the criteria for deciding whether an emotional episode is present or absent. Self-reports are often considered imperfect measures of emotion because they depend on subjective judgments and beliefs and require translation into words. In addition, people can experience an emotional event yet be unaware of it (i.e., conscious with no self-awareness) or unable to express emotion with words (a condition called alexithymia) and therefore unable to report on it. Despite questions about their validity, self-reports are the most common measure of emotion that scientists compare with facial AUs.

Human inference and assessing the presence of an emotional state. At this point, it should be obvious that any measure of an emotional state itself requires some degree of human inference; what varies is the amount of inference that is required. Herein lies a problem: To properly test the hypothesis that certain facial movements reliably and specifically express emotion, scientists (ironically) must first make a reverse inference that an emotional event is occurring—that is, they *infer* the **emotional instance** by observing changes in the body, brain, and behavior. Or they infer (a reverse inference) that an event or object evokes an instance of a specific emotion category (e.g., an electric shock elicits fear but not irritation, curiosity, or uncertainty). These reverse inferences are scientifically sound *only* if measures of emotion reliably, specifically, and validly characterize the instances of the emotion category. So, any clear, scientific test of the common view of emotional expressions rests on a set of more basic inferences about whether an emotional episode is present or absent, and any conclusions that come from such a test are only as sound as those basic inferences. (It is, of course, also possible simply to stipulate the emotion: For instance, a researcher could choose to *define* fear as the set of internal states caused by electric shock, an approach that becomes tautological if not further constrained.)

If all measures of emotion rest on human judgment to some degree, then, in principle, a scientist cannot be sure that an emotional state is present independently of that judgment, which in turn limits the observer-independent validity of any experiment designed to test whether a facial configuration validly expresses a specific emotion category. All face–emotion associations that are observed in an experiment reflect **human consensus**—that is, the degree of **agreement** between self-judgments (from the participants), expert judgments (from the scientist), and/or judgments from other observers (perceivers who are asked to infer emotion in the participants). These types of agreement are often referred to as accuracy, but this may or not be valid. We touch on this point again when we discuss studies that test whether certain facial configurations are routinely perceived as expressions of specific emotion categories.

Testing the common view of emotional expressions: interpreting the scientific observations. If a specific facial configuration *reliably* expresses instances of a certain emotion category in any given experiment, then we would expect measurements of the face (e.g., facial AU codes) to co-occur with other measures that indicate that participants are in the target emotional state. In principle, those measures might be more objective (e.g., ANS changes during an emotional event) or they might be more subjective (e.g., ratings provided by the participants themselves). In practice, however, the vast majority of experiments compare facial movements with subjective measures of emotion—a scientist's judgment about which

emotions are likely to be evoked by a particular stimulus, the judgments of other human observers about participants' emotional states, or participants' self-reports of emotional experience. For example, in an experiment, scientists might ask questions like these: Do the AUs that create a scowling facial configuration co-occur with self-reports of feeling angry? Do the AUs that create a pouting facial configuration co-occur with perceiver's judgments that participants are sad? Do the AUs that create a wide-eyed, gasping facial configuration co-occur when people are exposed to an electric shock? If such observations suggest that a configuration of muscle movements is reliably observed during episodes of a given emotion category, then those movements are said to express the emotion in question. As we will see, many studies show that some facial configurations occur more often than random chance would allow but are not observed with a high degree of reliability (according to the criteria from Haidt and Keltner, 1999, explained in Table 2 of the current article).

If a facial configuration specifically (i.e., uniquely) expresses instances of a certain emotion category in any given experiment, then we would expect to observe little co-occurrence between measurements of the face and measurements indicating the presence of emotional instances from *other* categories (again, see Table 2 and Fig. 3).

If a configuration of facial movements is observed in instances of a certain emotion category in a reliable, specific way within an experiment, so that we can infer that the movements are expressing an instance of the emotion in that study as hypothesized, then scientists can safely infer that the facial movements in question are *an* expression of that emotion category's instances in that situation. One more step is required before we can infer that the facial configuration is *the* expression of that emotion: We must observe a similar pattern of facial configuration–emotion co-occurrences across different experiments, to some extent generalizing across the specific measures and methods used and the participants and contexts sampled. If the facial configuration–emotion co-occurrences replicate across experiments that sample people from the same culture, then the facial configuration in question can reasonably be referred to as *an* emotional expression only in that culture; for example, if a scowling facial configuration co-occurs with measures of anger (and only anger) across most studies conducted on adult participants in the United States who are free from illness, then it is reasonable to refer to a scowl as an expression of anger in healthy adults in the United States. If facial configuration–emotion co-occurrences generalize across cultures—that is, if they are replicated across experiments that sample a variety of instances of that emotion category in people from different cultures—then the facial configuration in question can be said to universally express the emotion category in question.

Studies of healthy adults from the United States and other developed nations

We now review the scientific evidence from studies that document how people spontaneously move their facial muscles during instances of anger, disgust, fear, happiness, sadness, and surprise, as well as how they pose their faces when asked to indicate how they express each emotion category. We examine evidence gathered in the lab and in naturalistic settings, sampling healthy adults who live in a variety of cultural contexts. To evaluate the reliability, specificity, and generalizability of the scientific findings, we adapted criteria set out by Haidt and Keltner (1999), as discussed in Table 2.

Spontaneous facial movements in laboratory studies. A meta-analysis was recently conducted to test the hypothesis that the facial configurations in Figure 4 co-occur, as hypothesized, with the instances of specific emotion categories (Duran, Reisenzein, & Fernández-Dols, 2017). Thirty-seven published articles reported on how people moved their faces when exposed to objects or events that evoke emotion. Most studies included in the meta-analysis were conducted in the laboratory. The findings from these experiments were statistically summarized to assess the reliability of facial movements as expressions of emotion (see Fig. 6). In all emotion categories tested, other than fear, participants moved their facial muscles into the expected configuration more reliably than what we would expect by chance. Reliability levels were weak, however, indicating that the proposed facial configurations in Figure 4 have limited reliability (and to some extent, limited generalizability; i.e., a scowling facial configuration is *an* expression of anger, but not *the* expression of anger). More often than not, people moved their faces in ways that were not consistent with the hypotheses of the common view. An expanded version of this meta-analysis (Duran & Fernández-Dols, 2018) analyzed 131 effect sizes from 76 studies totaling 4,487 participants, with similar results: The hypothesized facial configurations were observed with average effect sizes (r) of .31 for the correlation between the intensity of a facial configuration and a measure of anger, disgust, fear, happiness, sadness, or surprise (corresponding to weak evidence of reliability; individual correlations for specific emotion categories ranged from .06 to .45, interpreted as no evidence of reliability to moderate evidence of reliability). The average proportion of the times that a facial configuration was observed during an emotional event (in one of those categories) was .22

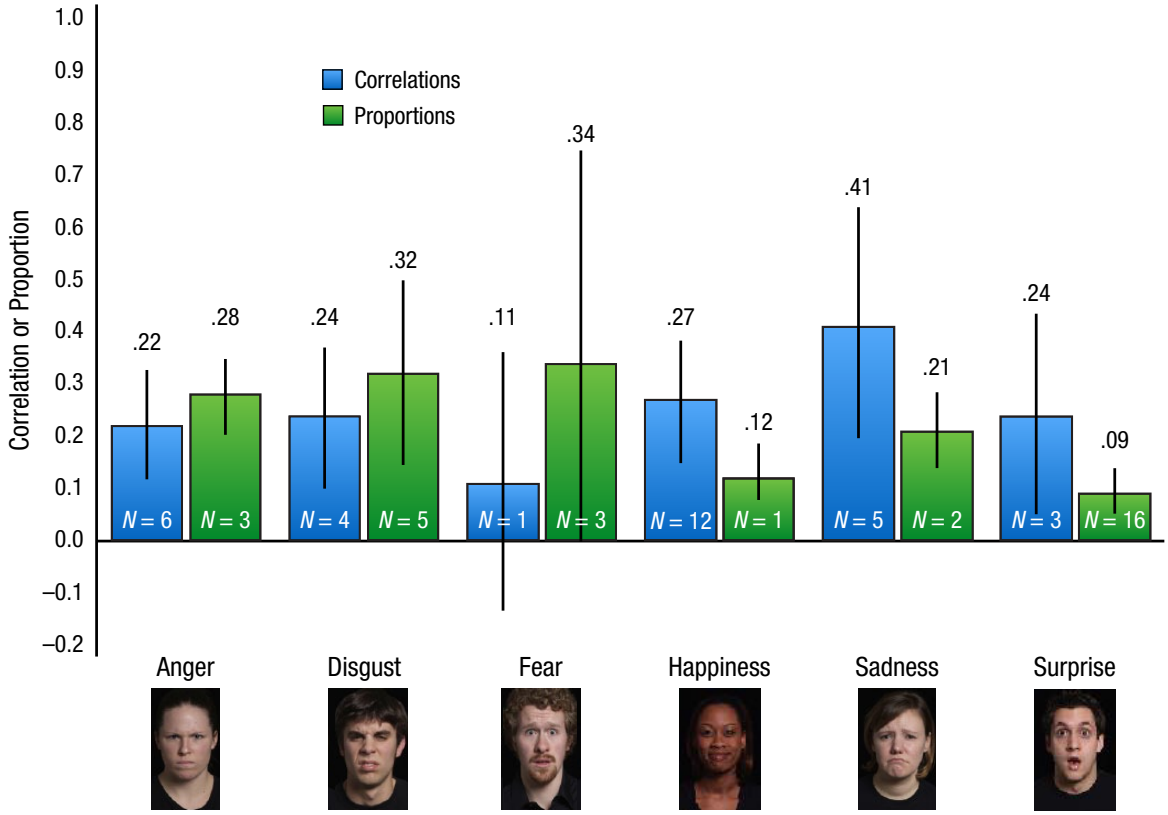


Fig. 6. Meta-analysis of facial movements during emotional episodes: a summary of effect sizes across studies (Duran, Reisenzein, & Fernández-Dols, 2017). Effect sizes are computed as correlations or proportions (as reported in the original experiments). Results include experiments that reported a correspondence between a facial configuration and its hypothesized emotion category as well as those that reported a correspondence between individual AUs of that facial configuration and the relevant emotion category; meta-analytic effect sizes that summarized only the effects for entire ensembles of AUs (the facial configurations specified in Fig. 4) were even lower than those reported here.

(proportions for specific emotion categories ranged from .11 to .35, interpreted as no evidence to weak evidence of reliability).[19]

No overall assessment of specificity was reported in either the original or the expanded meta-analysis because most published studies do not report the **false-positive rate** (i.e., the frequency with which a facial AU is observed when an instance of the hypothesized emotion category was not present; see Fig. 3). Nonetheless, some striking examples of specificity failures have been documented in the scientific literature. For example, a certain smile, called a Duchenne smile, is defined in terms of facial muscle contractions (i.e., in terms of **facial morphology**): It involves movement of the *orbicularis oculi*, which raises the cheeks and causes wrinkles at the outer corners of the eyes, in addition to movement of the *zygomaticus major*, which raises the corners of the lips into a smile. A Duchenne smile is thought to be a spontaneous expression of authentic happiness. Research shows, however, that a Duchenne smile can be intentionally produced when people are not happy (Gunnery & Hall, 2014; Gunnery, Hall, & Ruben, 2013; also see Krumhuber & Manstead, 2009), consistent with evidence that Duchenne smiles often occur when people are signaling submission or affiliation rather than reflecting *happiness* (Rychlowska et al., 2017).

Spontaneous facial movements in naturalistic settings. Studies of facial configuration–emotion category associations in naturalistic settings tend to yield results similar to those from studies that were conducted in more controlled laboratory settings (Fernández-Dols, 2017; Fernández-Dols & Crivelli, 2013). Some studies observe that people express emotions in real-world settings by spontaneously making the facial muscle movements proposed in Figure 4, but such observations are generally not replicable across studies (e.g., cf. Matsumoto & Willingham, 2006 and Crivelli, Carrera, & Fernández-Dols, 2015; cf. Rosenberg & Ekman, 1994 and Fernández-Dols, Sanchez, Carrera, & Ruiz-Belda, 1997). For example, two field studies of winning judo fighters recently demonstrated that so-called Duchenne smiles were better predicted by whether an athlete was interacting with an audience than the degree of happiness reported after winning their

matches (Crivelli et al., 2015). Only 8 of the 55 winning fighters produced a Duchenne smile in Study 1; all occurred during a social interaction. Only 25 of 119 winning fighters produced a Duchenne smile in Study 2, documenting, at best, weak evidence for reliability.

Posed facial movements. Another source of evidence comes from asking participants sampled from various cultures to deliberately pose the facial configurations that they *believe* they use to express emotions. In these studies, participants are given a single emotion word or a single, brief statement to describe each emotion category and are then asked to freely pose the facial configuration that they believe they make when expressing that emotion. Such research directly examines common beliefs about emotional expressions. For example, one study provided college students from Canada and Gabon (in Central Africa) with dictionary definitions for 10 emotion categories. After practicing in front of a mirror, participants posed the facial configurations so that "their friends would be able to understand easily what they feel" (Elfenbein, Beaupre, Levesque, & Hess, 2007, p. 134) and their poses were FACS coded. Likewise, a recent study asked college students in China, India, Japan, Korea, and the United States to pose the facial movements they believe they make when expressing each of 22 emotion categories (Cordaro et al., 2018). Participants heard a brief scenario describing an event that might cause anger ("You have been insulted, and you are very angry about it") and then were instructed to pose a facial (and nonverbal but vocal) expression of emotion, as if the events in the scenario were happening to them. Experimenters were present in the testing room as participants posed their responses. Both studies found moderate to strong evidence that participants across cultures share common beliefs about the expressive pose for anger, fear, and surprise categories; there was weak to moderate evidence for the happiness category, and weak evidence for the disgust and sadness categories (Fig. 7). Cultural variation in participants' beliefs about emotional expressions was also observed.

Neither study compared participants' posed expressions (their beliefs about how they move their facial muscles to express emotions) with observations of how they *actually* moved their faces when expressing emotion. Nonetheless, a quick comparison of the findings from the two studies and the proportions of spontaneous facial movements made during emotional events (from the Duran et al., 2017 meta-analysis) makes it clear that posed and spontaneous movements differ, sometimes quite substantially (again, see Fig. 7). When people pose a facial configuration that they believe expresses an emotion category, they make facial movements that more reliably agree with the hypothesized facial configurations in Figure 4. The same cannot be said of people's spontaneous facial movements during actual emotional episodes, however (for convergent evidence, see Motley & Camden, 1988; Namba, Makihara, Kabir, Miyatani, & Nakao, 2016). One possible interpretation of these findings is that posed and spontaneous facial-muscle configurations correspond to distinct communication systems. Indeed, there is some evidence that volitional and involuntary facial movements are controlled by different neural circuits (Rinn, 1984). Another factor that may contribute to the discrepancy between posed and spontaneous facial movements is that people's beliefs about their own behavior often reflect their **stereotypes** and do not necessarily correspond to how they actually behave in real life (see Robinson & Clore, 2002). Indeed, if people's beliefs, as measured by their facial poses, are influenced directly by the common view, then any observed relationship between posed facial expressions and hypothesized emotion categories is merely evidence of the beliefs themselves.

Summary. Our review of the available evidence thus far is summarized in the first through third data rows in Table 3. The hypothesized facial configurations presented in Figure 4 spontaneously occur with weak reliability during instances of the predicted emotion category, suggesting that they sometimes serve to express the predicted emotion. Furthermore, the specificity of each facial configuration as an expression of an emotion category is *largely unknown* (because it is typically not reported in many studies). In our view, this pattern of findings is most compatible with the interpretation that hypothesized facial configurations are not observed reliably or specifically enough to justify using them to infer a person's emotional state, whether in the lab or in everyday life. We are not suggesting that facial movements are meaningless and devoid of information. Instead, the data suggest that the meaning of any set of facial movements may be much more variable and context-dependent than hypothesized by the common view.

Studies of healthy adults living in small-scale, remote cultures

The emotion categories that are at the heart of the common view—anger, disgust, fear, happiness, sadness, and surprise—derive from modern U.S. English (Wierzbicka, 2014), and their proposed expressions (in Fig. 4) derive from observations of people who live in urbanized, Western settings. Nonetheless, it is hypothesized that these facial configurations evolved as emotion-specific expressions to signal socially relevant emotional information (Shariff & Tracy, 2011) in the challenging

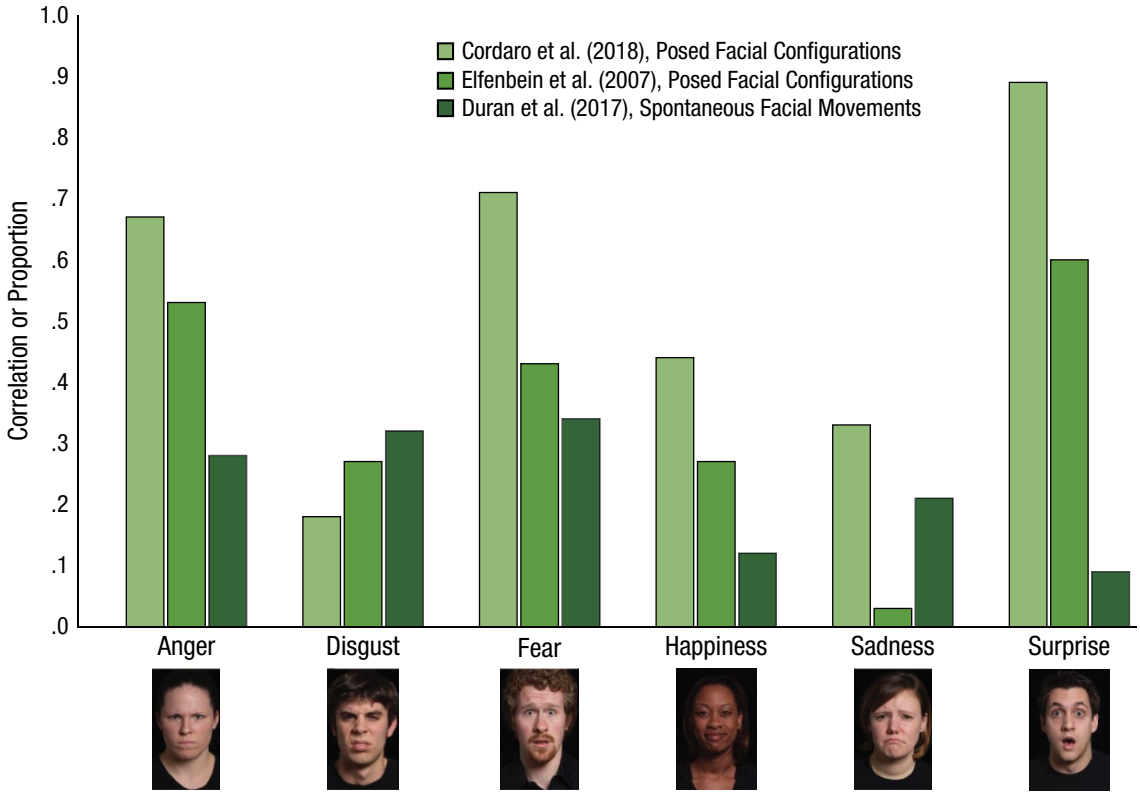


Fig. 7. Comparing posed and spontaneous facial movements. Correlations or proportions are presented for anger, disgust, fear, happiness, sadness, and surprise, separately for three studies. Data are from Table 6 in Cordaro et al. (2018), from Elfenbein, Beaupre, Levesque, and Hess (2007; reliability for the anger category is for AU4 + AU5 only), and from Duran, Reisenzein, and Fernández-Dols (2017; proportion data only).

situations that originated in our hunting-and-gathering hominin ancestors who lived on the African savannah during the Pleistocene era (Pinker, 1997; Tooby & Cosmides, 1990). It is further hypothesized that these facial configurations should therefore be observed during instances of the predicted emotion categories with strong reliability and specificity in people around the world, although the facial movements might be slightly modified by culture (Cordaro et al., 2018; Ekman, 1972). The strongest test of these hypotheses would be to sample participants who live in remote parts of the world with relatively little exposure to Western cultural norms, practices, and values (Henrich et al., 2010; Norenzayan & Heine, 2005) and observe their facial movements during emotional episodes.[20] In our evaluation of the evidence, we continued to use the criteria summarized by Haidt and Keltner (1999; see Table 2 in the current article).

Spontaneous facial movements in naturalistic settings. Our review of scientific studies that systematically measure the spontaneous facial movements in people of small-scale, remote cultures is brief by necessity: There are no such studies. At the time of publication, we were unable to identify even a single published report or manuscript registered on open-access, preprint services that measured facial muscle movements in people of remote cultures as they experienced emotional events. Scientists have almost exclusively observed how people from remote cultures *label* facial configurations as emotional expressions (i.e., studying emotion perception, not production) to test the hypothesis that certain facial configurations evolved to express certain emotion categories in a reliable, specific, and generalizable (i.e., universal) manner. Later in this article, we return to this issue and discuss the findings from these emotion-perception studies.

There are nonetheless several descriptive reports that provide support for the common view of universal emotional expressions (similar to what Valente, Theurel, & Gentaz, 2018, refer to as an "observational approach"). For example, the U.S. psychologist Paul Ekman and colleagues curated an archive of photographs of the Fore hunter-gatherers taken during his visits to Papua New Guinea in the 1960s (Ekman, 1980). The photographs were taken as people went about their daily activities in the small hamlets of the eastern highlands of Papua New Guinea. Ekman used his knowledge of the situation in which each photograph was taken to

Table 3. Reliability and Specificity: A Summary of the Evidence

Study type	Reliability	Specificity
Expression production		
Adults, developed, spontaneous, lab	Weak	Unknown
Adults, developed, spontaneous, naturalistic	Weak	Unknown
Adults, developed, posed	Weak to strong	Unknown
Adults, remote, spontaneous	Unknown	Unknown
Adults, remote, posed	Weak to strong	Unknown
Newborns, infants, toddlers	Unsupported	Unsupported
Congenitally blind	Unsupported to weak	Unsupported
Emotion perception		
Adults, developed, choice-from-array	Moderate to strong	Unknown
Adults, developed, reverse correlation (with choice-from-array)	Moderate	Moderate
Adults, developed, free labeling	Weak to moderate	Weak
Adults, developed, virtual humans	Unknown	Unknown
Adults, remote, choice-from-array (before 2008)	Moderate to strong	Unknown
Adults, remote, choice-from-array (after 2008)	Weak to moderate	Unsupported
Adults, remote, free labeling (before 2008)	Unsupported to strong	Variable
Adults, remote, free labeling (after 2008)	Unsupported	Unsupported
Infants, young children	Unsupported	Unsupported

Note. Criteria were adopted from Haidt and Keltner (1999), who suggest that reliability rates of 70% to 90% are considered strong evidence for universal emotion perception (following Ekman, 1994); presumably, this would also hold for studies of expression production. Weak evidence is in the range of 20% to 40% (Haidt & Keltner, 1999, citing Russell, 1994). By interpolation, reliability between 41% and 69% would be considered moderate evidence. Reliability estimates below 20% are interpreted as findings that clearly do not support the reliability hypothesis. We also adopted these criteria for specificity findings. Developed = studies of participants from the U.S. and other more urban countries; spontaneous = spontaneous facial movements; posed = posed facial configurations; remote = studies of participants from small-scale, remote samples.

assign each facial configuration to an emotion category, leading him to conclude that the Fore expressed emotions with the proposed facial configurations shown in Figure 4. Yet different scientific methods yielded a contrasting conclusion. When Trobriand Islanders living in Papua New Guinea were asked to infer emotions in facial configurations by labeling these same photographs in their native language, both by freely offering words and by choosing the best fitting emotion word from a list of nine choices, they did not label the facial configurations as proposed by Ekman and colleagues, at above-chance levels (Crivelli, Russell, Jarillo, & Fernández-Dols, 2017).[21] In fact, the proposed fear expression—the wide-eyed, gasping face—is reliably interpreted as an expression of threat (intent to harm) and anger by the Maori of New Zealand and by the Trobriand Islanders in Papua New Guinea (Crivelli, Jarillo & Fridlund, 2016).

A compendium of spontaneous human behavior published by the Austrian ethologist Irenäus Eibl-Eibesfeldt (1989) is sometimes cited as evidence for the hypothesis that certain facial movements are universal signals for specific emotion categories. No systematic coding procedure was used in his investigations, however. On close examination, Eibl-Eibesfeldt's detailed descriptions appear to be more consistent with results from the studies of people living in more industrialized cultures that we reviewed above: People move their faces in a variety of ways during episodes belonging to the same emotion category. For example, as reported by Eibl-Eibesfeldt, a rapid eyebrow raise (called an eyebrow flash) is thought to express friendly recognition in some cultures but not all. Likewise, particular facial muscle movements are not specific expressions of a given emotion category. For example, an eyebrow flash would be coded with FACS AU 1 (inner brow raise) and AU 2 (outer brow raise), which are part of the proposed expressions for surprise and fear (Ekman, Levenson, & Friesen, 1983), sympathy (Haidt & Keltner, 1999), and awe (Shiota, Campos, & Keltner, 2003). Even Eibl-Eibesfeldt acknowledged that eyebrow flashes were not unique expressions of specific emotion categories, writing that they also served as a greeting, as an invitation for social contact, as a sign of thanks, as an initiation of flirting, and as a general indication of "yes" in Samoans and other Polynesians, in the Eipo and Trobriand islanders in Papua New Guinea, and in the Yanomami of South America. In Japan, eyebrow flashes are considered an impolite way for adults to greet one another. In the United States and Europe, an eyebrow flash was observed when greeting friends but not when greeting strangers.

Posed facial movements. In the only study of expression production in a natural environment that we could find, researchers read a brief emotion story to people who live in the remote Fore culture of Papua New Guinea and asked each person to "show how his face would appear" (Ekman, 1972, p. 273) if he was the person described in the emotion story (sample size was not reported). Videotapes of 9 participants were shown to 34 U.S. college students who were asked to judge which emotion was being expressed. U.S. participants were asked to infer the emotional meaning of the facial poses by choosing an emotion word from six choices provided by the experimenter (called a **choice-from-array task**; discussed on page 31 of this article). Participants inferred the intended emotional meaning at above-chance levels for smiling (happiness, 73%), frowning (sadness, 68%), scowling (anger, 51%), and nose-wrinkling (disgust, 46%), but not for surprise and fear (27% and 18%, respectively).

Summary. Our review of the available evidence from expression-production studies in small-scale, remote cultures is inconclusive because there are no systematic, controlled observations that examine how people who live in these cultural contexts spontaneously move their facial muscles during emotional episodes. The evidence that does exist suggests that common beliefs about emotion may share some similarities across urban and small-scale cultural contexts, but more research is needed before any interpretations are warranted. These findings are summarized in the fourth and fifth data rows of Table 3.

Studies of healthy infants and children

The facial movements of infants and young children provide a valuable way to test common beliefs about emotional expressions because, unlike older children and adults, babies cannot exert voluntary control over their spontaneous expressive behaviors, meaning that they are unable to deliberately mask or portray instances of emotion in accordance with social demands. As a general rule, infants understand far more about the world than they can easily convey through their physical actions, making it difficult for experiments to distinguish between what infants understand and what they can do; the former often exceeds the latter (Pollak, 2009). Experiments must use human inference to determine when an infant is in an emotional state, as is the case in studies of adults (see *Human Inference and Assessing the Presence of an Emotional State*, above). The presence (or absence) of an instance of emotion is inferred (i.e., stipulated), either by a scientist (who exposes a child to something that is presumed to evoke an emotion episode) or by adult "raters" who infer the emotional meaning of the evoking situation or of the child's body movements and vocalizations (see *Subjective Measures of an Emotional Instance*, above). In the latter cases, inferences are measured by asking research participants to label the situation or the child's emotional state by choosing an emotion word or image from a small set of options, known as a choice-from-array task. We address the strengths and weaknesses of choice-from-array tasks (see Fig. 8) and the potential risk of **confirmatory bias** with the use of such methods (see *A Note on Interpreting the Data*, below).

There is also a risk, given the strong reliance on human inference, that scientists will implicitly confound the measurements made in an experiment with their interpretation of those measurements, in effect overinterpreting infant behavior as emotional, in part because these young research participants cannot speak for themselves. Some early and influential studies confounded the observation of facial movements with their interpreted emotional meaning, leading to the conclusions that babies as young as 7 months old were capable of producing an expression of anger. In fact, it is more scientifically correct to say that the babies were scowling. For example, in one study, infants' facial movements were coded as they were given a cookie, and then the cookie was taken away and placed out of reach, although it was still clearly visible. The babies appeared to scowl when the cookie was removed and not when it was in their mouths (Stenberg, Campos, & Emde, 1983). It is certainly possible that this repeated giving and taking away of the treat angered the infants, but the babies might also have been confused or just generally distressed. Without some independent evidence to indicate that a state of anger was induced, we cannot confidently conclude that certain facial movements in an infant reliably express a specific instance of emotion.

The Stenberg et al. (1983) study illustrates some of the concerning design issues that have historically plagued studies with infants. First, emotion-inducing situations are often defined with common-sense intuitions rather than objective evidence (e.g., an infant is assumed to become angry when a cookie is taken away). In fact, it is difficult to know how any individual infant at any point in time will construct and react to such an event. Second, when an infant produces a facial movement, a common assumption is used to infer its emotional meaning without additional measures or controls (e.g., when a scowling facial configuration is observed, it is assumed to necessarily be an expression of infant anger, even if there are no data to confirm that a scowl is specific to instances of anger in an infant). In fact, years later, as their research program progressed, Campos and his team revised their earlier interpretation of their findings, later concluding that the facial movements in question (infants lowering and drawing together their brows, staring straight ahead, or

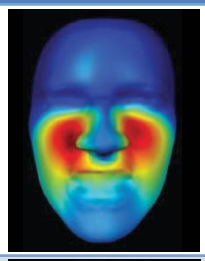
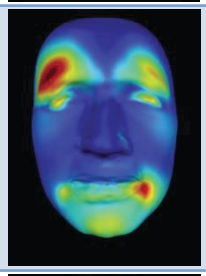
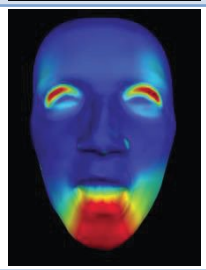
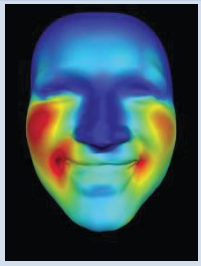
Facial Configuration	Action Unit Description	Associated Emotion Words	
		United Kingdom	China
	6 + 12 + 13 + 14	Delighted, Joy, Happy, Cheerful, Contempt, Pride	Joyful, Delighted, Happy, Glad, Feel Well, Pleasantly Surprised, Embarrassed, Pride
	4 + 20 + 24 + 43	Fear, Scared, Anxious, Upset, Miserable, Sad, Depressed, Shame, Embarrassed	Afraid, Anxious, Distressed, Broken-Hearted, Sorrow and Sadness, Having a Hard Time, Grief, Dismay, Anguish, Worry, Vexed, Unhappy, Shame, Despise
	2 + 5 + 26 + 27	Ecstatic, Excited, Surprised, Frightened, Terrified	Amazed, Greatly Surprised, Alarmed and Panicky, Scared, Fear
	7 + 9 + 16 + 22	Hate, Disgust, Fury, Rage, Anger	Disgusted, Bristle with Anger, Furious, Wild Wrath, Storm of Fury, Storm of Anger, Indignant, Rage

Fig. 8. Culturally common facial configurations extracted using reverse correlation from 62 models of facial configurations. Red coloring indicates stronger action unit (AU) presence and blue indicates weaker AU presence. Some words and phrases that refer to emotion categories in Chinese are not considered emotion categories in English. Adapted with permission of the American Psychological Association, from *Revealing Culturally Common Facial Expressions of Emotion*, by Jack, R. E., Sun, W., Delis, I., Garrod, O. G., and Schyns, P. G., in the *Journal of Experimental Psychology: General*, Vol. 145. Copyright © 2016; permission conveyed through Copyright Clearance Center, Inc.

pressing their lips together) were more generally associated with unpleasantness and distress and were not reliable expressions of anger (e.g., Camras et al., 2007).

The inference problem is particularly poignant when fetuses are studied. For example, in a 4-D ultrasonography study performed with fetuses at 20 gestational weeks, researchers observed the fetuses knitting their brows and described the facial movements as expressions of *distress* (Dondi et al., 2012). Yet the fetuses were producing these facial movements during situations in which fetal distress was unlikely. The brow-knitting was observed during noninvasive ultrasound scanning that did not involve perturbation of the fetus, and the pregnant women were at rest. Furthermore, the scans were brief, and the facial movements were interspersed with other movements that are typically not thought to express negative emotions, such as smiling and mouthing. This is an example of making a scientific inference

about the presence of an emotion solely on the basis of the facial movements without converging evidence that the organism in question (a fetus) was in a distressed state. Doing so highlights the common but unsound assumption that certain facial movements reliably and specifically index instances of the same emotion category.

The study of expression production in infants and children must deal with other design challenges—in addition to the reliance on human inference—that are shared by experiments with adult participants. In particular, most experiments observe facial movements in a restricted range of laboratory settings rather than in the wide variety of situations that naturally occur in everyday life. The frequent use of only a single stimulus or event to observe facial movements for each emotion category limits the opportunity to discover whether the expressions of an emotion category vary systematically with context.

Even with these design considerations, the scientific findings from studies of infants and children parallel those that we encountered from studies on adults: Weak to no reliability and specificity in facial muscle movements is the norm, not the exception (again, using the criteria from Haidt & Keltner, 1999, that are presented in Table 2 of the current article). Although some older studies concluded that infants produce invariant emotional expressions (e.g., Izard et al., 1995; Izard, Hembree, Dougherty, & Spirrizi, 1983; Izard, Hembree, & Huebner, 1987; Lewis, Ramsay, & Sullivan, 2006), these conclusions have been largely superseded by more recent work and in many cases have been reinterpreted and revised by the authors themselves (e.g., Lewis et al., 2006).

Facial movement in fetuses, infants, and young children. The most detailed research on facial movements in fetuses and newborns has focused on smiles. Human fetuses lower their brows (AU4), raise their cheeks (AU6), wrinkle their noses (AU9), crease their nasolabia (AU11), pull the corners of their lips (AU12), show their tongues (AU19), part their lips (AU25), and stretch their mouths (AU27)—all of which have been implicated, to some degree, in adult laughter. Infants sometimes produce facial movements that resemble adult laughter when other considerations suggest that they are in distress and pain (Dondi et al., 2012; Hata et al., 2013; Reissland, Francis, & Mason, 2013; Reissland, Francis, Mason, & Lincoln, 2011; Yan et al., 2006). Within 24 hr of birth, infants raise their cheek muscles in response to being touched (Cecchini, Baroni, Di Vito, & Lai, 2011). But these movements are not specific to smiling; neonates also raise their cheeks (contract the zygomatic muscle) during rapid eye movement (REM) sleep, when drowsy, and during active sleep (Dondi et al., 2007). A neonatal smile with raised cheeks is caused by brainstem activation (Rinn, 1984), and likely reflects internally generated arousal rather than expressing or communicating an emotion or even a more general feeling of pleasure (Emde & Koenig, 1969; Sroufe, 1996; Wolff, 1987). So, it remains unclear whether fetal or neonatal facial muscle movements have any relationship to specific emotional episodes as well as more generally to pleasant feelings or to other social meanings (Messinger, 2002).

In fact, it is not clear that fetal and neonatal facial movements always have a psychological meaning (consistent with a behavioral-ecology view of facial movements; Fridlund, 2017). Newborns appear to produce some combinations of facial movements for muscular reasons. For example, infants produce facial movements associated with the proposed expression for surprise (open mouth and raised eyebrows) in situations that are unsurprising, just because opening the mouth necessarily raises their eyebrows; conversely, infants do not consistently show the proposed expressive configuration for surprise in contexts that are likely to be surprising (Camras, 1992; Camras, Castro, Halberstadt, & Shuster, 2017). The facial movement that is part of the proposed expression for sadness (brows oblique and drawn together) occurs when infants attempt to lift their heads to direct their gaze (Michel, Camras, & Sullivan, 1992).

In addition, newborns produce many facial movements that co-occur with fussiness, distress, focused attention, and distaste (Oster, 2005). Newborns react to being given sweet versus sour liquids; for example, when given a sour liquid, newborns make a nose-wrinkle movement, which is part of the proposed expressive configuration for disgust (Granchrow, Steiner, & Daher, 1983). However, other studies show that newborns also make this facial movement when given sweet, salty, and bitter tastes (e.g., Rosenstein & Oster, 1988). Still other studies show that nose-wrinkling does not always occur when infants taste lemon juice (i.e., when that facial movement is expected; Bennett, Bendersky, & Lewis, 2002). More generally, infants rarely produce consistent facial movements that cleanly map onto any single emotion category. Instead, infants produce a variety of facial configurations that suggest a lack of emotional specificity (Matias & Cohn, 1993).

There are further examples that illustrate how infant facial movements lack strong reliability and specificity. In a study of 11-month-old babies from the United States, China, and Japan, infants saw a toy gorilla head that growled (to induce fear) or their arms were restrained (to induce anger; Camras et al., 2007). Observers judged the infants to be fearful or angry on the basis of their body movements, yet the infants produced the same facial movements in the two situations.[22] In another study, 1-year-old infants were videotaped in situations in which they were tickled (to elicit joy), tasted sour

flavors (to elicit disgust), watched a jack-in-the box (to elicit surprise), had an arm restrained (to elicit anger), and were approached by a masked stranger (to elicit fear; Bennett et al., 2002). Infants whose arms were restrained (to purportedly induce an instance of anger) produced the facial actions associated with the proposed facial configuration for an anger expression only 24% of the time (low reliability); instead, 80 infants (54%) produced the facial actions proposed as the expression of surprise, 37 infants (25%) produced the facial actions proposed as the expression of joy, 29 infants (19%) produced the facial actions proposed as the expression of fear, and 28 (18%) produced the facial actions proposed as the expression of sadness. This dramatic lack of specificity was observed for all emotion categories studied. An equal number of babies produced facial movements that are proposed as the expressions of joy, surprise, anger, disgust, and fear categories when a sour liquid was placed on infants' tongues to elicit disgust. When infants faced a masked stranger, only 20 (13%) produced facial movements that corresponded to the proposed expression for fear, compared with 56 infants (37%) who produced facial actions associated with the proposed expression for instances of joy.[23]

Taken together, these findings suggest that infant facial movements may be associated with **affect** (i.e., the affective features of experience, such as distress or arousal), as originally described by Bridges (1932), or may communicate a desire to approach or avoid something (e.g., Lewis, Sullivan, & Kim, 2015). Affective features such as valence (ranging from pleasantness to distress) and arousal (ranging from activated to quiescent) are continuous properties of experience, just as **approach/avoidance** is an affective property of action. These affective features are shared by many instances of different emotion categories, as well as with mental events that are not considered emotional (as discussed in Box 9 in the Supplemental Material), but this does not diminish their importance or effectiveness for infants.[24] Over time, infants likely learn to differentiate mental events with simple affective features into episodes of emotion with additional psychological features that are specific to their sociocultural contexts, making them maximally effective at eliciting needed responses from their caregivers (L. F. Barrett, 2017b; Holodynski & Friedlmeier, 2006; Weiss & Nurcombe, 1992; Witherington, Campos, & Hertenstein, 2001).

The affective meaning of an infant's facial movements may, in fact, be what makes these movements so salient for adult observers. When infants move their lips, open their mouths, or constrict their eyes, adults view infants as feeling more pleasant or unpleasant depending on the context (Bolzani Dinehart et al., 2005). Infant expressions thus do have a reliable link to instrumental effects in the adults who observe them—playing an important role in parent–infant interaction, attachment, and the beginnings of social communication (Atzil, Gao, Fradkin, & Barrett, 2018; Feldman, 2016). For example, if an infant cries with narrowed eyes, adults infer that the infant is feeling negative, is having an unwanted experience, or is in need of help, but if the infant makes that same eye movement while smiling, adults infer that the infant is experiencing more positive emotion. These data consistently point to the usefulness of facial movements in the communication of arousal and valence, particularly when combined or with other communicative features such as vocalizations (properties of affect; see Box 9 in the Supplemental Material). Even when episodes of more specific emotions start to emerge, we do not yet have evidence that facial movements map reliably and regularly to a specific emotion category.

Young children begin to produce adult-like facial configurations *after* the first year of life. Even then, however, children's facial movements continue to lack strong reliability and specificity (Bennett et al., 2002; Camras & Shutter, 2010; Matias & Cohn, 1993; Oster, 2005). Examples of a wide-eyed, gasping facial configuration, proposed as the expression of fear (see Fig. 4), have rarely been observed or reported in young infants (Witherington, Campos, Harriger, Bryan, & Margett, 2010). Nor do infants reliably produce a scowling facial configuration, proposed as the expression of anger (again, see Fig. 4). Infants scowl when they cry or are about to cry (Camras, Fatani, Fraumeni, & Shuster, 2016). A frown (mouth corner depression, AU15) is not reliably and specifically observed when infants are frustrated (Lewis & Sullivan, 2014; Sullivan and Lewis, 2003). A smile (cheek raising and lip corner pulling, AU6 and AU12) is not reliably observed when infants are in visually engaging or mastery situations, or even when they are in pleasant social interactions (Messinger, 2002).

Experiments that observe young children's facial movements in naturalistic settings find largely the same results as those conducted in controlled laboratory settings. For example, one study trained ethnographic videographers to record a family's daily activities over 4 days (Sears, Repetti, Reynolds, & Sperling, 2014). Coders judged whether or not the child from each participating family made a scowling facial configuration (referred to as an expression of anger), a frowning facial configuration (referred to as an expression of sadness), and so on, for the six (presumed) emotion categories included in the study—happiness, sadness, surprise, disgust, fear, and anger. During instances that were coded as anger (defined as situations that included verbal disagreements or sibling bickering, requests for

compliance and/or reprimands from parents, parent refusal of child requests, during homework, and sibling provocation), a variety of facial movements were observed, including frowns, furrowed brows, and eye-rolls, as well as a variety of vocalizations, including shouts and whining, and both nonaggressive and aggressive physical behaviors.

Perhaps the most telling observation for our purposes is that expressions of anger were more often vocal than facial. During anger situations, children raised their voices 42% of the time, followed by whining about 21% of the time. By contrast, children made scowling facial configurations only 16.2% of the time.[25] Yet even during anger situations, the facial movements were predominantly frowning, which can be part of many different proposed facial configurations. The authors reasoned that children engage in specific behaviors to obtain specific goals, and that behaviors such as whining are more likely to attract attention and possibly change parental behavior than is a facial movement. Indeed, it is easier for parents to ignore a negative facial expression than a whining child in the room! Similar findings for low reliability and specificity of the facial configurations presented in Figure 4 were recently observed in a naturalistic study that videotaped 7- to 9-year-old children and their mothers discussing a conflict during their visit to the laboratory related to homework, chores, bedtime, or interactions with siblings (Castro, Camras, Halberstadt, & Shuster, 2018).

Summary. Newborns and infants react to the world around them with facial movements. There is not yet sufficient evidence, however, to conclude that these facial movements reliably and specifically express the instances of any emotion category (findings are summarized in the sixth data row of Table 3). When considered alongside vocalizations and body movements, there is consistent evidence that infant facial movements reliably signal distress, interest, and arousal and perhaps serve as a call for help and comfort. In young children, instances of the same emotion category appear to be expressed with a variety of different muscle movements, and the same muscle movements occur during instances of various emotion categories, and even during nonemotional instances. It may be the case that reliability and specificity emerges through learning and development (see Box 10 in the Supplemental Material), but this remains an open question that awaits future research.

Studies of congenitally blind individuals

Another source of evidence to test the common view comes from observations of facial movements in people who were born blind. The assumption is that people who are blind cannot learn by watching others which facial muscles to move when expressing emotion. On the basis of this assumption, several studies have claimed to find evidence that congenitally blind individuals express emotions with the hypothesized facial configurations in Figure 4 (e.g., blind athletes were reported to show expressions that are reliably interpreted as shame and pride; Tracy & Matsumoto, 2008; see also Matsumoto & Willingham, 2009). People who are born blind learn through other **sensory modalities**, however (for a review, see Bedny & Saxe, 2012), and therefore can learn whatever regularities exist between emotional states and facial movements from hearing descriptions in conversation, in books and movies, and by direct instruction.[26] As an example of such learning, Olympic athletes who won medals smiled only when they knew they were watched by other people, such as when they were on the podium facing the audience; in other situations, such as while they waited behind the podium or while they were on the podium facing away from people but toward a flag, they did not smile (but presumably were still very happy; Fernández-Dols & Ruiz-Belda, 1995). Such findings are consistent with the behavioral ecology view of facial expressions (Fridlund, 2017, 1991) and with more recent sociological evidence that smiles are social cues that can communicate different social messages depending on the cultural context (J. Martin, Rychlowska, Wood, & Niedenthal, 2017).

The limitations that apply to studies of emotional expressions in sighted individuals, reviewed throughout this article, are even more applicable to scientific studies of emotional expressions in the blind.[27] Participants are given predetermined emotion categories that constrain their possible responses, and facial movements are often quantified by human judges who have their own biases when inferring the emotional meaning of facial movements (e.g., Galati, Miceli, & Sini, 2001; Galati, Scherer, & Ricci-Bitti, 1997; Valente et al., 2018). In addition, people who are blind make additional, often unusual movements of the head and the eyes (Chiesa, Galati, & Schmidt, 2015) to better hear objects or echoes. These unusual movements might influence expressive facial movements. More importantly, they reveal whether a participant is blind or sighted, and this knowledge can bias human raters who are judging the presence or absence of facial movements in emotional situations.

Helpful insights about the facial expressions of congenitally blind individuals comes from a recent review (Valente et al., 2018) that surveyed 21 studies published between 1932 and 2015. These studies observed how blind participants move their faces during instances of emotion and then compared those movements with

both the proposed expressive forms in Figure 4 and the facial movements of sighted people. Both spontaneous facial movements and posed movements were tested. Eight older studies (published between 1932 and 1977) reported that congenitally blind individuals spontaneously expressed emotions with the proposed facial configurations in Figure 4, but Valente et al. (correctly) questioned the objectivity of these studies because the data were based largely on subjective impressions offered by researchers or their assistants.

The 13 studies published between 1980 and 2015 were better designed: Researchers videotaped participants' facial movements and described them using a formal facial coding system for adults (e.g., FACS) or a similar coding system for children. There are too few of these studies and the sample sizes are insufficient to conduct a formal meta-analysis, but taken together they suggest that, in general, congenitally blind individuals *spontaneously* moved their faces in ways similar to sighted individuals during instances of emotion: Both groups expressed instances of anger, disgust, fear, happiness, sadness, or surprise with the proposed expressive configurations (or their individual AUs) in Figure 4, with either weak reliability or no reliability, and neither group produced any of the configurations with any specificity (e.g., Galati et al., 2001; Galati et al., 1997; Galati, Sini, Schmidt, & Tinti, 2003). The lack of specificity is not surprising given that, on closer inspection, several of the studies discussed in Valente et al. (2018) compared emotion categories that systematically differ in their prototypical affective properties, contrasting facial movements in pleasant and unpleasant circumstances (e.g., Cole et al., 1989), or observed facial movements only in pleasant circumstances without distinguishing the facial AUs for the happiness category from other positive emotion categories (e.g., Chiesa et al., 2015). As a consequence, the findings from these studies cannot be interpreted unambiguously as evidence specifically pertaining to emotional expressions, per se.

Congenitally blind and sighted individuals were similar to one another in the variety of their spontaneous facial movements, but they differed in their posed facial configurations. After listening to descriptions of situations that were assumed to elicit an instance of anger, sadness, fear, disgust, surprise, and happiness, sighted participants posed their faces with the proposed expressive forms for the negative emotion categories in Figure 4 at higher levels of reliability and specificity than did blind participants (Galati et al., 1997; Roch-Levecq, 2006). These findings suggest that sighted individuals share common beliefs about emotional expressions, replicating other findings with posed expressions (see Table 3, third data row), whereas congenitally blind individuals may share these beliefs to a lesser degree; their knowledge of social rules for producing those configurations on command differs from those of sighted individuals.

Taken together, the evidence from studies of blind individuals is consistent with the other scientific evidence reviewed so far (see Table 3). Even in the absence of visual experience, blind individuals, like sighted individuals, develop the ability to spontaneously make a variety of facial movements to express emotion, but those movements do not reliably and specifically configure in the manner proposed by the common view of emotion (depicted in Fig. 4). Learning to voluntarily pose the proposed expressions in Figure 4 does seem to covary with vision, however, further emphasizing that posed and spontaneous expressions should be treated as different phenomena. Further scientific attention is warranted to examine how congenitally blind individuals learn, via other sensory modalities, to express emotions.

Summary of scientific evidence on the production of facial expressions

The scientific findings we have reviewed thus far—dealing with how people actually move their faces during emotional events—does not strongly support the common view that people reliably and specifically express instances of emotion categories with spontaneous facial configurations that resemble those proposed in Figure 4. Adults around the world, infants and children, and congenitally blind individuals all show much more variability than commonly hypothesized. Studies of posed expressions further suggest that people believe that particular facial movements express particular emotions more reliably and specifically than is warranted by the scientific evidence. Consequently, it is misleading to refer to facial movements with commonly used phrases such as "emotional facial expression," "emotional expression" or "emotional display." More neutral phrases that assume less, such as "facial configuration" or "pattern of facial movements" or even "facial actions," are more scientifically accurate and should be used instead.

We next turn our attention to the question of whether people reliably and specifically infer certain emotions from certain patterns of facial movements, shifting our focus from studies of production to studies of perception. It has long been assumed that emotion perception provides an indirect way of testing the common view of expression production, because facial expressions, when they are assumed to be displays of emotional states, are thought to have coevolved with the ability to recognize and read them (Ekman, Friesen, & Ellsworth, 1972). For example, Shariff and Tracy (2011) have suggested that emotional expression and emotion perception likely coevolved as an integrated

signaling system (for additional discussion, see Jack, Sun, Delis, Garrod, & Schyns, 2016).[28] In the next section, we review the scientific evidence on emotion perception.

Perceiving Emotions From Facial Movements: A Review of the Scientific Evidence

For over a century, researchers have directly examined whether people reliably and specifically infer emotional meaning in the facial configurations presented in Figure 4. Most of these studies are interpreted as evidence for people's ability to *recognize* or *decode* emotion in facial configurations, on the assumption that the configurations *broadcast* or *signal* emotional information to be recognized or detected. This is yet another example of confusing what is known and what is being tested. A more correct interpretation is that these studies evaluate whether or not people reliably and specifically *infer*, *attribute*, or *judge* emotion in those facial configurations. The pervasive tendency to confuse inference and recognition may explain why very few studies have actually investigated the processes by which people detect the onset and offset of facial movements and infer emotions in those movements (i.e., few studies consider the mechanisms by which people infer emotional states from detecting and perceiving facial movements; for discussion, see Lynn & Barrett, 2014; Martinez, 2017a, 2017b). In this section, we first review the design of typical emotion-perception experiments that are used to test the common view that emotions can be reliably and specifically "read out" from facial movements. We also examine the emotions people infer from the facial movements in dynamic, computer-generated faces, a class of studies that offers an interesting alternative way to study emotion perception, and in virtual humans, which provides the opportunity for a more implicit approach to studying emotion perception.

The anatomy of a typical experiment designed to test the common view

For a person—a perceiver—to infer that another person is in an emotional state by looking at that person's facial movements, the perceiver must have many competencies. People move their faces continuously (i.e., real human faces are never still), so a perceiver must notice or detect the relevant facial movements in question and **discriminate** them from other facial movements (that is, the perceiver must be able to set a perceptual boundary to know when the movements begin and end, and, for example, that a scowl is different from a sneer). To do this, the perceiver must be able to identify (or segment) the movements as an ensemble or pattern (i.e., bind them together and distinguish them from other movements that are normally inferred to be irrelevant). And the perceiver must be able to infer similarities and differences between different instances of facial movements, as specified by the task (e.g., categorize a group of facial movements as instances expressing anger). This **categorization** might involve merely labeling the facial movements, referred to as action identification (describing how a face is moving, such as smiling) or it might involve inferring that a particular mental state caused the actions, referred to as **mental inference** or **mentalizing** (inferring why the action is performed, such as a state of happiness; Vallacher & Wegner, 1987). In principle, the categorization could also involve inferring a situational cause for the actions, but in practice, this question is rarely investigated in studies of emotion perception. The overwhelming majority of studies ask participants to make mental inferences, although, as we discuss later in this section, there appears to be important cultural variation in whether emotions are perceived as situated actions or as mental states that cause actions.

The use of posed configurations of facial movements in assessments of emotion perception. In the majority of the experiments that study emotion perception, researchers ask participants to infer emotion in photographs of posed facial configurations (such as those in Fig. 4, but without the FACS codes). In most studies, the configurations have been posed by people who were not in an emotional state when the photos were taken. In a growing number of studies, the poses are created with computer-generated humans who have no actual emotional state. As a consequence, it is not possible to assess the accuracy (i.e., validity) of perceivers' emotional inferences and, correspondingly, data from emotion-perception studies should not be interpreted as support for the validity of the common view of emotional expressions (except insofar as these are simply stipulated to be the consensus). As is the case in expression-production studies, it is more appropriate to interpret participants' responses in terms of their agreement (or consensus) with common beliefs (which may vary by language and culture).

Even more serious is the fact that the proposed expressive facial configurations in Figure 4, which are routinely used as stimuli in emotion-perception studies, do not capture the wider range of muscle movements that are observed when people actually express instances of these emotion categories in the lab or in everyday life. A recent study that mined more than 7 million images from the Internet (Srinivasan & Martinez, 2018; for method, see Box 7 in the Supplemental Material) identified multiple facial configurations associated with the same

emotion-category label and its synonyms—17 distinct facial configurations were associated with the word *happiness*, five with *anger*, four with *sadness*, four with *surprised*, two with *fear*, and one with *disgust*. The different facial configurations associated with each emotion word were more than mere variations on a universal core expression—they were distinctive sets of facial movements.[29]

Measuring emotion perception. The typical emotion perception experiment takes one of several forms, summarized in Table 4. Choice-from-array tasks, in which participants are asked to match photos of facial configurations and emotion words (with or without brief stories), have dominated the study of emotion perception since the 1970s. For example, a meta-analysis of emotion-perception studies published in 2002 summarized 87 studies, 83 (95%) of which exclusively used a choice-from-array response method (Elfenbein & Ambady, 2002). This method has been widely criticized for more than 2 decades, however, because it limits the possibility of observing evidence that could disconfirm the common view. Choice-from-array tasks strongly constrain the possible meanings that participants can infer in a facial configuration, such as a photograph of a scowl, because they can choose only the options provided in the experiment (usually a small number of emotion words). In fact, the preponderance of choice-from-array tasks in the scientific study of emotion perception has been identified as one important factor that has helped perpetuate and sustain the common view (Russell, 1994). Other tasks exist for assessing emotion perception (see Table 4), including those that use a **free-labeling** method, where participants are asked to freely nominate words to label photographs of posed facial configurations, rather than choosing a word from a small set of predefined options. For example, on viewing a scowling configuration, participants might offer responses like "angry," "sad," "confused," "hungry," or even "wanting to avoid a social interaction." By allowing participants more freedom in how they infer meaning in a facial configuration, free labeling makes it equally possible to observe evidence that could either support or disconfirm the common view.

Recent innovations in measuring emotion perception use computer-generated faces or heads rather than photographs of posed human faces. One method, called **reverse correlation,** measures participants' internal model of emotional expressions (i.e., their mental representations of which facial configurations are likely to express instances of emotion) by observing how participants label an avatar head that displays random combinations of animated facial action units (Yu, Garrod, & Schyns, 2012; for a review, see Jack, Crivelli, & Wheatley, 2018; Jack & Schyns, 2017). As each pattern appears on the computer screen (on a given test trial), participants infer its emotional meaning by choosing an emotion label from a set of options (a choice-from-array response). After thousands of trials, researchers estimate the statistical relationship between the dynamic patterns of facial movements and each emotion word (e.g., disgust) to reveal participants' beliefs about which facial configurations are likely to express different emotion categories.

A second approach using computer-generated faces has participants interact with more fully developed *virtual humans* (Rickel et al., 2002), also known as *embodied conversational agents* (Cassell et al., 2000). Software-based virtual humans look like and act like people (for examples, see Fig. 9). They are similar to characters in video games in their surface appearance and are designed to interact face-to-face with humans using the same verbal and nonverbal behaviors that people use to interact with one another. The underlying technologies used to realize virtual humans vary considerably in approach and capability, but most virtual-human models can be programmed to make context-sensitive, dynamic facial actions that would, when used by a person, typically communicate emotional information to other people (see Box 11 in the Supplemental Material for discussion). The majority of the scientific studies with virtual humans were not designed to test whether human participants infer specific emotional meaning in a virtual human's facial movements, but their design makes them useful for studying when and how facial movements take on meaning as emotional expressions: Unlike all the other ways of assessing emotion perception discussed so far, which ask participants to make explicit inferences about the emotional cause of facial configurations, interactions with virtual humans offers the possibility of learning how a participant implicitly infers emotional meaning during social interactions.

Testing the common view of emotion perception: interpreting the scientific observations. Traditionally, in most experiments, if participants reliably infer the hypothesized emotional state from a facial configuration (e.g., inferring anger from a scowling configuration) at levels that are greater than what would be expected by chance, then this is taken as evidence that people "recognize that emotional state in its facial display." It is more scientifically correct, however, to interpret such observations as evidence that people *infer* an emotional state (i.e., they consistently make a reverse inference) at greater-than-chance levels. Only when reverse inferences are observed in a reliable and specific way within an experiment can scientists reasonably infer that participants are perceiving an instance of a certain emotion category in a certain facial configuration; technically, the inference holds only for emotion perception as it occurs in the particular situations contained in the experiment (because situations

Table 4. Pros and Cons of Common Tasks for Measuring Explicit Emotion Perception

General considerations

Participants are typically asked to infer emotional meaning in posed, rather than spontaneous, facial configurations. Spontaneous or candid facial configurations typically produce much lower levels of agreement in emotion-perception studies (e.g., Kayyal & Russell, 2013; Naab & Russell, 2007).

Participants are typically asked to infer emotional meaning in static, nonmoving facial configurations (i.e., in photographs rather than movies). This reduces the ecological validity of the findings for how people infer emotional meaning in faces in the real world. In the real world, people have to infer when a set of movements begin and end; this is called discrimination or detection. Moreover, there is information in the dynamics of facial movements (Jack & Schyns, 2017; Krumhuber, Kappas, & Manstead, 2013), but dynamic facial movements, particularly when they are spontaneous, do not always produce higher levels of agreement in emotion-perception studies. Dynamic movements add realism and intensity and improve levels of agreement primarily when movements are degraded or are artificial.

Participants are typically asked to infer emotional meaning in exaggerated facial configurations, which are said to have greater "source clarity" (Ekman, Friesen, & Ellsworth, 1972). They reduce the ecological validity of the findings for how people infer emotional meaning in faces in the real world. The facial configurations used in most experiments (see Fig. 4) are caricatures—they are exaggerated to maximally distinguish one from the another. Caricatures are easier to label (categorize) than are typical stimuli, particularly when the categories in question are highly interrelated (Goldstone, Steyvers, & Rogosky, 2003).

Participants are typically asked to infer emotional meaning in highly selected facial configurations. In early studies, a smaller set of exaggerated facial configurations were culled from much larger sets of posed faces (involving several thousand faces; for a discussion, see Gendron & Barrett, 2017; Russell, 1994).

Only a single task used in most experiments (i.e., participants are asked to infer emotion in facial configurations via one method of responding). Ideally, multiple tasks should be used with the same population of participants to determine whether convergent results are obtained. This approach is rarely taken, but for an example, see Crivelli, Jarillo, Russell, & Fernández-Dols, 2016; Gendron, Roberson, van der Vyver, & Barrett, 2014b; Gendron, Hoemann, et al., 2018).

Test–retest reliability is rarely evaluated but is critical. A number of contextual factors are known to influence judgments, including a perceiver's internal state. Test–retest assessments are rarely done for practical reasons.

Most experiments ask participants to infer emotion in a disembodied face, alone, without context. This reduces the ecological validity of the findings for how people infer emotional meaning in faces in the real world. In addition, a growing number of experiments now show that context is an important, and sometimes dominant, source of information when people infer emotional meaning in a facial configuration. (See Box 3 in the Supplemental Material.) For example, situational information tends to dominate perception of emotion in faces in common, everyday events (Carrera-Levillain & Fernandez-Dols, 1994), even when situations are more ambiguous than the exaggerated facial configurations being judged (Carroll & Russell, 1996, Study 3).

Most studies do not report evidence about the specificity of emotion perceptions or the frequency with which people infer the nonintended emotional meaning from a facial configuration.

Until recently, the large majority of experiments included only one pleasant emotion category (happiness) among several unpleasant emotion categories (anger, fear, sadness, etc.). This may be one reason that agreement rates are so high for smiles. In the past few years, experiments have included a larger variety of pleasant emotion categories (pride, awe, gratitude, etc.), but there continues to be debate over whether these emotion categories are expressed with consistent, specific facial configurations.

Choice-From-Array Task: matching photos of facial configurations and emotion words (with or without brief stories). Response options are limited to those provided in the task.

Participants are (a) shown a facial configuration and asked to infer its emotional meaning by choosing an emotion word from a small set of words or (b) presented with an emotion word that labels an emotion category (e.g., sadness) or a brief story about a typical instance of an emotion category (e.g., "the boy's much loved dog just died and he is sad") along with two or three photographs of faces (typically posed into one of the configurations presented in Fig. 4) and then asked to choose the facial configuration that they judge best matches the emotional episode described in the word or vignette. Typically, each emotion category is represented by a single scenario.

Words influence how the brain processes visual inputs from faces (e.g., Doyle & Lindquist, 2018; Gendron, Lindquist, Barsalou, & Barrett, 2012). Stories can prime action perceptions, as well. More generally, choice-from-array tasks have been shown to encourage biased perceptual responding using a signal detection analysis (e.g., DeCarlo, 2012). Choice-from-array tasks are still commonly used, however, because they are easy and efficient and straightforward for participants to understand. Choice-from-array responses are easy for scientists to score. Most studies using continuous judgments (rather than forced choice) find that participants do not infer emotional meaning in facial configurations in a yes/no or on/off sort of way (Russell, 1994).

The fact that participants are exposed to the same facial configurations and emotion words over and over allows them to learn the intended pairings even if they do not know them to begin with (N. L. Nelson & Russell, 2016).

An emotion word does not necessarily have a unique correspondence to a single emotion category for all people in a given culture (i.e., they may differ in **emotional granularity**; L. F. Barrett, 2004, 2017b; Lindquist & Barrett, 2008) or people from different cultures. Concerns about individual word meaning is why choice-from-array using stories might be preferable to those using single words.

(continued)

Table 4. (Continued)

A small range of answers is predetermined by the experimenter, making it easier for participants to provide the answers scientists expect. For example, by constraining which words participants were allowed to choose from, frowns were consensually labeled as fear, wide-eyed, gasping faces were labeled as surprise (Russell, 1993). Scowling faces are more likely to be perceived as fearful when paired with the description of danger (Carroll & Russell, 1996, Study 1) and appear determined or puzzled depending on the story they are presented with (Carroll & Russell, 1996, Study 2).

Participants are asked to make yes/no decisions when assigning a facial configuration to an emotion category. Multiple emotion words may apply to a single configuration (i.e., people might infer more than one emotional meaning in a face), but the option to infer multiple emotional meanings is rarely given to participants. Continuous judgments using, for example, a Likert-type scale ranging from 1 to 7 would solve both of these problems and also allow analysis of the similarity among facial configurations (which evidence shows is important; e.g., Jack, Sun, Delis, Garrod, & Schyns, 2016; Kayyal & Russell, 2013). Similarity allows scientists to discover the emotional meanings that people implicitly assign to a facial configuration, rather than having people explicitly state them (see further discussion of similarity below).

A participant might decide that no emotion word provided applies to a facial configuration, but the option to respond this way is rarely given to participants (they are usually forced to choose an emotion word; for discussion, see Frank & Stennett, 2001). See Cordaro et al. (2017) for an example of this design feature.

If a participant hears a story and is asked to choose between two faces (e.g., a scowl and smile), he or she can give the expected answer (e.g., scowl) simply by figuring out that "smile" is NOT correct. For example, after hearing a story about anger, a participant is shown a scowl and a smile and can choose the scowl merely by realizing the smile is not correct (on the basis of valence). This is similar to getting an answer right on a multiple-choice test by eliminating all the alternatives—you do not actually know the right answer, but you figured it out because of the structure of the task. A similar point can be made about showing a single face and asking participants to label it with a word by selecting from among a small set of options. Participants use a process of elimination strategy: Words that are not chosen on prior trials are selected more frequently, inflating agreement levels (DiGirolamo & Russell, 2017).

If participants hear a story about anger and must choose between a scowl and a smile, they can figure out that the scowl is correct merely because they are distinguishing between negative (scowl) and positive (smile). If participants hear a story about anger and must choose between a scowl and a frown, they can figure out that the scowl is correct merely because they is distinguishing between high arousal (scowl) and low arousal (frown).

In tasks that involve brief stories or vignettes about emotion, only one typical story is offered for each emotion category, making it more difficult to observe any variation within a category.

Free-Sorting Task: photos of facial configurations are sorted into groupings, such that each grouping represents a perceived category. Cue-to-Cue Matching: matching photos of facial configurations to a recording of posed vocalization.

Most participants still spontaneously use words to guide their sorting and organize their groupings. Free sorting and cue matching are ideal for preverbal participants or those with semantic deficits (e.g., Lindquist, Gendron, Barrett, & Dickerson, 2014).

Similarity Task: Judgments between pairs of facial configurations. Perceptual Matching: Indicating whether two photos of facial configurations belong to the same emotion category.

It is inefficient and time-consuming to judge the similarity of all pairs of facial configurations. For a set of 100 faces, this requires $(100 \times 100)/2 = 5{,}000$ different similarity judgments. Participants can arrange face stimuli on a computer screen, and all pairwise similarity judgments can be computed (the SPAM method proposed by Goldstone, 1994; e.g., see Hout, Goldinger, & Ferguson, 2013). This procedure also solves the problem that the same pair of stimuli will have a different judged similarity depending on which item is presented first if face pairs are presented sequentially (the judged similarity of two objects, A and B, can depend on the order in which they are presented; the similarity between A and B is not always judged to be the same as that between B and A; Tversky, 1977). Other advantages are that categories can be discovered, rather than prescribed, and verbal associations are minimized. Analyses of similarity judgments typically yield more continuous similarity relations between emotion categories along affective dimensions (see Russell & Barrett, 1999).

Free-Labeling Task: photos of facial configurations are labeled with words offered by participants (unconstrained by experimenter).

Forcing people to translate faces into words is not a good idea, because much of the information from faces cannot be easily captured in words (Ekman, 1994). In addition, facial expressions did not evolve to represent specific verbal labels (Ekman, 1994, p. 270). "Regardless of the language, of whether the culture is Western or Eastern, industrialized or preliterate, these facial expressions are labeled with the same emotion terms: happiness, sadness, anger, fear, disgust, and surprise" (Ekman, 1972, p. 278). These are not special criticisms of free-labeling studies—it applies to all studies that ask people to label a face with words, including the choice-from-array tasks.

(continued)

Table 4. (Continued)

There is no widely accepted method for scoring (i.e., categorizing) freely provided responses. (Ekman, 1994, p. 274). Most scientists group together similar words (synonyms), so that a variety of words can be used to show evidence of a correct response (e.g., a frowning face, which is the proposed expression for sadness, could be labeled as *sad*, *grieving*, *disappointed*, *blue*, *despairing*, and so on. Scientists routinely use databases that indicate synonyms (e.g., WORDNET; used in Srinivasan & Martinez, 2018). In addition, it is possible to do data-driven groupings of emotion words into semantic categories (e.g., Jack et al., 2016; Shaver, Schwartz, Kirson, & O'Connor, 1987). The more serious problem is that early studies using free labeling (e.g., Boucher & Carlson, 1980; Izard, 1971) did not provide enough information in the method sections about how freely provided labels were grouped.
Using freely chosen labels in a study of different cultures is difficult because it may be hard to find adequate translations (Ekman, 1994, p. 274). A given emotion word, such as *sadness*, can correspond to different emotion concepts (with different features) in different languages (e.g., Wierzbicka, 1986, 2014). A single emotion word in one language can refer to more than one concept in another language (e.g., Pavlenko, 2014). Some languages have no one-to-one translation for English emotion words, and some emotion concepts in other languages are not directly translatable into English emotion words (see L. F. Barrett, 2017b; Russell, 1991; Jack et al., 2016). This is not a special criticism of free-labeling studies, however; it holds for any experiment that uses emotion words requiring translation, including choice-from-array tasks. A standard solution to this problem is to use both forward and backward translation (e.g., a word spoken in Hadzane is translated into English and then back translated into Hadzane; this process estimates whether or not the translation has fidelity). An even better method is to elicit features for the emotion words in question, including typicality of those features, to determine the fidelity of translation (e.g., de Mendoza, Fernández-Dols, Parrott, & Carrera, 2010).
Scientifically, issues with translation are manageable if scientists allow phrases to stand in for specific words.
Using only single words will always fail to capture much of the rich information in faces. Participants often provide multiple words or even longer descriptions of situations, behaviors, or behaviors in situations (e.g., see Gendron et al., 2014b; Russell, 1994). Such data are time consuming to code and analyze.
Even when participants are told that photographs are of people trying to express an emotion, they often offer nonemotion labels. For example, Izard (1971) found that people offered labels such as *deliberating*, *clowning*, *skepticism*, *pain*, and so on (as reported in Russell, 1994).This is not necessarily evidence that participants did not understand the task asked of them. It might be evidence that these facial configurations are not specific for expressing emotions.

Note: Response tasks are arrayed in order from those that most constrain participants' responses (making it difficult to observe evidence that can disconfirm common beliefs about emotion) to those that least constrain participants' responses (making it easier to observe variation and disconfirm common beliefs). For detailed design concerns about choice-from-array tasks, see Russell (1994, 1995).

are never randomly sampled). If the emotion-perception evidence is *replicated* across experiments that sample people from the same culture, then the interpretation can be generalized to emotion perceptions in that culture. Only when the findings generalize across cultures—that is, are replicated across experiments that sample people from different cultures—is it reasonable to conclude that people universally infer a specific emotional state when perceiving a specific facial configuration. These findings can be interpreted as evidence about the reliability and specificity of *producing* emotional expressions if the coevolution assumption is valid (i.e., that emotional expressions and their perception coevolved as an integrated signaling system; Ekman et al., 1972; Jack et al., 2016; Shariff & Tracy,

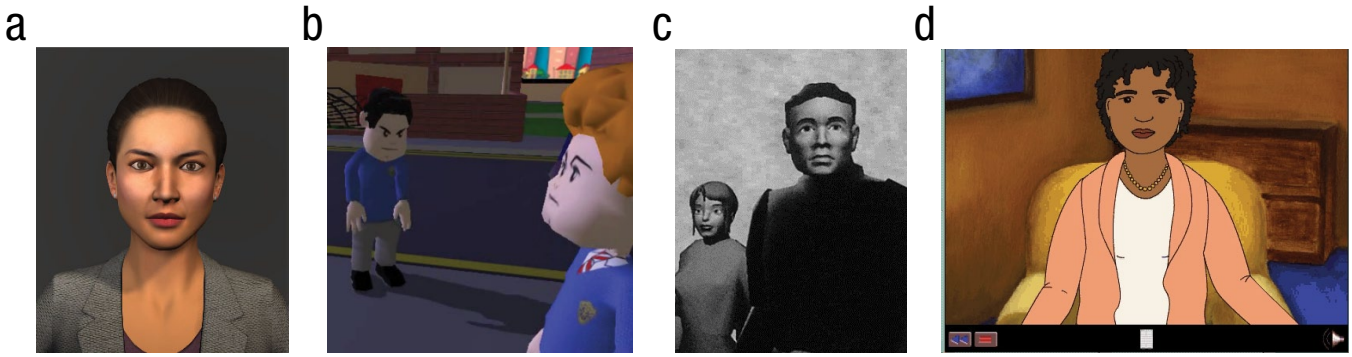


Fig. 9. Examples of virtual humans. Virtual humans are software-based artifacts that look like and act like people. (a) The system that used this virtual human is described in Feng, Jeong, Krämer, Miller, and Marsella (2017). (b) This virtual human is reproduced from Zoll, Enz, Aylett, and Paiva (2006). (c) This virtual human is reproduced from Hoyt, C., Blascovich, J., and Swinth, K. (2003). Social inhibition in immersive virtual environments. *Presence, 12*(2), 183–195, courtesy of The MIT Press. (d) The system that was used to create this virtual human is described in Marsella, Johnson, and LaBore (2000).

2011). The findings can be interpreted as evidence about emotion **recognition** only if the reverse inference has been verified as valid (i.e., if it can be verified that the person in the photograph is, indeed, in the expected emotional state).

Studies of healthy adults from the United States and other developed nations

Studies that measure emotion perception with choice-from-array tasks. The most recent meta-analysis of emotion-perception studies was published by Elfenbein and Ambady (2002). It statistically summarized 87 experiments in which more than 22,000 participants from more than 20 cultures around the world inferred emotional meaning in facial configurations and other stimuli (e.g., posed vocalizations). The majority of participants were sampled from larger-scale or developed countries, including Argentina, Brazil, Canada, Chile, China, England, Estonia, Ethiopia, France, Germany, Greece, Indonesia, Ireland, Israel, Italy, Japan, Malaysia, Mexico, the Netherlands, Scotland, Singapore, Sweden, Switzerland, Turkey, the United States, Zambia, and various Caribbean countries. The majority of studies (95%) used posed facial configurations; only four studies had participants label spontaneous facial movements, a dramatic example of the challenges facing validity that we discussed earlier. All but four studies used a choice-from-array response method to measure emotion inferences, a good example of the challenges facing hypothesis disconfirmation that we discussed earlier.

The results of the meta-analysis, presented in Figure 10, reveal that perceivers inferred emotions in the facial configurations of Figure 4 in line with the common view, well above chance levels (using the criteria set out by Haidt and Keltner, 1999, presented in Table 2 of the current article). Results provided strong evidence that, when participants viewed posed facial configurations made by people from their own culture, they reliably perceived the expected emotion in those configurations: Scowling facial configurations were perceived as anger expressions, wide-eyed facial configurations were perceived as fear expressions, and so on, for all six emotion categories. Moderate levels of reliability were observed when perceivers were labeling facial configurations posed by people from other cultures; this difference in reliability between same- and cross-culture differences is referred to as an **in-group advantage** (see Box 12, in the Supplemental Material). The majority of emotion-perception studies did not report whether the hypothesized facial configurations were perceived with any specificity (e.g., how likely was a scowl to be perceived as expressing an instance of emotion categories other than anger, or as an instance of a mental category that is not considered emotional). Without information about specificity, no firm conclusions can be drawn about the emotional meaning of the facial configurations in Figure 4, especially for the translational purpose of inferring someone's emotional state from their facial comportment in real life.

Nonetheless, most of the studies cited in the Elfenbein and Ambady (2002) meta-analysis interpret their reliability findings alone (i.e., inferring anger from a scowling face, disgust from a nose-wrinkled face, fear from a wide-eyed, gasping face, etc.) as evidence of accurate reverse inferences. Such interpretations may explain why many scientists who study emotion, when surveyed, indicated that they believe compelling evidence exists for the hypothesis that certain emotion categories are each expressed with a unique, universal facial configuration (see Ekman, 2016) and interpret variation in emotional expressions to be caused by cultural learning that modifies what are presumed to be inborn universal expressive patterns (e.g., Cordaro et al., 2018; Ekman, 1972; Elfenbein, 2013). Cultural learning has also been hypothesized to modify how people "decode" facial configurations during emotion perception (Buck, 1984).

Studies that measure emotion perception with free-labeling tasks. Experimental methods that place fewer constraints on participants' inferences (Table 4) provide considerably less support for the common view of emotional expressions. In the least constrained experimental task, called free labeling, perceivers freely volunteer a word (emotion or otherwise) that they believe best captures the meaning in a facial configuration, rather than choosing from a small set of experimenter-provided options. In urban samples, participants who freely label facial configurations produce the expected emotion labels with weak reliability (when labeling spontaneously produced facial configurations) to moderate reliability (when labeling posed facial configurations). Participants' responses usually reveal weak specificity when specificity is assessed at all (for examples and discussion, see Russell, 1994; also see Naab & Russell, 2007).

For example, participants in a study by Srinivasan and Martinez (2018) were sampled from multiple countries. They were asked to freely provide emotion words in their native languages (English, Spanish, Mandarin Chinese, Farsi, Arabic, and Russian) to label each of 35 facial configurations that had been cross-culturally identified. Their labels provided evidence of a moderately reliable correspondence between facial configurations and emotion categories, but there was no evidence of specificity (see Fig. 11).[30] Multiple facial configurations were associated with the same emotion category label (e.g., 17 different facial configurations were associated with the expression of happiness, five with

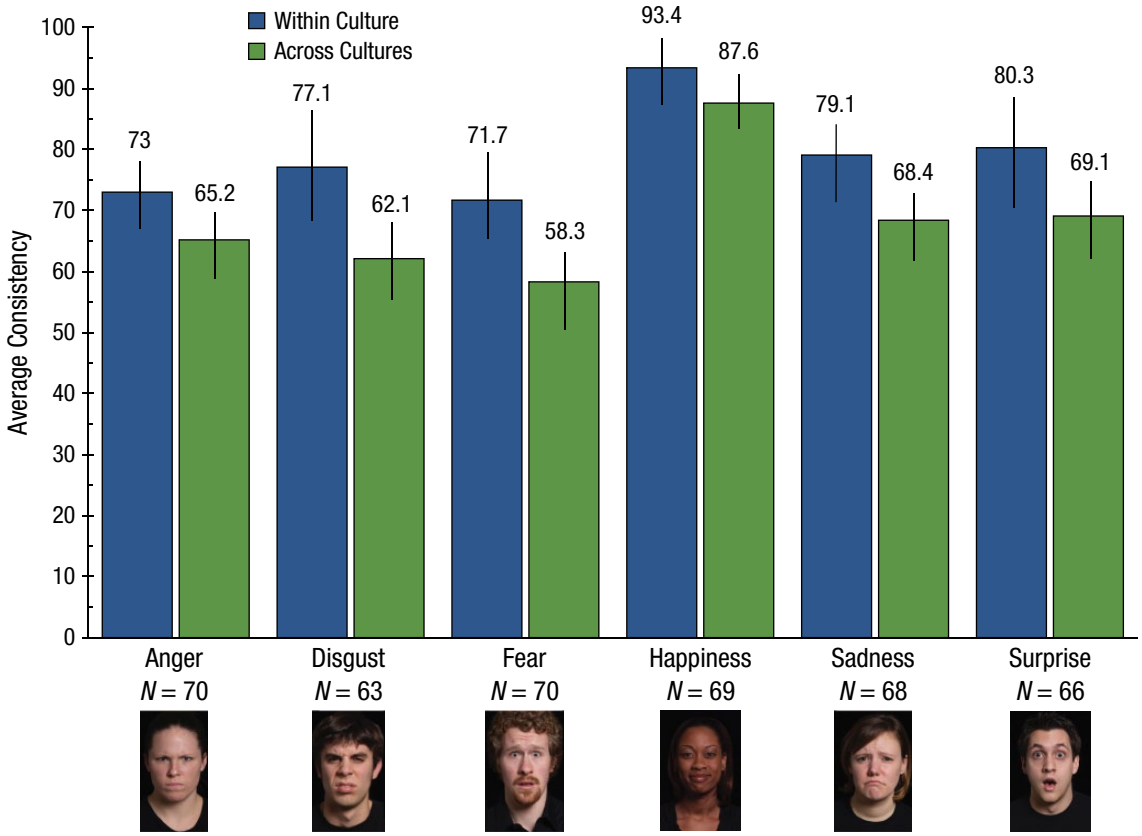


Fig. 10. Emotion-perception findings. Average effect sizes for perceptions of facial configurations in which 95% of the articles summarized used choice-from-array to measure participants' emotion inferences. Data are from Elfenbein and Ambady (2002). The images presented on the *x*-axis are for illustrative purposes only and were not necessarily used in the articles summarized in this meta-analysis.

anger, four with sadness, four with surprise, two with fear, and one with disgust). This many-to-many mapping is inconsistent with the common view that the facial configurations in Figure 4 are universally recognized as expressing the hypothesized emotion category, and they give evidence of variation that is far beyond what is proposed by the basic-emotion view. Some of this variability may come from different cultures and languages, but there is variability even within a single culture and language. Evidence of this many-to-many mapping is also apparent in free-labeling tasks in small-scale, remote samples as well (Gendron, Crivelli, & Barrett, 2018), which we discuss in the next section.

Studies that measure emotion perception with the reverse-correlation method. Using a choice-from-array response method with the reverse-correlation method is an inductive way to learn people's beliefs about which facial configurations express the instances of an emotion category (for reviews, see Jack et al., 2018; Jack & Schyns, 2017). In such studies, participants view thousands of random combinations of AUs that are computer generated on an avatar head and label each one by choosing an emotion word from a set of predefined options. All of the facial configurations labeled with the same emotion word (e.g., anger) are then statistically combined for each participant to estimate that person's belief about which facial movements express instances of the corresponding emotion category. One recent study using the reverse correlation method with participants from the United Kingdom and China found evidence of variation in the facial movements that were judged to express a single emotion category as well as similarity in the facial movements that were judged to express different categories (Jack et al., 2016). The study first identified groupings of emotion words that are widely discussed in the scientific literature (which, we should note, is dominated by English): 30 English words grouped into eight emotion categories for the sample from the United Kingdom (happy/excited/love, pride, surprise, fear, contempt/disgust, anger, sad, and shame/embarrassed) and 52 Chinese words grouped into 12 categories in the Chinese

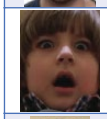
	Emotion Label Offered (% of Responses)							
Face	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Nonaffective	Action
	39.92		7.19		7.93		12.92	3.96
	11.38	24.94		2.63	10.05		12.39	
	4.14	4.01	34.75	4.12	4.13	18.61	2.73	
				48.84		3.17	1.84	10.01
	8.31		8.60		37.41		13.40	
	2.16	6.02	12.23	6.01		31.90		1.82

Fig. 11. Free labeling of facial configurations across five language groups. Data are from Srinivasan and Martinez (2018). The proportion of times participants offered emotion-category labels (or their synonyms) are reported. The facial configurations presented were chosen by researchers to represent the best match to the hypothetical facial configurations in Figure 4 on the basis of the action units (AUs) present. No configuration discovered in this study exactly matches the AU configurations proposed by Darwin or documented in prior research. According to standard scientific criteria, universal expressions of emotion should elicit agreement rates that are considerably higher than those reported here, generally in the 70% to 90% range, even when methodological constraints are relaxed (Haidt & Keltner, 1999). Specificity data are not available for the Elfenbein and Ambady (2002) meta-analysis.

sample (joyful/excitement, pleasant surprise, great surprise/amazement, shock/alarm, fear, disgust, anger, sad, embarrassment, shame, pride, and despise). The reverse-correlation method revealed 62 separate facial configurations: The same emotion category in a given culture was associated with multiple models of facial movements because synonyms of the same emotion category were associated with distinctive models of facial movements.

Amidst this variability, Jack and colleagues also found that these 62 separate facial configurations could be summarized as four prototypes, which are presented in Figure 8 along with the corresponding emotion words with which they were frequently associated. Each prototype was described with a unique set of affective features (combinations of valence, arousal and dominance). A comparison of the four estimated configurations with the common view presented in Figure 4 and with the basic-emotion hypotheses listed in Table 1 reveals some striking similarities: Configuration 1 in Figure 8 most closely resembles the proposed expression for happiness, Configuration 2 is similar to a combination of the proposed expressions for fear and anger, Configuration 3 most closely resembles the proposed expression for surprise, and Configuration 4 is similar to a combination of the proposed expressions for disgust and anger.[31] Taken together, these findings suggest that, at the most general level of description, participants' beliefs about emotional expressions (i.e., their internal models of which facial movements expressed which emotions) were consistent with the common view (indeed, they could be taken to constitute part of the common view); when examined in finer detail with more granularity, however, the findings also give evidence of substantial within-category variation in beliefs about the facial movements that express instances of the same emotion category. This observation suggests that the way the common view is often described in scientific reviews, depicted in the media, and used in many applications does not, in fact, do justice to people's more varied beliefs about facial expressions of emotion.

Studies that implicitly assess emotion perception during interactions with virtual humans. Designers typically study how a virtual human's expressive movements influence an interaction with a human participant. Much of the early research modeling expressive movements in virtual humans focused on endowing them with the facial expressions proposed in Figure 4. A number of studies have endowed virtual humans with blends of these configurations (Arya, DiPaola, & Parush, 2009; Bui, Heylen, Poel, & Nijholt, 2004). Designers are also inspired by people's beliefs about how emotions are expressed. Actors, for example, have been asked to pose facial configurations that they believe express emotions, which are then processed by graphical and machine-learning algorithms to craft the relation between emotional states and expressive movements (Alexander, Rogers, Lambeth, Chiang, & Debevec, 2009). In another study, human subjects used a specially designed software tool to craft animations of facial movements that they believed express certain mental categories, including emotion categories. Then, other human subjects judged the crafted facial configurations (Ochs, Niewiadomski, & Pelachaud, 2010). Increasingly, data-driven methods are used that place people in emotion-eliciting conditions, capture the facial and body motion, and then synthesize animations from those captured motions (Ding, Prepin, Huang, Pelachaud, & Artie`res, 2014; Niewiadomski et al., 2015; N. Wang, Marsella, & Hawkins, 2008).

In general, studies with virtual humans show nicely how the situational context influences people's inferences about the meaning of facial movements (de Melo, Carnevale, Read, & Gratch, 2014). For example, in a game that allowed competition and cooperation (Prisoner's Dilemma, Pruitt & Kimmel, 1977), a virtual human who smiled after making a competitive move evoked more competitive and less cooperative responses from human participants compared with a virtual human using an identical strategy in the game (tit-for-tat) but who smiled after cooperating. Virtual humans who make a verbal comment about a film that is inconsistent with their facial movements, such as saying they enjoyed the film but grimacing, quickly followed by a smile, were perceived as less reliable, trustworthy, and credible (Rehm & Andre, 2005).

The dynamics of the facial actions, including the relative timing, speed, and duration of the individual facial actions, as well as the sequence of facial muscle movements over time, offer information over and above the mere presence or absence of the movements themselves and have an important influence on how human perceivers interpret facial movements (e.g., Ambadar, Cohn, & Reed, 2009; Jack & Schyns, 2017; Keltner, 1995; Krumhuber, Kappas, & Manstead, 2013) and how much they trust a virtual human during a social interaction (Krumhuber, Manstead, Cosker, Marshall, & Rosin, 2009). Research with virtual humans has shown that the dynamics of facial muscle movements are critical for them to be perceived as emotional expressions (Niewiadomski et al., 2015; Ochs et al., 2010). These findings are consistent with research showing that the temporal dynamics carry information about the emotional meaning of facial movements that are made by real humans (e.g., Kamachi et al., 2001; Krumhuber & Kappas, 2005; Sato & Yoshikawa, 2004; for a review, see Krumhuber et al., 2013).[32]

Summary. Whether people can reliably perceive emotions in the expressive configurations of Figure 4, as predicted by the common view, depends on how participants are asked to report or register their inferences (see Table 3). Hundreds of experiments have asked participants to infer the emotional meaning of posed, exaggerated facial configurations (such as those presented in Figure 4) by choosing a single emotion word from a small number of options offered by scientists, called choice-from-array-tasks. This experimental approach tends to generate moderate to strong evidence that people reliably label scowling facial configurations as angry, frowning facial configurations as sad, and so on for all six emotion categories that anchor the common view. Choice-from-array tasks severely limit the possibility of observing evidence that can disconfirm the common view of emotional expressions, however, because they restrict participants' options for inferring the psychological meaning of facial configurations by offering them a limited set of emotion labels. (As we discuss below, when people are provided with labels other than angry, sad, afraid, and so on, they routinely choose them; also see Carroll & Russell, 1996; Crivelli et al., 2017). In addition, the specificity of emotion-perception judgments is largely unreported.

Scientists often go further and interpret the better-than-chance reliability findings from these studies as evidence that scowls are expressions of anger, frowns are expressions of sadness, and so on. Such inferences are not sound, however, because most of these studies ask participants to infer emotion from posed, static faces that are likely limited in their validity (i.e., people posing facial configurations such as those depicted in Figure 4 are unlikely to be experiencing the hypo thesized emotional state). Furthermore, other ways of assessing emotion perception, such as the reverse-correlation method and free-labeling tasks, find much weaker evidence for reliability of emotion inferences. Instead, they suggest that what people actually infer and believe about facial movements incorporates considerable variability: In short, the common view depicted in many reviews, summaries, the media, and used in numerous applications is not an accurate

reflection of what people believe about facial expressions of emotion, when these beliefs are probed in more detail (in a way that makes it possible to observe evidence that could disconfirm the common view). In the next section, we discuss scientific evidence from studies of emotion perception in small-scale remote cultures, which further undermines the common view.

Studies of healthy adults living in small-scale, remote cultures

A growing number of studies examine emotion perception in people from remote, nonindustrialized cultural groupings. A more in-depth review of these studies can be found in Gendron, Crivelli, and Barrett (2018). Our goal here is to summarize the trends found in this line of research (see Table 5).

Studies that measure emotion perception with choice-from-array tasks. During the period from 1969 to 1975, between five and eight small-scale samples from remote cultures in the South Pacific were studied with choice-from-array tasks; the goal was to investigate whether these participants perceived emotional expressions in facial movements in a manner similar to that of people from the United States and other industrialized countries of the Western world (see Fig. 12a). Our uncertainty in the number of samples stems from reporting inconsistencies in the published record (see note to Table 5). We present the findings here according to how the original authors reported their findings, despite the inconsistencies. Five samples performed choice-from-array tasks, three in which participants chose a photographed facial configuration to match one brief vignette that described each emotion category (Ekman, 1972; Ekman & Friesen, 1971; Sorenson, 1975) and two in which they chose a photograph to match an emotion word (Ekman, Sorenson, & Friesen, 1969). All five samples performed some version of a choice-from-array task that provided strong evidence in support of cross-cultural reliability of emotion perception in small-scale societies. Evidence for specificity was not reported. Until 2008, *all claims* that anger, sadness, fear, disgust, happiness, and surprise are universally recognized (and therefore are universally expressed) were based largely on three articles (two of them peer reviewed) reporting on four samples (Ekman, 1972; Ekman & Friesen, 1971; Ekman et al., 1969).[33]

Since 2008, 10 verifiably separate experiments observing emotional inferences in small-scale societies have been published or submitted for publication. These studies involve a greater diversity of social and ecological contexts, including sampling five small-scale societies across Africa and the South Pacific (see Fig. 12b) that were tested with a greater diversity of research methods listed in Table 4, including tasks that allow for the possibility of observing cross-cultural variation in emotion perception and therefore the possibility of disconfirming the common view. Six samples registered their emotion inferences using a choice-from-array task, in which participants were given an emotion word and asked to choose the posed facial configuration that best matched it or vice versa (Crivelli, Jarillo, Russell, & Fernández-Dols, 2016; Crivelli, Russell, Jarillo, & Fernández-Dols, 2016; Crivelli et al., 2017, Study 2; Gendron, Hoemann, et al., 2018, Study 2; Tracy & Robins, 2008).

Only one study (Tracy & Robins, 2008) reported that participants selected an emotion word to match the facial configurations similar to those in Figure 4 more reliably than what would be expected by chance, and effects ranged from weak (anger and fear) to strong (happiness) with surprise and disgust falling in the moderate range.[34] Information about the specificity of emotion inferences was not reported. A close examination of the evidence from four studies by Crivelli and colleagues suggest weak to moderate levels of reliability for inferring happiness in smiling facial configurations (all four studies), sadness in frowning facial configurations (all four studies), fear in gasping, wide-eyed facial configurations (three studies), anger in scowling facial configurations (two studies), and disgust in nose-wrinkled facial configurations (three studies). A detailed breakdown of findings can be found in Box 13 in the Supplemental Material. None of the studies found specificity for any facial configuration, however, except that smiling was reported as unique to happiness, but that finding was not replicated across samples.[35]

The final study using a choice-from-array task with people from a small-scale, remote culture is important because it involves the Hadza hunter-gatherers of Tanzania (Gendron, Hoemann, et al., 2018, Study 2).[36] The Hadza are a high-value sample for two reasons. First, universal and innate emotional expressions are hypothesized to have evolved to solve the recurring fitness challenges of hunting and gathering in small groups on the African savanna (Pinker, 1997; Shariff & Tracy, 2011; Tooby & Cosmides, 2008); the Hadza offer a rare opportunity to study hunters and foragers who are currently living in an ecosystem that is thought to be similar to that of our Paleolithic ancestors.[37] Second, the population is rapidly disappearing (Gibbons, 2018). Before this study, the Hadza had not participated in any studies of emotion perception, although they have been the subject of social cognition research more broadly (H. C. Barrett et al., 2016; Bryant et al., 2016).

After listening to a brief story about a typical instance of anger, disgust, fear, happiness, sadness, and surprise, Hadza participants chose the expected facial

Table 5. Summary of Cross-Cultural Emotion Perception in Small-Scale Societies

Task	Culture	*N*	Citation	Level of support
Free labeling	Fore, PNG[a]	100	Sorenson (1975), Sample 2[b]	None
	Bahinemo, PNG	71	Sorenson (1975), Sample 3	None
	Hadza, Tanzania	43	Gendron, Hoemann, et al. (2018), Study 1	None
	Trobrianders, PNG	32[c]	Crivelli et al. (2017), Study 1	None
	Sadong, Borneo[d]	15	Sorenson (1975), Sample 4[e]	Strong
Cue-to-cue matching	Shuar, Ecuador	23	Bryant & Barrett (2008), Study 2	Weak
	Himba, Namibia[f]	65	Gendron et al. (2014)	None
Choice-from array: matching face and words	Fore, PNG[a]	32	Ekman, Sorenson, & Friesen (1969)[b]	None
	Mwani, Mozambique	36[c,g]	Crivelli, Jarillo, Russell, & Fernández-Dols (2016), Study 2	None
	Trobrianders, PNG	24[c]	Crivelli et al. (2017), Study 2	None
	Trobrianders, PNG	68[c,g]	Crivelli, Jarillo, Russell, & Fernández-Dols (2016), Study 1	None
	Trobrianders, PNG	36[c]	Crivelli, Russell, et al. (2016), Study 1a	None
	Dioula, Burkina Faso[a]	39	Tracy & Robins (2008), Study 2	Weak
	Sadong, Borneo[d]	15	Ekman et al. (1969)[e]	Strong
Choice-from array: matching face and scenario	Hadza, Tanzania	54	Gendron, Hoemann, et al. (2018), Study 2	None
	Dani, New Guinea[a]	34	Described in Ekman (1972)[h]	Moderate
	Fore, PNG[d]	189, 130[g]	Sorenson (1975), Sample 1[i]	Moderate
	Fore, PNG[d]	189, 130[g]	Ekman and Friesen (1971)[i]	Strong

Note. Findings for anger, disgust, fear, sadness, and surprise are summarized; happiness is the only pleasant category tested in all studies except for Tracy and Robins (2008). Therefore, in most studies, inferences of happiness in smiling faces do not uniquely reflect emotion perception and may be driven by valence perception (distinguishing pleasant from unpleasant). All studies used photographs of posed facial configurations that are similar to those in Figure 4, except Crivelli, Jarillo, Russell, and Fernández-Dols, 2016, Study 2, which used dynamic as well as static posed configurations, and Crivelli et al., 2017, Study 1, which used static spontaneous configurations. The Bryant and Barrett (2008) study was designed to examine emotion perception from vocalizations but is included because perceivers matched them to faces; in addition, participants were tested in a second language (Spanish) in which they received training. Choice-from-array studies (except Gendron et al. 2014a, Study 2, and Gendron, Hoemann, et al., 2018, Study 2) did not carefully control whether target facial configurations and foils could be distinguished by valence and/or arousal. All participants were adults unless otherwise specified. Levels of support: "None" indicates that reliability and specificity were at chance levels or that any level of reliability above chance was combined with evidence of no specificity. "Weak" indicates that reliability was between 20% and 40% (weak) for at least a single emotion category other than happiness combined with above chance specificity for that category or reliability between 41% and 70% (moderate reliability) for at least a single category other than happiness with unknown specificity. "Moderate" indicates that reliability was between 41% and 70% combined with any evidence of above-chance specificity for a category other than happiness or reliability above 70% (strong reliability) for at least a single category other than happiness with unknown specificity. "Strong" indicates strong evidence of reliability (above 70%) and strong evidence of specificity for at least a single emotion category other than happiness. It is questionable whether the Sadong and the Fore subgroup with more other-group contact should be considered isolated (see Sorenson, 1975, pp. 362 and 363), but we include them here to avoid falsely dichotomizing cultures as "isolated from" versus "exposed to" one another (Fridlund, 1994; Gewald, 2010). PNG = Papua New Guinea.

[a]Specificity levels were not reported. [b]Sorenson (1975), Sample 2, included three groups of Fore participants (those with little, moderate, and most other-group contact). The pattern of findings is nearly identical for the subgroup with the most contact and the data reported for the Fore in Ekman et al. (1969); Sorenson described using a free-labeling method, whereas Ekman et al. (1969) described using a choice-from-array method. [c]Ekman (1994) indicated, however, that he did not use a free-labeling method, implying that the samples may be distinct. Participants were adolescents. [d]Specificity was inferred from reported results. [e]The sample size, marginal means, and exact pattern of errors reported for the Sadong samples are identical in Sorenson (1975), Sample 4, and Ekman et al. (1969); Sorenson described using a free-labeling method and Ekman et al. (1969) described using a choice-from-array method in which participants were shown photographs and asked to choose a label from a small list of emotion words. [f]Traditional specificity and consistency tests are inappropriate for this method, but the results are placed here based on the original author's interpretation of multidimensional scaling and clustering results. [g]Participants were children. [h]The Dani sample reported in Ekman (1972) is likely to be a subset of the data from an unpublished manuscript. [i]Sorenson (1975), Sample 1 and Ekman and Friesen (1971) may be the same sample because the sample sizes and pattern of data are identical for all emotion categories except for the fear category, which is extremely similar, and for the disgust category, which includes responses for contempt in Ekman and Friesen (1971) but was kept separate in Sorenson (1975).

configuration more often than chance only when the target and foil could be distinguished by the affective property referred to as valence. The finding that Hadza participants were successfully inferring pleasantness and unpleasantness is consistent with anthropological studies of emotion (Russell, 1991), linguistic studies (Osgood, May, & Miron, 1975), and findings from other recent studies of participants from small-scale societies,

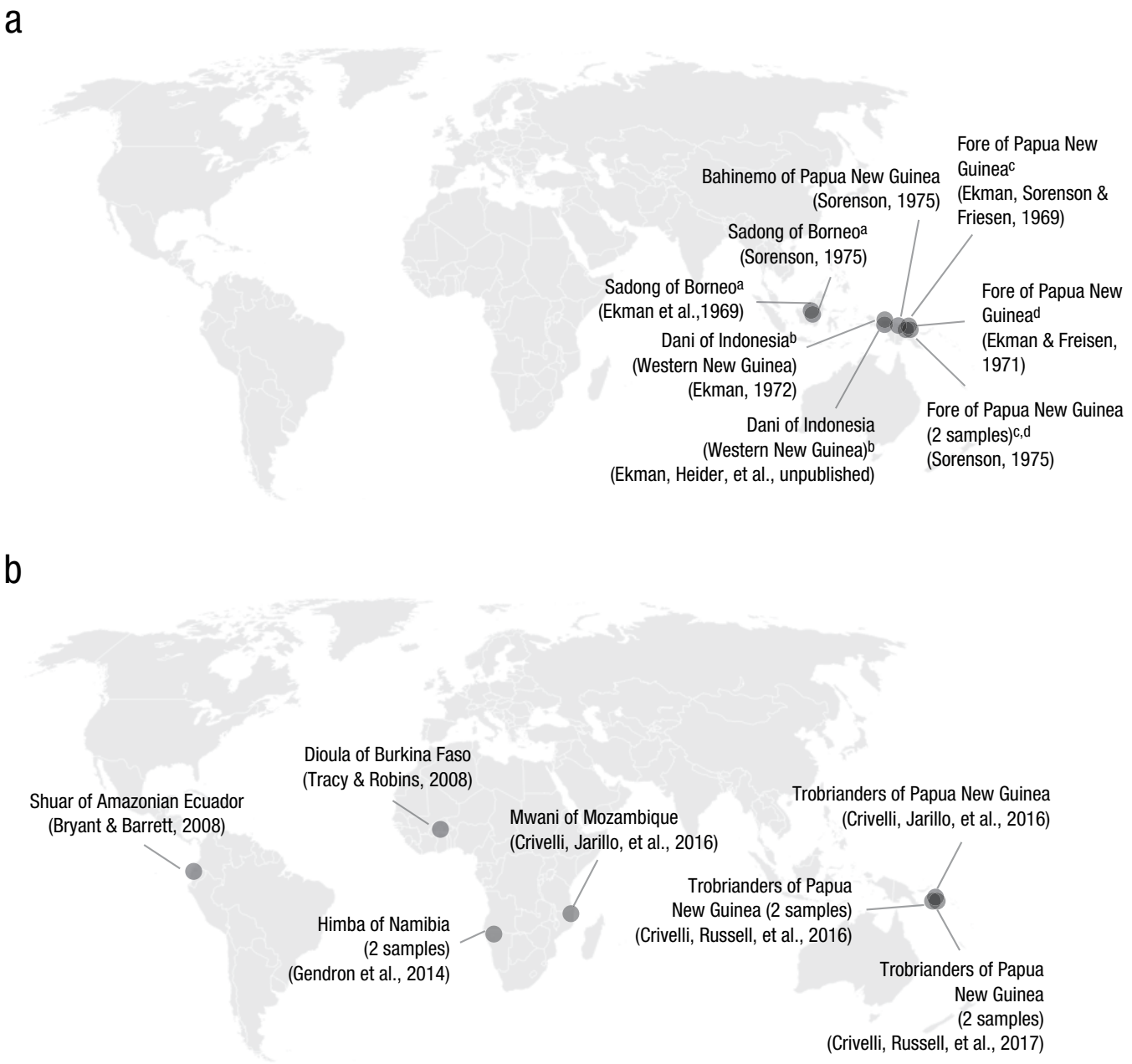


Fig. 12. Map of cross-cultural studies of emotion perception in small-scale societies. People in small-scale societies typically live in groupings of several hundred to several thousand people who maintain autonomy in social, political and economic spheres. (a) Epoch 1 studies, published between 1969 and 1975, were geographically constrained to societies in the South Pacific. Studies that share the same superscript letter may share the same samples. (b) Epoch 2 studies, published between 2008 and 2017, sample from a broader geographic range including Africa and South America and are more diverse in the ecological and social contexts of the societies tested. This type of diversity is a necessary condition for discovering the extent of cultural variation in psychological phenomena (Medin, Ojalehto, Marin, & Bang, 2017). Adapted from Gendron, Crivelli, and Barrett (2018).

such as the Himba (Gendron, Roberson, van der Vyver, & Barrett, 2014a, 2014b) and the Trobriand Islanders (Crivelli, Jarillo, et al., 2016; also see Srinivasan & Martinez, 2018, described in Box 7 in the Supplemental Material); these studies showed that perceivers can reliably infer valence but not arousal in facial configurations. In addition, Hadza participants who had some contact with people from other cultures—they had some formal schooling or could speak Swahili, which is not their native language—were more consistently able to choose

the hypothesized facial configuration than were those with no formal schooling who spoke minimal Swahili (for a similar finding with Fore participants in a free-labeling study, see Table 2 in Sorenson, 1975). Of the 27 Hadza participants who had minimal contact with other cultures, only 12 reliably chose the wide-eyed, gasping facial configuration at above chance levels to match the fear story. (Compare this finding with the observation that the hypothesized universal expression for fear—a wide-eyed, gasping facial configuration—is understood as an aggressive, threatening display by Trobriand Islanders; Crivelli, Jarillo, & Fridlund, 2016; Crivelli, Russell, Jarillo, & Fernández-Dols, 2016, 2017).

Studies that measure emotion perception with free-labeling tasks. During the period from 1969 to 1975, between one and three small-scale samples from remote cultures in the South Pacific were studied with free labeling to investigate emotion perception (three samples were reported in Sorenson, 1975; see Table 5 in the current article). From 2008 onward, two additional studies were conducted, one asking participants from the Trobriand Islands to infer emotions in photographs of spontaneous facial configurations (Crivelli et al., 2017, Study 1) and the other asking Hadza participants to infer emotions in photographs of posed facial configurations (Gendron et al., 2018, Study 2). Taken together, these five studies provide little evidence that the facial configurations in Figure 4 are universally judged to specifically express certain emotion categories. The three free-labeling studies reported in Sorenson (1975) produced variable results. The only replicable finding appears to be that participants labeled smiling facial configurations uniquely as happiness in all studies (as the only pleasant emotion category tested). The two newer free-labeling studies both indicated that participants rarely spontaneously labeled facial configurations with the expected emotion labels (or their synonyms) above chance levels. Trobriand Islanders did not label the proposed facial configurations for happiness, sadness, anger, surprise, or disgust with the expected emotion labels (or their synonyms) at above chance levels (although they did label the faces consistently with other words; Crivelli et al., 2017, Study 1). Hadza participants labeled smiling and scowling facial configurations as happiness (44%) and anger (65%), respectively, at above chance levels (Gendron, Hoemann, et al., 2018, Study 2). The word *anger* was not used to uniquely label scowling facial configurations, however, and it was frequently applied to frowning, nose-wrinkled, and gasping facial configurations.

Facial movements carry meaningful information, even if they do not reliably and specifically display emotional states. The more recent studies of people living in small-scale, remote cultures suggest two interesting and noteworthy observations. First, even though people may not routinely infer anger from scowls, sadness from frowns, and so on, they do reliably infer other social meanings for those facial configurations, because facial movements often carry important information about social motives and other psychological features (Crivelli, Jarillo, Russell, & Fernández-Dols, 2016; Crivelli et al., 2017; Rychlowska et al., 2015; Wood, Rychlowska, & Niedenthal, 2016; Yik & Russell, 1999; for a discussion, see Fridlund, 2017; J. Martin et al., 2017). For example, as we mentioned earlier, Trobriand Islanders consistently labeled wide-eyed, gasping faces (the proposed expressive facial configuration for the fear category) as signaling an intent to attack (i.e., a threat; for additional evidence in carvings and masks in a variety of cultures, including Maori, !Kung Bushmen, Himba, and Eipo, see Crivelli, Jarillo, & Fridlund, 2016; Crivelli, Jarillo, Russell, & Fernández-Dols, 2016).

Second, people do not always infer internal psychological states (emotions or otherwise) from facial movements. People who live in non-Western cultural contexts, including Himba and Hadza participants, are more likely to assume that other people's minds are not accessible to them, a phenomenon called *opacity of mind* in anthropology (Danziger, 2006; Robbins & Rumsey, 2008). Instead, facial movements are perceived as actions that predict future actions in certain situations (e.g., a wide-eyed, gasping face is labeled as "looking"; Crivelli et al., 2017; Gendron, Hoemann, et al., 2018; Gendron et al., 2014b). Similar observations were unavailable for the earlier studies conducted by Ekman, Friesen, and Sorenson because, according to Sorenson (1975), they directed participants to provide emotion terms. When participants spontaneously offered an action label (e.g., "she is just looking") or a social evaluation (e.g., "he is ugly," or "he is stupid"), they were asked to provide an "affect term." Such findings suggest that there may be profound cultural variation in the type of inferences human perceivers typically make when looking at other human faces in general, an observation that has been raised by a number of anthropologists and historians.

A note on interpreting the data. To properly interpret the scientific evidence, it is crucial to consider the constraints placed on participants by the experimental tasks that they are asked to complete, summarized in Table 4. In most urban and in some remote samples, experiments using choice-from-array tasks produce evidence supporting the common view: Participants reliably label scowling facial configurations as angry, smiling facial configurations as happy, and so on. (We do not yet know whether perceivers are uniquely labeling each facial configuration as a

specific emotion because most studies do not report that information.)

It has been known for almost a century that choice-from-array tasks help participants obtain a level of reliability in their emotion perceptions that is not routinely seen in studies using methods that allow participants to respond more freely, and this is one reason they were chosen for use in the first place (for a discussion, see Gendron & Barrett, 2009, 2017; Russell, 1994; Widen & Russell, 2013). When participants are offered words for happiness, fear, surprise, anger, sadness, and disgust to register their inferences for a scowling facial configuration, they are prevented from judging a face as expressing other emotion categories (e.g., confusion or embarrassment), nonemotional mental states (e.g., a social motive, such as rejection or avoidance), or physical events (e.g., pain, illness, or gas), thus inflating reliability rates within the task. When people are provided with other options, they routinely choose them. For example, participants label scowling faces as "determined" or "puzzled," wide-eyed faces as "hopeful," and gasping faces as "pained" when they are provided with stories about those emotions rather than with stories of anger, surprise, and fear (Carroll & Russell, 1996; also see Crivelli et al., 2017). The problem is not with the choice-from-array task per se—it is more with failing to consider alternative explanations for the observations in an experiment and therefore drawing unwarranted conclusions from the data.

Choice-from-array tasks may do more than just limit response options, making it difficult to disconfirm common beliefs. The emotion words provided during the task may actually encourage people to see anger in scowls, sadness in pouts, and so on, or to learn associations between a word (e.g., *anger*) and a facial configuration (e.g., a scowl) during the experiment (e.g., Gendron, Roberson, & Barrett, 2015; Hoemann et al., in press). The potency of words is discussed in Box 14, in the Supplemental Material.

Summary. The pattern of findings from the studies conducted with remote samples replicates and underscores the pattern observed in samples of participants from larger, more urban cultural contexts: Asking perceivers to infer an emotion by matching a facial configuration to an emotion word selected from a small array of options, or telling participants a brief story about a typical instance of an emotion category and asking them to pick a facial configuration from an array of two or three photos, generally inflates agreement rates, producing evidence that is more likely to support the hypothesis of reliable emotion perception compared with data coming from less constrained response methods, such as free labeling (see Table 3). This is particularly true for studies that include only one pleasant emotion category (i.e., happiness) so that all foils differ from the target in valence. The robust reliability and specificity for inferring happiness from smiling observed in these studies may be the result of participants classifying valence rather than classifying emotion categories per se. Studies that use less constrained tasks that are designed to more freely discover how people perceive emotion instead yield evidence that generally fails to find support for the common view. Less constrained studies suggest that perceivers infer more than one emotion category from the same facial configuration, infer the same emotion category in a variety of different configurations, and often disagree about the set of emotion categories that they infer. Cultural variation in emotion perception is consistent with the variation we observed in studies of expression production (again, see Table 3) and is even consistent with the research on face perception, which itself is determined by experience and cultural factors (Caldara, 2017).

Studies of healthy infants and children

Some scientists concur with the common view that infants can read specific instances of emotion in faces from birth (Haviland & Lelwica, 1987; Izard, Woodburn, & Finlon, 2010; Leppänen & Nelson, 2009; Walker-Andrews, 2005). However, it is difficult to ascertain whether infants and young children possess the various capacities required to perceive emotion per se: Simply detecting and discriminating facial movements is not the same as categorizing them to infer their emotional meaning. It is challenging to design well-controlled experiments that do a good job of distinguishing these two capacities. Infants are preverbal, so scientists use other measurement techniques, such as the amount of time an infant looks at a stimulus, to infer whether infants can discriminate one facial configuration from another, and ultimately, whether infants categorize those configurations as emotionally meaningful (for a brief explanation, see Box 15 in the Supplemental Material).

This "looking" approach introduces several possible confounds because of the stimuli used in the experiments: Infants and children are typically shown photographs of the proposed expressive forms (similar to those presented in Figure 4; e.g., Leppänen, Richmond, Vogel-Farley, Moulson, & Nelson, 2009; Peltola, Leppänen, Palokangas, & Hietanen, 2008). Infants are more familiar with some of these configurations than with others (e.g., most infants are more familiar with smiling faces than with scowls or frowns), and familiarity is known to influence perception (see Box 15, in

the Supplemental Material), making it difficult to know which features of a face are holding an infant's attention (familiarity or novelty) and which might be the basis of categorization in terms of emotional meaning. The configurations proposed for each emotion category also differ in their perceptual features (e.g., the proposed expressions for fear and surprise contain widened eyes, whereas the proposed expression for sadness does not), contributing more ambiguity to the interpretation of findings. For example, when an infant discriminates smiling and scowling facial configurations, it is tempting to infer that the child is discriminating expressions of anger and happiness when in fact that target of **discrimination** may be the presence or absence of teeth in a photograph (Caron, Caron, & Myers, 1985). Moreover, the facial configurations in question are usually posed as exaggerated facial movements that are not typical of the expressive variation that children actually observe in their everyday lives (Grossmann, 2010). Furthermore, unlike adults, infants may have had little or no experience with viewing photographs of anything, including heads of people with no bodies and no context.

The most important and pervasive confound in developmental studies of emotion perception is that most studies are not designed to distinguish between whether infants and children (a) discriminate facial configurations according to their emotional meaning and whether they (b) discriminate affective features (pleasant vs. unpleasant; high arousal vs. low arousal; see Box 9 in the Supplemental Material). Often, a facial configuration that is intended to depict a pleasant instance of emotion (smiling in happiness) is compared with one that is intended to depict an unpleasant instance of emotion (e.g., scowling in anger, frowning in sadness, or gasping in fear), or these configurations are compared with a neutral face at rest (e.g., Leppänen et al., 2007, 2009; Montague & Walker-Andrews, 2001). (This problem is similar to the one encountered earlier in our discussion of emotion-perception studies in adults from small-scale societies, in which perceptions of valence can be confused with perceptions of emotion categories.) For example, in one study, 16- to 18-month-olds preferred toys paired with smiling faces and avoided toys paired with scowling and gasping faces (N. G. Martin, Maza, McGrath, & Phelps, 2014); this type of study cannot distinguish whether infants are differentiating pleasant from unpleasant, approach versus avoidance, or something about a specific emotion.

Another study (Soken & Pick, 1999) reported that 7-month-olds distinguished sadness and anger when looking at faces, but only when the faces were paired with vocalizations. What is unclear is the extent to which the level of arousal or activation conveyed in the acoustic signals were most salient to infants. A recent study suggested that 10-month-old infants can differentiate between the high arousal, unpleasant scowling and nose-wrinkled facial configurations that are proposed as expressions of anger and disgust, suggesting that they can categorize these two facial configurations separately (Ruba et al., 2017). Yet the scowling and nose-wrinkled facial configurations also differed in properties besides their proposed emotional meaning: scowling faces showed no teeth, but nose-wrinkled faces were toothy, and it is well known that infants use perceptual features such as "toothiness" to categorize faces (see Caron et al., 1985). If an infant looks longer at a (pleasant) smiling facial configuration after viewing several (unpleasant) scowling faces, this does not necessarily mean that the infant has discriminated between and understands "happiness" and "anger"; the infant might have discriminated positive from negative, affective from neutral, familiar from novel, the presence of teeth from the absence, less eye sclera from more, or even different amounts of contrast in the photographs. In the future, to provide a sound basis to infer that infants are processing specific emotional meaning, experiments must be designed to rule out the possibility that infants are categorizing facial configurations into different groupings using factors other than emotion.

As a consequence of these confounds, there is still much to learn about the developmental course of emotion-perception abilities. By 3 months of age, infants can distinguish the facial features (the morphology) in the proposed expressive configurations for happiness, surprise, and anger; by 7 months, they can discriminate the features in proposed expressive configurations for fear, sadness, and interest. Left uncertain is whether, beyond just discriminating between the mere appearance of particular facial features, infants also understand the emotional meaning that is typically inferred from those features within their culture. By 7 months of age, infants can reliably infer whether someone is feeling pleasant or unpleasant when facial configurations are accompanied by sensory information from the voice (Flom & Bahrick, 2007; Walker-Andrews & Dickson, 1997). Only a handful of studies have attempted to test whether infants can infer emotional meaning in facial configurations rather than just discriminating between faces with different physical appearances, but they report conflicting results (Schwartz, Izard, & Ansul, 1985; Serrano, Iglesias, & Loeches, 1992). One promising future direction involves measuring the electrical signals (event-related potentials) in infant brains as they view the proposed

expressive configurations for anger and fear categories (e.g., Hoehl & Striano, 2008; Kobiella, Grossmann, Reid, & Striano, 2008). Both of these studies reported differential brain responses to the proposed facial configurations for anger and fear, but their findings did not replicate one another (and for certain measurements, they observed opposing effects; for a broader review, see Grossmann, 2015).

Studies that measure a child's ability to use an adult caregiver's facial movements to resolve ambiguous or threatening situations, referred to as *social referencing*, have been interpreted as evidence of emotion perception in infants. One-year-olds use social referencing to stay in close physical proximity to a caregiver who is expressing negative affect, whereas infants are more likely to approach novel objects if the caregiver expresses positive affect (Carver & Vaccaro, 2007; Moses, Baldwin, Rosicky, & Tidball, 2001; Saarni, Campos, Camras, & Witherington, 2006). Similar results emerge from the caregiver's tone of voice (Hertenstein & Campos, 2004; Mumme, Fernald, & Herrera, 1996). In fact, by 14 months of age, the positive or negative tone of a caregiver's voice influences what an infant will touch even more than will a caregiver's facial movements or the content of what the adult is actually saying (Vaish & Striano, 2004; Vaillant-Molina & Bahrick, 2012). These studies clearly suggest that infants can infer the valenced meaning of facial movements, at least when made by live (as opposed to virtual) people with whom they are familiar. But, again, these data do not help resolve what, if anything, infants infer about the emotional meaning of facial movements.

Learning to perceive emotions. Children grow up in emotionally rich social environments, making it difficult to run experiments that are capable of testing the common view of emotion perception while also taking into account the possible roles for learning and social experience. Nonetheless, several themes have emerged in the scientific literature, all of which suggest a clear role for learning and context in children's developing emotion-perception capacities.

One hypothesis that continues to be strongly supported by experiments is that children's capacity to infer emotional meaning in facial movements depends on context (the conditions surrounding the face that may convey information about a face's meaning). For example, emotion-concept learning, as a potent source of internal context, shapes emotion-perception capacity (discussed in Boxes 10 and 16 in the Supplemental Material). There are also developmental changes in how people use context to shape their emotional inferences about facial movements. Children as young as 19 months old can detect facial movements that are emotionally incongruent with a context (Walle & Campos, 2014). For example, when presented with adult facial configurations that are placed on bodies posing an emotional context (e.g., a scowling facial configuration placed on a body holding a soiled diaper), children (ages 4, 8, and 12 years) moved their eyes back and forth between faces and bodies when deciding how to label the emotional meaning of the faces, whereas adult participants directed their gaze (and overt visual attention) to the face alone, judging its emotional meaning in a way that was independent of the bodily context (Leitzke & Pollak, 2016). The youngest children were equally likely to base their labeling of the scene on face or context. The results of this experiment suggest that younger children devote greater attention to contextual information and actively cross-reference facial and contextual cues, presumably to better learn about and understand the emotional meaning those cues.[38]

Another important source of context that shapes the development of emotion perception in children involves the broader environment in which children grow. Children who grow up in neglectful or abusive environments in which their emotional interactions with caregivers are highly atypical have a different developmental trajectory than do those growing up in more consistently nurturing environments (Bick & Nelson, 2016; Pollak, 2015). Parents from these high-risk families produce unclear or context-inconsistent expressions of emotion (Shackman et al., 2010). Neglected children, who often do not receive sufficient social feedback, show delays in perceiving emotions in the ways that adults do (Camras, Perlman, Fries, & Pollak, 2006; Pollak et al., 2000), whereas children who are physically abused learn to preferentially attend to and identify facial movements that are associated with threat, such as a scowling facial configuration (Briggs-Gowan et al., 2015; Cicchetti & Curtis, 2005; da Silva Ferreira, Crippa, & de Lima Osório, 2014; Pollak, Vardi, Bechner, & Curtin, 2005; Shackman & Pollak, 2014; Shackman, Shackman, & Pollak, 2007). Abused children require less perceptual information to infer anger in a scowling configuration (Pollak & Sinha, 2002) and more reliably track the trajectory of facial muscle activations that signal threat (Pollak, Messner, Kistler, & Cohn, 2009). Children raised in physically abusive environments also more readily infer anger and threat in ambiguous facial configurations (Pollak & Kistler, 2002) and then require greater effortful control to disengage their attention from signs of threat (Pollak & Tolley-Schell, 2003) compared with children who have not been maltreated. This close attention to scowling faces with knitted eyebrows shapes how abused children understand what facial movements mean. For example, one study

found that 5-year-old abused children tended to believe that almost any kind of interpersonal situation could result in an adult becoming angry; by contrast, most nonabused children understand that anger is likely to be particular to interpersonal circumstances (Perlman, Kalish, & Pollak, 2008).

By 3 years of age, North American children not only start to show reliability in their emotion perceptions but also begin to show evidence of specificity. They understand that facial movements do not necessarily map on to emotional states, and how someone really feels can be faked or masked. Moreover, they know what facial movements are expected in a particular context and try to produce them despite their feelings. For example, the "disappointing gift" experiments developed by psychologist Pamela Cole and her colleagues demonstrate this well. In one study, preschool-age children were told they would be rewarded with a gift after they completed a task. Later, children received a beautifully wrapped package that contained a disappointing item, such as a broken pair of cheap sunglasses. When facing a smiling unfamiliar adult who had presented them with a gift, children forced themselves to smile (lip corner pull, cheek raise, and brow raise) and to thank the experimenter. Yet, although the children were smiling, they often kept their eyes focused down, slumped their shoulders, and made negative statements about the object, indicating that they did not, in fact, feel positive about the situation (Cole, 1986). Moreover, there was no difference in the behavioral responses of visually impaired children when receiving a disappointing gift (Cole, Jenkins, & Shott, 1989). Studies like this one provide a more implicit way of assessing children's knowledge about emotion perception (i.e., it illustrates the inferences that children expect others to make from their own facial movements).

It is possible that the frequency and type of facial input that people encounter influences their emotion categorizations. To test whether the statistical distribution of emotion input would influence how people construed boundaries between emotion categories, Plate, Wood, Woodard, and Pollak (2018) manipulated the frequency of this information to perceivers. Participants were asked to categorize facial morphs (from neutral to scowling) as being either "calm" or "upset." A third of participants saw more scowling faces, a third saw more neutral faces, and the others saw faces that were equally distributed across scowling and neutral. Both school-age children and adults adjusted their emotion categories based on the frequency of the input they encountered. Those exposed to more scowling faces increased their threshold for categorizing a face as upset (therefore narrowing their category of "anger"). Those exposed to more calm faces decreased their threshold for categorizing a face as angry. These data are consistent with the idea that the frequency or commonness of a facial configuration in an observer's environment influences his or her conception of an emotion (Levari et al., 2018; Oakes & Ellis, 2013), as well as the more general findings that expertise with faces influences identity perception (Beale & Keil, 1995; Jenkins, White, Monfort, & Burton, 2011; McKone, Martini & Nakayama, 2001; Viviani, Binda & Bosato, 2007; for a discussion of how familiarity is important for face perception, see Young & Burton, 2017). As a result, individual differences in emotion perception may be influenced by early experience that differs according to emotional input, reflecting the malleability of these categories.

Summary. There is currently no clear evidence to support the hypothesis that infants and young children reliably and specifically infer emotion in the proposed expressive configurations for the anger, disgust, fear, happiness, sadness, and surprise categories (findings summarized in Table 3). A more plausible interpretation of the existing evidence is that young infants infer affective meaning, such as valence and arousal, from facial configurations. Data from infants and young children obtained using a variety of methods further suggest that emotion-perception abilities emerge and are shaped through learning in a social environment. These findings are consistent with the idea that the human face plays an important and privileged role to communicate importance or salience. But it is not clear that the expressive configurations proposed for specific emotion categories are similarly privileged in this way.

Summary of scientific evidence on the perception of emotion in faces

The scientific findings on perception studies generally replicate those from production studies in failing to strongly support the common view. The one exception to this overall pattern of findings is seen in studies that ask participants to match a posed face to an emotion word or brief scenario. This method produces evidence that can support the common view, even when it is applied to completely novel emotion categories with made-up expressive cues (Hoemann et al., 2018), opening up interesting questions about the psychological potency of the elements that make up choice-from-array designs (e.g., the emotion words embedded in the task or the choice of foils on a given trial). These findings reinforce our earlier conclusion that such terms as "facial configuration" or "pattern of facial movements" or even "facial actions" are preferred to more

loaded terms such as "emotional facial expression," "emotional expression," or "emotional display," which can be misleading at best and incorrect at worst.

Summary and Recommendations

Evaluation of the empirical evidence

The common view that humans around the world reliably express and recognize certain emotions in specific configurations of facial movements continues to echo within the science of emotion, even as scientists increasingly acknowledge that anger, sadness, happiness, and other emotion categories are more variable in their facial expressions. This entrenched common view does more than guide the practice of science. It influences public understanding of emotion and hence education, clinical practice, and applications in industry. Indeed, it reaches into almost every facet of modern life, including emoticons and movies. However, there is insufficient evidence to support it. People do express instances of anger, disgust, fear, happiness, sadness, and surprise with the hypothesized facial configurations presented in Figure 4 at above chance levels, suggesting that those facial configurations sometimes serve as expressions of emotion as proposed. However, the reliability of this finding is weak, and there is evidence that the strength of support for the common view varies systematically with the research methods used. The strongest support for the common view—found in data from urban, industrialized, or developed samples completing choice-from-array tasks—does not show robust generalizability. Evidence for specificity is lacking in almost all research domains. A summary of the scientific evidence is presented in Table 3.

These research findings do not imply that people move their faces randomly or that the configurations in Figure 4 have no psychological meaning. Instead, they reveal that the facial configurations in question are not "fingerprints" or diagnostic displays that reliably and specifically signal particular emotional states regardless of context, person, and culture. It is not possible to confidently infer happiness from a smile, anger from a scowl, or sadness from a frown, as much of current technology tries to do when applying what are mistakenly believed to be the scientific facts.

Instead, the available evidence from different populations and research domains—infants and children, adults living in industrialized countries and in remote cultures, and even individuals who are congenitally blind—overwhelmingly points to a different conclusion: When facial movements do express emotional states, they are considerably more variable and dependent on context than the common view allows. There appear to be many-to-many mappings between facial configurations and emotion categories (e.g., anger is expressed with a broader range of facial movements than just a scowl, and scowls express more than anger). A scowling facial configuration may be *an* expression of anger in the sense of being a part of anger in a given instance. But a scowling facial configuration is not *the* expression of anger in any generalizable or universal way (there appear to be no prototypical facial expressions of emotions). Scowling facial configurations and the others in Figure 4 belong to a much larger repertoire of facial movements that express more than one emotion category, and also nonemotional psychological meanings, in a way that is tailored to specific situations and cultural contexts. The face is a powerful tool for social communication (Jack & Schyns, 2017). Facial movements, like reflexive and voluntary motor movements (L. F. Barrett & Finlay, 2018), are strongly context-dependent. Recent evidence suggests that people's categories for emotions are flexible and responsive to the types and frequencies of facial movements to which they are exposed in their environments (Plate, Wood, Woodard, & Pollak, 2018).

The degree of variation suggested by the published evidence goes well beyond the hypothesis that the facial configurations in Figure 4 are prototypes or *typical* expressions and that any observed variations are merely the result of cultural accents, display rules, suppression or other regulatory strategies, differences in induction methods, measurement error, or stochastic noise (as proposed by various scientists, including Ekman & Cordaro, 2011; Elfenbein, 2013, 2017; Levenson, 2011; Matsumoto, 1990; Roseman, 2001; Tracy & Randles, 2011). Instead, the facial configurations in Figure 4 are best thought of as Western gestures, symbols or stereotypes that fail to capture the rich variety with which people spontaneously move their faces to express emotions in everyday life. A stereotype is not a prototype. The distinction is an important one, because a prototype is the most frequent or typical instance of a category (Murphy, 2002), whereas a stereotype is an oversimplified belief that is taken as generally more applicable than it actually is.

The conclusion that emotional expressions are more variable and context-dependent than commonly assumed is also mirrored by the evidence from physiological changes (e.g., heart-rate and skin-conductance measures; see Box 8 in the Supplemental Material) and even in evidence on the brain basis of human emotion (Clark-Polner et al., 2017). The task of science is to systematically document these context-dependent patterns, as well as to understand the mechanisms that cause them, so that we can explain and predict them. Clearly, the face is a rich source of information that

plays a crucial role in guiding social interaction. Facial movements, when measured in a high-dimensional dynamic context (i.e., in a deeply multivariate way, sampling across many measurement domains within an emoter and the spatiotemporal context), may serve the diagnostic purpose that many consumers of emotion science are looking for (in which context can be a cultural context, a specific situation, a person's learning history or momentary physiological state, or even the temporal context of what just took place a moment ago; L. F. Barrett, 2017b; L. F. Barrett, Mesquita, & Gendron, 2011; Gendron, Mesquita, & Barrett, 2013).

A note on the scientific literature

Our review identified several broad problems that lurk within the scientific research on facial expressions and that may cause considerable misunderstanding and confusion for consumers of this research. First, statistical standards are commonly adopted that do not translate well for applying emotion research to other domains, applied or scientific. Showing that people frown when sad or scowl when angry with greater statistical reliability than would be expected by chance may be a scientific finding that warrants publication in a peer-reviewed journal, but above-chance responding is often low in absolute terms, making broad conclusions impossible, particularly for translation to domains of life in which a person's outcomes can be influenced by the emotional meaning that perceivers infer. Making inferences on the basis of statistical reliability without properly accounting for actual effect sizes, specificity, and generalizability, is similarly problematic.

Second, even studies that surmount these common shortcomings often have a mismatch between what is claimed in their conclusions (or what others claim in reviews or citations of those primary research studies) and what inferences can, in fact, be reasonably supported by the results. This is particularly problematic, because the perpetuation of the common view, and its applications, may be the result of superficial readings of abstracts or secondary sources rather than in-depth evaluation of the primary research.

Third, the mismatch between observations and interpretations often results from problems in how studies are designed—the particular stimuli used, the tasks used, and the statistical analyses are critically important and constrain what can be observed and inferred in the first place. Unfortunately, the published research on emotional expressions and emotion perception is rarely designed to systematically assess the degree of expressive variation. Furthermore, this research often confounds the measurements made in an experiment with the interpretation of those measurements, referring to facial movements as "emotional displays," "emotional expressions," or even "facial expressions" rather than "facial configurations," "facial movements," or "facial actions"; referring to people "detecting" or "recognizing" emotion rather than "perceiving" or "inferring" an emotional state on the basis of some set of cues (facial movements, vocal acoustics, body posture, etc.); and referring to "accuracy" rather than "agreement" or "consensus."

A note on other emotion categories

Our conclusions most directly challenge what we have termed the "common view": that a scowling facial configuration is *the* expression of anger, a nose-wrinkled facial configuration is *the* expression of disgust, a gasping facial configuration is *the* expression of fear, a smiling facial configuration is *the* expression of happiness, a frowning facial configuration is *the* expression of sadness, and that a startled facial configuration is *the* expression of surprise (see Fig. 4). By necessity, we focused our review of evidence on these 6 emotion categories, rather than the more than 20 emotion categories that are currently being studied, because studies on these 6 are far more numerous than studies of other emotion categories. Nonetheless, some scientists claim that each of these other emotion categories has a prototypical, universal expression, facial or otherwise, that is modified or accented by culture (e.g., Cordaro et al., 2018; Keltner et al., 2019). In our view, such claims rest on evidence that is subject to the same critique that we offered for the research reviewed in detail here. In short, even though our review focused on the six emotion categories that are sometimes referred to as "basic emotions," our observations and conclusions generalize to studies of other emotion categories that use similar methods.

Recommendations for consumers of emotion research on applying the scientific findings

Presently, many consumers of emotion research assume that certain questions about emotional expressions have been answered satisfactorily when in fact this is not the case. Technology companies, for example, are spending millions of research dollars to build devices to read emotions from faces, erroneously taking the common view as a fact that has strong scientific support. A more accurate description, however, is that such technology detects facial movements, not emotional expressions.[39] Corporations such as Amazon are exploring virtual-human technology to interface with consumers. Virtual humans are used to educate children, train physicians, and train the military as well as infer psychological disorders, and perhaps will eventually even be used to offer treatments for psychiatric illnesses. At the moment,

Table 6. Recommendations for Reading Scientific Studies About Emotion

1. Take note of whether an experiment is studying expressive stereotypes or more variable facial movements.
2. Take note of data on specificity and generalizability; do not focus solely on reliability at above chance levels.
3. Make a distinction between the data in an experiment (what was measured) and how those data are interpreted.
4. Translate "emotional expressions" or "emotional displays" into "facial movements."
5. Translate "emotion recognition" into "emotion perception" or "emotion inference."
6. Translate "accuracy" to "agreement," "consensus," or "reliability."
7. Give more weight to studies that measure facial movements or study the perception of facial movements in more naturalistic settings.
8. Take note of studies that measure or manipulate context.
9. Field studies of people from small-scale, remote cultures are often less well-controlled than studies conducted in the laboratory, but they are invaluable in the information that they provide and should be appreciated.
10. Remember that emotions are not understood as internal states in all cultures. In some cultures they are understood as situated actions.
11. Do not skip the Method and Results sections and go directly to the Discussion to learn the results of an experiment. It is important to know what was measured and observed, not just how scientists interpreted their measurements.

the science of emotion is ill-equipped to support any of these initiatives. So-called emotional expressions are more variable and context-dependent than originally assumed, and most of the published research was not designed to probe this variation and characterize this context dependence. As a consequence, as of right now, the scientific evidence offers less actionable guidance to consumers than is commonly assumed.

In fact, our review of the scientific evidence indicates that very little is known about how and why certain facial movements express instances of emotion, particularly at a level of detail sufficient for such conclusions to be used in important, real-world applications. To help consumers navigate the science of emotion, we offer some tips for how to read experiments and other scientific articles (Table 6).

More generally, tech companies may well be asking a question that is fundamentally wrong. Efforts to simply "read out" people's internal states from an analysis of their facial movements alone, without considering various aspects of context, are at best incomplete and at worst entirely lack validity, no matter how sophisticated the computational algorithms. These technology developments are powerful tools to investigate the expression and perception of emotions, as we discuss below. Right now, however, it is premature to use this technology to reach conclusions about what people feel on the basis of their facial movements—which brings us to recommendations for future research.

Recommendations for future scientific research

Specific, concrete recommendations for future research to capitalize on the opportunity offered by current challenges can be found in Table 7, but we highlight a few general points here. First, the expressive stereotypes that summarize the common view, such as those depicted in Figure 4, are ubiquitous in published research. It is time to move beyond a science of stereotypes to develop a science of how people actually move their faces to express emotion in real life, and the processes by which those movements carry information about emotion to someone else (a perceiver). (For a discussion of information theory as applied to emotional communication, see Box 16 in the Supplemental Material). The stereotypes of Figure 4 must be replaced by a thriving scientific effort to observe and describe the lexicon of context-sensitive ways in which people move their facial muscles to express emotion, and the discovery of when and how people infer emotions in other people's facial movements.

New research on emotion should consider sampling individuals deeply, with high dimensional measurements, across many different situations, times of day, and so forth: a Big Data approach to learning the expressive repertoires of individual people. The diagnosis of an instance of emotion might be improved by combining many features, even those that are weakly diagnostic on their own, particularly if the analysis is conducted in a person-specific (idiographic) way (e.g., Rudovic, Lee, Dai, Schuller & Picard, 2018; Yin, et al., 2018). In the ideal case, videos of people in natural situations could be quantified by automated algorithms for various physical features, such as facial movements, posture, gait, and tone of voice. To this, scientists could add the sampling of other physical features, such as ambulatory monitoring of ANS changes to sample the internal milieu of people's bodies as they dynamically change over time, ambulatory eye-tracking to assess gaze and attention, ambulatory brain imaging (e.g., electroencephalography), and optical brain imaging (e.g., functional near-infrared spectroscopy). Only a highly multivariate set of measures is

likely to work to classify instances of emotion with high reliability and specificity.

The failure to find reliable "fingerprints" for emotion categories, including the lack of reliable facial movements to express these categories, may stem, at least in part, from the same source: Scientific approaches have ignored *substantial*, meaningful variability attributable to context (for recent computer innovations, see Kosti, Alvarez, Recasens & Lapedriza, 2017). There is Bluetooth technology to capture the physical spaces people inhabit (which can be quantified for various structural and social descriptive features, such as the extent of their exposure to light and noise), whether they are with another person, how that person reacts, and so on. In principle, rich, multimodal observations could be available from videos; when time-synchronized with the other physical measurements, such video could be extremely useful in understanding the conditions when certain facial movements are made and what those movements might mean in a given context. Naturally, Big Data in the absence of hypotheses is not necessarily helpful.

Participants could be offered the opportunity to annotate their videos with subjective ratings of the features that describe their experiences (whether or not they are identified as emotions). Candidate features are affective properties such as valence and arousal (see Box 9 in the Supplemental Material), **appraisals** (i.e., descriptions of how a situation is experienced; e.g., Clore & Ortony, 2013; see L. F. Barrett, Mesquita, Ochsner, & Gross, 2007; Gross & Barrett, 2011), and emotion-related goals. These additional psychological features have the potential to add higher dimensional details to more specifically characterize facial movements and what they mean.[40] Such an approach introduces various technical and modeling challenges, but this sort of deeply inductive approach is now within reach.

Another opportunity for high dimensional sampling of emotional events involves interactions with virtual humans. Because virtual humans can realize contingent behavior in rich social interactions under strict and precise experimental control, they can provide a richer, more natural context in which to study emotional expressions and emotion perception than may be true for traditional laboratory studies. In addition, they do not suffer from the loss of experimental control that limits causal inferences from ethological studies.

To date, this potential has not yet been exploited to explore the reliability and specificity in context-sensitive relations between facial movements and mental events. As we noted earlier, most of the virtual systems are now designed to teach people a variety of skills, where the goal is not to assess how well participants perceive emotions in facial movements under realistic, socially ambiguous conditions, but instead to program expressive behaviors into virtual humans that will motivate people to learn the needed skills. In these experiments, the psychological realism of facial movements is often secondary to the primary goals of the experiment. A scientist might even program a virtual human with behavior or appearance that is unnatural or infeasible for a human (i.e., that are supernormal) so that a participant can unambiguously interpret and be influenced by the agent's actions (for relevant discussion, see D. Barrett, 2010; Tinbergen, 1953).

Nonetheless, the scientific approach of observing people as they interact with artificial humans holds great promise for understanding the dynamics and mechanisms of emotion perception and may get us closer to understanding human emotion perception in everyday life. Virtual humans are vivid. Unlike more passive approaches to evoking emotion, such as viewing videos or images of facial configurations, a virtual human engages a human participant in a direct, social interaction to elicit perceptual judgments that are either directly reported or inferred from behaviors measured in the participant. Virtual humans are also highly controllable, allowing for precise experimentation (Blascovich et al., 2002). A virtual human's facial movements and other details can be repeated across participants, offering the potential for robust and replicable observations. Numerous studies have demonstrated that humans are influenced by them (e.g., Baylor & Kim, 2008; Krumhuber et al., 2007; McCall, Blascovich, Young, & Persky, 2009). For example, human learners are more engaged by virtual agents who move their faces (and modulate their voices), leading the learners to an increased sense of self-efficacy (Y. Kim, Baylor, & Shen, 2007). As a consequence, virtual humans potentially allow for the study of emotion in a rich virtual ecology, a form of synthetic in vivo experimentation (Marsella & Gratch, 2016).

When combined with the high dimensional sampling we described earlier, there is the potential to revolutionize our understanding of emotional expressions by asking questions that are different from those encouraged by the common view. Automated algorithms using data captured from videos offer substantial improvements with a data-driven, unsupervised approach. The result could be robust descriptions about the context-sensitive nature of emotional expressions that is currently missing, and that would set the stage for a more mechanistic, causal account of emotions and their expressions.

An ethology of emotions and their expressions can also be pursued in the lab. Experiments can go beyond a study of how people move their faces in a single situation chosen to be most typical of a given emotion category. Most studies to date have been designed to observe facial movements in only the most stereotypic

Table 7. Recommendations for Future Research

General Recommendations

- Take chances on studies that attempt to go beyond merely supporting the common view of emotion.
- Support papers that attempt to study facial movements in real life, measuring context, sampling across cultures (even though these studies are often less well controlled than studies in the laboratory), or using facial stimuli that are less familiar to reviewers than common-view stimulus sets.
- Prioritize multidisciplinary studies that combine classical psychology methods with cognitive neuroscience, machine learning, and so forth.
- Support larger scale studies that bridge the lab and the world, that study individual people across many contexts, and that measure emotional episodes in high dimensional detail, including physical, psychological, and social features.
- Support the development of computational approaches.
- Create teams that pair psychologists and cognitive scientists trained in the psychology of emotion with engineers and computer scientists.
- Increase opportunities to test innovative methods and novel hypotheses, with the acknowledgment that such approaches are likely to elicit resistance from established scientists in the field of emotion.
- Generate more studies to identify the underlying neural mechanisms of the production and perception of facial movements.
- Direct funding to thornier but necessary new questions and be critical of projects that perpetuate past errors in emotion research.
- Direct healthy skepticism to tests, measures, and interventions that rely on assumptions about "reading facial expressions of emotion" that seem to ignore published evidence and/or ignore integration of contextual information along with facial cues.
- Develop systematic, precise ways to describe and/or manipulate the dynamics of specific facial actions.

Problem	Recommendation
	Stimulus selection
Limitations in stimulus selection can bias results.	• For production studies, ensure that multiple stimuli per emotion category are used to evoke an emotion. • Measure emotional episodes in a multimodal way and attempt to discover explicit criteria for when an instance of emotion is present or absent. Such discovery may require within-person approaches. Consider quantifying the presence or absence probabilistically. • For perception studies, incorporate images from the wild (e.g., from multiple Internet sources) to capture the full range of facial movements that humans produce in their everyday lives. • For both production studies (in which stimuli are designated to evoke emotion) and perception studies, build variation into stimulus sets so conclusions about emotion categories are more generalizable to the real world. Consider randomly sampling a variety of stimuli for a given category and treating stimuli as a random variable.
Little is known about the dynamics of emotional expression and emotion inferences.	• For production studies, code the temporal dynamics of facial movements. • Attempt to determine the apex of facial movements, changes to AUs as movements emerge and recede, and whether the kinematics of distinct AUs are similar or different across sequences or phases of facial movements. • Ensure sufficient temporal resolution to allow for event segmentation to be assessed in perception studies. • For perception studies, use dynamic images rather than rely on still images. • Quantify, as best as possible, participants' degree of exposure to and knowledge of Western cultural practices and norms, as well as the amount and type of formal schooling made available to participants.
The role of context is hotly debated but rarely measured in sufficient detail.	• Measure or manipulate the context in which facial movements occur. • Manipulate (or at least measure) the context in which facial movements are perceived to evaluate whether data are truly stimulus-specific or influenced by the entire scene (i.e., present facial movements in their spatiotemporal context). • Explicitly acknowledge that the experimental method is itself a psychologically potent context that is likely to influence responses.
	Sample selection
Cross-cultural studies can provide powerful insights but are limited in number and scope.	• Harness technology to collect larger numbers of images and video sequences of facial movements across cultures. • Remember that emotions are understood as situated actions and not as mental events or feelings in many cultures; also, mental inferences may be understood differently in different cultures.

(continued)

Table 7. (Continued)

	Task, method, and design
Measurement and interpretation of emotion are often blurred in research studies.	• Contrast more than one emotion category with a baseline, so that conclusions about a specific emotion category are not drawn from a comparison of an emotion condition and a no-emotion condition. • Compare multiple emotion categories to no-emotion categories in a given study. • Unless a study design is completely data-driven, explicitly state the theoretical priors of the research team. Consider stating whether a study is designed to discover new knowledge or test existing hypotheses about the traditional English categories. Both approaches are valid but should be clearly articulated. • Explicitly hypothesize how task design and method might influence responses.
New insights about emotion are constrained by reliance on, and assumptions about, traditional English emotion categories.	• Sample a broader number of emotion categories those used in prior research (move beyond the 20 or so English categories that are now becoming common in research). Consider sampling non-English emotion categories. Test for variations within these categories and similarity across categories.
	Data analysis and interpretation
Findings are limited by a failure to consider issues related to forward and reverse inference.	• Test both reliability and specificity when presenting results about facial movements (expression production) and emotion inferences (emotion perception). • Use formal signal detection analytics and information theory measures rather relying on frequency or levels of agreement. • Consider using Bayesian methods so that the null hypothesis can be tested directly. • Use unlabeled classification approaches to discover emotion categories and their expressive forms, rather than continuing to ask whether other cultures express and perceive emotions in a manner that is similar to that of people in the United States.

Note: AU = action unit.

situations. Future studies should examine emotional expression and perception across a range of situations that vary systematically in their physical, psychological, and social features. Furthermore, scientists should aim to understand the various ways that humans acquire the skills to express and perceive emotion, as well as the conditions that can impair the development of these processes.

The shift toward more context-sensitive scientific studies of emotion has already begun (see Box 3 in the Supplemental Material), but it currently falls short of what we are recommending. Nonscientists (and some scientists) still anchor on the common view and only slowly shift away from it (Tversky & Kahneman, 1974; T. D. Wilson, Houston, Etling, & Brekke, 1996). The pervasiveness of the common view supports strong convictions about what faces signal, and people often continue to hold to those convictions even when they are demonstrably wrong (L. F. Barrett, 2017b; Todorov, 2017). Such convictions reflect cultural beliefs and stereotypes, however. This state of affairs is not unique to the science of emotional expression or to the science of emotion more generally (Kuhn, 1962).

In our view, the scientific path forward begins with the explicit acknowledgment that we know much less about emotional expressions and emotion perception than we thought we did, providing an opportunity to cultivate the spirit of discovery with renewed vigor and take scientific discovery in a new direction (Firestein, 2012). With this context of discovery comes the sobering realization that those of us who cultivate the science of emotion and the consumers who use this research should seriously question the assumptions of the common view and step back from what we think we know about reading emotions in faces. Understanding how best to infer someone's emotional state or predict someone's future actions from their facial movements awaits the outcomes of future research.

Appendix A: Glossary

Accuracy/accurate: The extent to which a participant's performance corresponds to what is hypothesized in a given experimental task. Critically, this requires that the hypothesized performance can be measured in a perceiver-independent way that is not subject to the inferences of the experimenter.

Affect: A general property of experience that has at least two features: pleasantness or unpleasantness (valence) and degree of arousal. Affect is part of every waking moment of life and is not specific to instances of emotion, although all emotional experiences have affect at their core.

Agreement: The extent to which two people provide consistent responses; high agreement produces high intersubject **consistency**. Percentage agreement is not the same as percentage accuracy, because the former is more **perceiver-dependent** than the latter.

Appraisal: A psychological feature of experience (e.g., a situation is experienced as novel). Some scientists use the

word *appraisal* to additionally refer to a literal cognitive mechanism that causes a feature of experience (e.g., an evaluation or judgment of whether a situation is novel).

Approach/avoidance: A fundamental dimension of motivated behavior. It is different from valence, which is a dimension of experience rather than of behavior.

Category/categorization: The psychological grouping of objects, people, or events that are perceived to be similar in some way. Categorization may occur consciously or unconsciously. May be explicit (as when applying a verbal label to instances of the grouping) or implicit (treating instances the same way or behaving toward them in the same way).

Choice-from-array tasks: Any judgment task that asks research participants to pick a correct answer from a small selection of options provided by the experimenter. For example, in the study of emotion perception, participants are often shown a posed facial configuration depicting an emotional expression (e.g., a scowl), along with a small selection of emotion words (e.g., "angry," "sad," "happy") and asked to pick the word that best describes the face.

Common view: In this article, the most predominant view about how emotions are related to facial movements. Although it is difficult to quantify, we characterize the common view through examples (e.g., an Internet Google search—see Box 1 in the Supplemental Material). The common view holds that (a) certain emotion categories reliably cause specific patterns of facial muscle movements, and (b) specific configurations of facial muscle movements are diagnostic of certain emotions categories. See Figure 4.

Conditional probability: The probability that an event X will occur given that another event Y has already occurred, or $p(X|Y)$. If X is a frown and Y is sadness, then $p(\text{frown}|\text{sadness})$ is the conditional probability that a person will frown when sad. See also **consistency**, **forward inference**, and **reverse inference.**

Configuration of facial-muscle movements/facial configuration: A pattern of visible contractions of multiple muscles in the face. Configurations can be described with FACS coding. Not synonymous with **facial expression**, which requires an inference about the causes or meaning of the facial configurations.

Confirmatory bias: The tendency to search for, remember, or believe evidence that is consistent with one's existing beliefs or hypotheses rather than ramain open to evidence inconsistent with one's priors.

Congenitally blind: People who are born without vision. The use of this term in the literature is considerably heterogeneous. Some people are truly blind from the moment they are born, but others have severe visual impairments short of complete blindness or they become blind in infancy. If the cause is peripheral (in the eyes rather than the brain), such individuals may still be able to think and imagine very similarly to sighted individuals.

Consistency: An outcome that does not vary greatly across time, context, and different individuals. Consistency is not accuracy (e.g., a group of people can consistently believe something that is wrong). Also referred to as **reliability**.

Discrimination: In psychophysics, the action of judging that two stimuli are different from one another. This is separate from pinpointing what they are (identification) or what they mean (recognition).

Emotional episode: A window of time during which an **emotional instance** unfolds. Often, but not always, accompanied by an experience of emotion, and sometimes, but not always, involves an **emotional expression.**

Emotional expression: A facial configuration, bodily movement, or vocal expression that reliability and specifically communicates an emotional state. Many perceived emotional expressions are in fact errors of **reverse inference** on the part of perceivers (e.g., an actor crying when not sad).

Emotional granularity: Experiencing or perceiving emotions according to many different **categories.** For instance, low emotional granularity involves understanding terms such as *angry, sad*, and *afraid* as synonyms of *unpleasant*; high emotional granularity involves understanding terms such as *frustrated, irritated*, and *enraged* as distinct from each other and from *angry*.

Emotional instance/instance of emotion: An event categorized as an emotion. For example, an instance of anger is the categorization of an **emotional episode** of anger. In cognitive science, an instance is called a *token* and the category is called a *type*. So, an instance of anger is a token of the category anger. (See **emotional episode**.)

Facial action coding system (FACS): A system to describe and quantify visible human facial movements.

Facial expression: A **facial configuration** that is inferred to express an internal state.

Facial movement: A **facial configuration** that is objectively described in a **perceiver-independent** way. This description is agnostic about whether the movement expresses an emotion and does not use **reverse inference**. **FACS** coding is used to describe facial movements.

Forward inference: Inferring an effect from knowing its cause. An example would be the **conditional probability** of observing a frown if we know somebody is angry, $p(\text{frown}|\text{anger})$.

Free-labeling task: An experimental task that is not a forced choice, but in which the participants generate words or other responses of their choosing.

Generalize/generalizability: The replication of research findings across different settings, samples, or methods. Generalizability can be weak (when a finding can be replicated to a limited extent) or strong (when it can be replicated across very different methods and cultures).

Mental inference/mentalizing: Assigning a mental cause to actions; also sometimes referred to as *theory of mind*. The reverse inference of inferring emotions from seeing facial movements can be an example of mentalizing.

Meta-analysis: A method for statistically combining findings from many studies.

Multimodal: Combining information from more than one of the senses (e.g., vision and audition).

Null hypothesis: The hypothesis or default position that there is no relationship between a set of variables. Equivalent to observing effects that would occur by chance (i.e., what would obtain if observations are random or permuted). Consequently, if the null hypothesis is true, the distribution of p values is uniform (every possible outcome has an equal chance).

Perceiver-dependent: An observation that depends on human judgment. Perceiver dependency can produce conclusions that are consistent across people but **consistency** does not assure accuracy or validity.

Perceiver-independent: An observation that does not depend on human judgment (although its interpretation will depend on human inference). Some philosophers argue that all observations require human judgment, there are degrees of dependency. Judging whether a flower vase is rectangular or oval is relatively perceiver-independent, whereas judging whether it looks nice is perceiver-dependent.

Perceptual matching task: An experimental task that requires research participants to judge two stimuli, such as two facial configurations, as similar or different. This requires only **discrimination**, not **categorization**, recognition, or naming.

Prototype: The most frequent or most typical instance of a category. Distinct from **stereotype**: A group of people may have a **perceiver-dependent** stereotype that is an inaccurate representation of the prototype.

Recognize/recognition: Acknowledging something's existence (which is confirmed to exist by **perceiver-independent** means). Contrasted with perception (which involves inference and interpretation).

Reliable/reliability: An observation that is repeatable across time, context, and individuals. See **consistency**.

Replicable: The extent to which new experiments come to the same conclusions as a previous study. Strong replications **generalize** well: Similar conclusions are obtained even when the new experiments use different subject samples, stimuli, or contexts.

Reverse correlation: A psychophysical, data-driven technique for deriving a representation of something (e.g., an image of a facial configuration) by averaging across a large number of judgments.

Reverse inference: Inferring a cause from having observed its purported effect. For instance, inferring that a scowl means someone is angry—the conditional probability, $p(\text{anger}|\text{frown})$. In general, reverse inference is poorly constrained because multiple causes are usually compatible with any observation.

Sensory modalities: The different senses: vision, hearing, etc.

Specific/specificity: Research conclusions that include positive as well as negative statements. For instance, concluding that a scowl signals anger and no emotion categories other than anger. High specificity is required for valid **reverse inference**.

Stereotype: A widely held belief about a category that is generally believed to be more applicable than it actually is.

Universal: Something that is common or shared by all humans. The source of this commonality (innate or learned) is a separate issue. If an effect is universal, it generalizes across cultures.

Validity: Whether an observed variable actually measures what is claimed—for example, whether a facial movement reliably expresses an emotion (convergent validity) and specifically that emotion (discriminative validity)—where the presence of the emotional instance can be verified by objective means.

Acknowledgments

We thank Jose-Miguel Fernández-Dols and James Russell for providing a copy of their edited volume before its publication, and we additionally thank Jose-Miguel and Juan Durán for providing comments on our description of their meta-analytic work. We are grateful to Jennifer Fugate for her assistance with constructing Figure 4 and, in particular, for her FACS coding of the facial configurations presented in Figure 4. Many thanks also to Linda Camras, Vanessa Castro, and Carlos Crivelli for providing additional details of their published experiments. Thanks to Jeff Cohn for his guidance on acceptable interrater reliability levels for FACS coding. This article benefited from discussions with Linda Camras, Pamela Cole, Maria Gendron, Alan Fridlund, and Tuan Le Mau. We are also deeply grateful to those friends and colleagues who invested their time and efforts in commenting on an early draft of the manuscript (though the summaries of published works and conclusions drawn from them reflect only the views of the authors): Linda Camras, Maria Gendron, Rachael Jack, Judy Hall, Ajay Satpute, and Batja Mesquita. The views, opinions, and/or findings contained in this article are those of the authors and shall not be construed as an official U.S. Department of the Army position, policy, or decision, unless so designated by other documents.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by U.S. Army Research Institute for the Behavioral and Social Sciences Grant W911NF-16-1-019 (to L. F. Barrett); National Cancer Institute Grant U01-CA193632 (to L. F. Barrett); National Institute of Mental Health Grants R01-MH113234 and R01-MH109464 (to L. F. Barrett), 2P50-MH094258 (to R. Adolphs), and R01-MH61285 (to S. D. Pollak); National Science Foundation Civil, Mechanical and Manufacturing Innovation Grant 1638234 (to L. F. Barrett and S. Marsella); National Institute on Deafness and Other Communication Disorders Grant R01-DC014498 (to A. M. Martinez); National Eye Institute Grant R01-EY020834 (to A. M. Martinez); Human Frontier Science Program Grant RGP0036/2016 (to A. M. Martinez); *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Grant U54-HD090256 (to S. D. Pollak); a James McKeen Cattell Fund Fellowship (to S. D. Pollak); and Air Force Office of Scientific Research Grant FA9550-14-1-0364 (to S. Marsella).

Notes

1. Decades of research in social psychology show that humans automatically try to predict other people's behavior by inferring a mental state—this is called *mental-state inference* or *mentalizing*, such as when inferring someone's emotional state (e.g., for a review, see Gilbert, 1998). This research suggests that inference and prediction are not separate steps (E. R. Smith & DeCoster, 2000).
2. Words in boldface type appear in the glossary in Appendix A.
3. To be clear, teaching children how to infer emotions in others is not a problem because this skill is related to efficient communication with others. The question is whether children are being taught information that is scientifically valid and generalizable.
4. A website for the Detego Group (2018) indicated that "The methods developped [*sic*] by Paul Ekman are based on 40 years of research and are being taught to the FBI, CIA, Scotland Yard and more forensics specialists around the world."
5. This empirical emphasis is largely consistent with scientists' explicit reports of what they believe, according to a recent survey (Ekman, 2016). Two-hundred forty-eight scientists who published peer-reviewed articles on the topic of emotion were asked about their views on what the scientific evidence shows. Of the 149 (60%) who responded, 119 (80%) indicated that they believed compelling evidence exists for the hypothesis that certain emotion categories are expressed with universal facial configurations or vocal signals; no questions about variability were included in the survey.
6. In social psychology, this is the distinction between identifying an action and making an inference about the mental cause of the action (Gilbert, 1998; Vallacher & Wegner, 1987).
7. This corresponds to the null hypothesis for the true positive (in Fig. 3).
8. To test the specificity hypothesis, we test something called the false positive: that people frequently scowl when not angry, meaning that they scowl more frequently than chance would allow when fearful, sad, confused, hungry, and so on (see Fig. 3). Retaining the null hypothesis for the false positive (i.e., that people do not scowl more frequently than they would by chance when fearful, sad, confused, hungry, etc.) is equivalent to rejecting the null hypothesis for (i.e., finding support for) the specificity hypothesis. Rejecting the null hypothesis for the false positive (because people scowl when they are fearful, sad, confused, hungry, etc., in addition to when they are angry) is evidence of no specificity (i.e., retaining the null hypothesis for the test of specificity).
9. Our decision to focus on the anger, disgust, fear, happiness, sadness, and surprise categories was reinforced by two observations. First, consider a recent poll that asked scientists about their beliefs (Ekman, 2016). Two-hundred forty-eight scientists who published peer-reviewed articles on the topic of emotion were given a list of 18 emotion labels and were asked to indicate which, according to available empirical evidence, have been established as biological categories with universal expressions. Of the 149 (60%) who responded,

> There was high agreement about five emotions . . .: anger (91%), fear (90%), disgust (86%), sadness (80%), and happiness (76%). Shame, surprise, and embarrassment were endorsed by 40%–50%. Other emotions, currently under study by various investigators drew substantially less support: guilt (37%), contempt (34%), love (32%), awe (31%), pain (28%), envy (28%), compassion (20%), pride (9%), and gratitude (6%). (Ekman, 2016, p. 32)

Second, there is no smoking gun in the published research on these additional emotion categories—that is, thus far, there are no scientific findings related to the production or perception of facial expressions for those emotion categories that challenge the general conclusions of this article. Simply put, regardless of how many emotion categories we evaluated, the pattern of findings was the same.
10. Different numbers of facial muscles are reported in various sources depending on how muscles are grouped or divided.
11. Regarding facial expressions:

> Scientists often refer to a set of actions that occur on the face simultaneously as "facial events," rather than calling them facial expressions. It is more descriptive. The word "expression" suggests that something from the inside becomes observable on the outside. Yet not every facial behavior expresses an internal state – most probably do not. (Rosenberg, 2018)

12. Box 6 in the Supplemental Material presents a summary of computer-vision algorithms for automatically detecting facial actions.
13. Changes in illumination and face orientation are currently major hurdles.
14. Thirty-eight groups, each with its own face-reading algorithm, announced their intention to participate in the challenge (Benitez-Quiroz, Srinivasan, et al., 2017). Groups tuned their algorithms on the set of training images that were provided 2 weeks before the challenge deadline. Final evaluations were done on the testing set only. Of the original 38 groups, only 4 submitted results before the challenge ended.
15. These accuracy levels might be considered an upper estimate because of the characteristics of the training and test-image databases. The methods for choosing the database are described

in Benitez-Quiroz et al. (2016). Note that a number of images are posed and professionally taken. Some facial configurations are exaggerated. Under these idealized circumstances, manual verification of these faces was estimated at 81% accuracy.

16. It is also possible that an individual person has a repertoire of probabilistic physical changes that reliably and specifically occur during the instances of a single emotion category; for a number of reasons, however, this hypothesis has not yet been scientifically tested. Specific studies to address this question would be very helpful.

17. There are ways to get around this circularity by using unsupervised, data-driven methods to discover categories, but to date, studies have used supervised approaches in which categories are prescribed by human inference.

18. By relying on their own beliefs, scientists are using human consensus to identify when an emotional episode is occurring and which emotion category it belongs to (i.e., when they agree that fear or some other emotion is present, then it is said to be present). It is important to realize that every single experiment dealing with emotion to date relies on human inference in this way. Consensus inferences are made in many areas of science. In physics and astronomy, consensus emerges from expert scientists whose beliefs and assumptions often challenge the common-sense view, such as in the case of quantum mechanics, dark matter, and black holes. In other areas of psychology, consensus is used to define many mental categories, such as memory and attention; consensus is also used to define psychiatric categories, such schizophrenia and autism. Even defining depression as a mental as opposed to a physical illness is a matter of consensus rather than objective ground truth. But it is noteworthy that when it comes to emotions, scientists use exactly the same categories as nonscientists, which may give us cause for concern (as forewarned by William James, 1890/2007, 1894). For example, compare the findings in Box 8 with the recent survey of scientists who study emotion (Ekman, 2016): Of 149 scientists who responded, 88 continue to believe that certain emotion categories have universal physiological markers, despite meta-analyses showing otherwise.

19. These meta-analytic findings are consistent with an earlier summary published by Matsumoto et al. (2008): Of 14 studies using rigorous FACS coding by human experts, only 5 reported that participants spontaneously displayed some or all of the hypothesized AUs during emotions. This is in contrast to the 9 studies that used the less reliable emotion FACS (EMFACS) coding, all of which reported support. These findings suggest that some type of perceptual bias creeps in when observers make judgments of whether an AU is present or not (e.g., indicating whether a participant is smiling, or displaying "happiness") than when AUs are coded independently, one at a time.

20. Remote, small-scale cultures are not untouched by Western influences. All cultures have some minimal contact with Western cultures (and this was also the case for the seminal articles published by Ekman and his colleagues in the 1970s; Crivelli & Gendron, 2017).

21. The Trobriand Islanders and the Fore are different ethnic groups; Trobrianders are subsistence fisherman and horticulturalists living in a small archipelago of islands located 200 km from the mainland where the Fore (who were photographed) lived; recall that the Fore photos were judged by Trobrianders in Crivelli et al., 2017). As Crivelli et al. make clear in their article, these findings are a within-nation rather than a within-culture comparison.

22. The value of this particular study is that the researchers not only coded infants' facial movements but also measured a range of concurrent movements that could support inferences about the infants' feelings of pleasantness, unpleasantness, and level of arousal, termed *affect* (see Box 9 in the Supplemental Material), including increased respiration, withdrawal/leaning away with the body, stilling/freezing, struggling, turning toward the mother, extreme withdrawal, hiding of their faces, squirming, self-stimulation, looking toward mother, pointing at the object, doing a "double-take," and banging on the table.

23. Bennett et al. (2002) note that when they observed facial actions that were thought to be associated with more than one emotion category—for example, when an infant produced a facial configuration that was a combination of scowling (anger) and pouting (sadness)—they interpreted the expression using the facial actions in only the upper region of the face, which indicates that infants' facial movements were even more variable than reported in the data tables. A footnote in the article further indicates that infants produced facial movements that were interpreted to reflect "interest" across all of the eliciting situations, but these facial actions were not included in any data analyses (Bennett et al., 2002, footnote 1). Any facial configuration that included AUs stipulated as interest and AUs for another emotion category was coded as an expression of the other emotion category.

24. In addition, it is not clear that children find sour foods disgusting (e.g., Rozin, Hammer, Oster, Horowitz, & Marmora, 1986; Stein, Ottenberg, & Roulet, 1958). Young children appear to be *attracted* to many things that adults find disgusting, whereas by the age of five, children have more adult-like behavioral responses and reject them (Rozin et al., 1986). For a discussion of how disgust is learned, see Widen and Russell (2013).

25. In another naturalistic study, videos of children ages 4 through 7 were downloaded from the Internet and FACS coded (Shuster, Camras, Grabell, & Perlman, 2018). The children were playing "the scary maze game": A child solves maze after maze of increasing difficulty, only to encounter a screaming, demonic girl from the movie *The Exorcist* (filmed in 1973). The game is generally thought to evoke an instance of fear (hence the name "scary"), but it may also evoke surprise as the scary stimulus makes a sudden unexpected appearance. Children produced the wide-eyed, gasping configuration (the proposed facial expression of fear) and/or a startled configuration (the proposed facial expression of surprise) with only weak reliability (38% and 10%, respectively).

26. By analogy, people who have been blind since birth learn color concepts and the relation between these concepts, such as "red," "blue," and "green" are similar to those of sighted people (e.g., congenitally blind individuals understand the U.S. concept for "blue" is more similar to "green" than to "red"; Shepard & Cooper, 1992). The structure of brain regions in the visual cortex that represent visual concepts are also virtually

indistinguishable in sighted and congenitally blind individuals (Koster-Hale, Bedny, & Saxe, 2014; X. Wang et al., 2015).

27. The onset and severity of blindness varies hugely across studies. Even a small amount of visual experience in infancy or early childhood will influence brain development and provide experiences for learning about emotions (see earlier section on emotion-concept development in infants). Helen Keller, for example, could see and hear until she was 19 months old, providing some initial scaffolding for her later ability to communicate.

28. For example, Ekman (2017) recently wrote,

> Another challenge to the findings of universality came from the anthropologist, Margaret Mead . . . Establishing that posed expressions are universal, she said, does not necessarily mean that spontaneous expressions are universal. I replied (Ekman, 1977) that it seemed illogical to presume that people can readily interpret posed facial expressions if they had not seen those facial expressions and experienced them in actual social life. (p. 46)

29. Although these findings are instructive, they probably provide a lower limit of the possible real-world variation in the facial configurations that express the varied instances of a given emotion category. After all, the Internet is a curated version of reality, and some frequent facial configurations are likely missing because they are rarely uploaded to the Internet. Likewise, some configurations commonly found on the Internet might not be commonly observed in the real world.

30. Compare these findings with those from a study that mined images from the Internet using a similar but narrower approach, and who had two raters use a choice-from-array method to label the images (Mollahosseini et al., 2016).

31. Configuration 3 also resembles people's beliefs about the configurations that express fear and awe (i.e., the "international core patterns" reported by Cordaro et al., 2018).

32. More generally, participants are more likely to perceive the intended emotion in the hypothesized facial configurations of Figure 4 when they are displayed on dynamically moving, synthetic faces (Wehrle, Kaiser, Schmidt, & Scherer, 2000), in video footage of posed facial muscle movements (e.g., Ambadar, Schooler, & Cohn, 2005; Cunningham & Wallraven, 2009), and even in point-light displays of motion created by facial muscle movements (Bassili, 1979). This "dynamic advantage" sometimes disappears when participants are viewing real human faces (e.g., Fiorentini & Viviani, 2011; Gold et al., 2013; Miles & Johnston, 2007; N. L. Nelson & Russell, 2011).

33. Ekman and Friesen (1971) was chosen as one of the 40 studies that changed psychology (Hock, 2009) and, along with Ekman et al. (1969), is routinely discussed in introductory psychology textbooks.

34. Dioula participants from Burkina Faso in West Africa showed strong reliability for labeling smiling facial configurations as happiness, moderate reliability for labeling frowning facial configurations as sadness, startled facial configurations as surprise, and nose-wrinkled facial configurations as disgust. They showed weak reliability for labeling scowling facial configurations as anger and wide-eyed, gasping facial configurations as fear.

35. For example, a sample of Trobriand Islanders, who are subsistence horticulturalists and fishermen living in the Trobriand Islands of Papua New Guinea, labeled a scowling facial configuration as anger with above chance reliability (29% of the time), but also labeled that facial configuration more frequently with "feels like avoiding a social interaction" (50% of the time; Crivelli et al., 2017, Study 2). In fact, the wide-eyed, gasping facial configuration that is thought to be the expression for fear (Fig. 4) is understood as an expression of aggression or threat in the Trobriand culture (Crivelli et al., 2017; Crivelli, Jarillo, & Fridlund, 2016; Crivelli, Russell, Jarillo, & Fernández-Dols, 2016). Trobrianders uniquely labeled smiling facial configurations as happiness across two studies but this finding was not replicated in a third sample or in a sample of Mwani participants who are subsistence fisherman living on Matemo Island in Mozambique, Africa.

36. The ancestors of the Hadza are thought to have been continuously practicing a hunting-and-gathering lifestyle for at least the past 50,000 years in their current region of East Africa. Furthermore, Hadza social structure, mobility, residential patterns, and language have thus far remained largely buffered from their interactions with other ethnic groups (Apicella & Crittenden, 2016; Crittenden & Marlowe, 2008) which have been sustained for at least the past 100 years (Jones, 2016).

37. It has been suggested that the wide-eyed, gasping stereotype for fear evolved for enhanced sensory sampling that supports efficient threat detection (Susskind et al., 2008). Likewise, the nose-wrinkle stereotype for disgust is thought to have evolved to limit exposure to noxious stimuli (Chapman & Anderson, 2013; Chapman, Kim, Susskind, & Anderson, 2009).

38. Note that adult perceivers may have done less overt looking at the postures, but other evidence with the same stimuli suggest that different body contexts influenced how adult participants visually scanned the exact same facial configurations; Aviezer et al., 2008). At the other end of the age spectrum, older adults are also more influenced by context than are young adults when inferring emotional meaning in facial configurations (Ngo & Isaacowitz, 2015).

39. Some applications will not be affected by context because they are not aiming to use facial movements to infer an individual's underlying emotional state. These initiatives have very specific applications in mind. For example, detecting pain in patients (Apple), driver drowsiness (Google), creating virtual facial expression stickers or animojis from one's own facial poses (Facebook, iPhone X), or Alibaba's "smile to pay."

40. The word "appraisal" has two meanings in the science of emotion. Here, appraisals simply to refer to the descriptive features of how a situation is experienced (e.g., novelty, goal relevance) without any inference about how those experiential features are caused (e.g., Clore & Ortony, 2008, 2013). The other meaning of appraisal refers to the mechanisms that cause the experiential features as components of emotion (e.g., the component process model of emotion, in which appraisals are considered evaluative "checks" that the human mind uses in a serial fashion; e.g., Scherer, Mortillaro, & Mehu, 2017). There is very little evidence that appraisals are, in fact, causal mechanisms (for a discussion, see Parkinson, 1997). In some studies, for example, participants are presented with a written scenario that is assumed to automatically trigger a specific sequence of appraisal checks (i.e., cognitive evaluations), which in turn is hypothesized to produce a specific pattern of facial muscle movements. Notice that the main causal mechanisms here—appraisal checks—are not measured directly but are inferred to

have occurred. In other studies, participants are asked to explicitly report on the appraisals they experience, on the assumption that the corresponding "checks" are active. There is abundant evidence, however, that although people can explicitly report on *what* they are feeling or doing, they are very inaccurate at reporting on the *causes* of their feelings and actions (i.e., they cannot accurately report on psychological mechanisms; Nisbett & Wilson, 1977; T. D. Wilson, 2002). Emerging scientific evidence links appraisals, as descriptive features, to facial movements, although the evidence to date suggests that these relationships are not as reliable or as specific as hypothesized (a summary of this research program can be found in Scherer et al., 2017).

References

Affectiva.com. (2018). Solutions. Retrieved from https://www.affectiva.com/what/products/

Adolphs, R. (2002). Neural mechanisms for recognizing emotions. *Current Opinion in Neurobiology*, *12*, 169–178.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*, 669–672.

Alexander, O., Rogers, M., Lambeth, W., Chiang, M., & Debevec, P. (2009) Creating a photoreal digital actor: The digital Emily project. In *Proceedings of the 2009 European Conference on Computer Vision for Media Production (CVMP)* (pp. 176–187). New York, NY: IEEE. doi:10.1109/CVMP.2009.29

Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior*, *33*, 17–34.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*, 403–410.

Apicella, C. L., & Crittenden, A. N. (2016). Hunter-gatherer families and parenting. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 797–827). Hoboken, NJ: John Wiley & Sons.

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience*, *20*, 1404–1412.

Arya, A., DiPaola, S., & Parush, A. (2009). Perceptually valid facial expressions for character-based applications. *International Journal of Computer Games Technology*, *2009*, Article 462315. doi:10.1155/2009/462315

Atzil, S., Gao, W., Fradkin, I., & Barrett, L. F. (2018). Growing a social brain. *Nature Human Behavior*, *2*, 624–636.

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., . . . Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*, 724–732.

Bandes, S. A. (2014). Remorse, demeanor, and the consequences of misinterpretation. *Journal of Law, Religion and State*, *3*, 170–199. doi:10.1163/22124810-00302004

Baron-Cohen, S., Golan, O., Wheelwright, S., & Hill, J. J. (2004). *Mind reading: The interactive guide to emotions*. London, England: Jessica Kingsley.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*, 241–251.

Barrett, D. (2010). *Supernormal stimuli: How primal urges overran their evolutionary purpose*. New York, NY: W.W. Norton.

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., . . . Pisor, A. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences, USA*, *113*, 4688–4693.

Barrett, L. F. (2004). Feelings or words? Understanding the content in self-report ratings of emotional experience. *Journal of Personality and Social Psychology*, *87*, 266–281. doi:10.1037/0022-3514.87.2.266

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*, 28–58.

Barrett, L. F. (2011). Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science*, *20*, 400–406.

Barrett, L. F. (2017a). Facial action coding. Retrieved from https://how-emotions-are-made.com/notes/Facial_action_coding

Barrett, L. F. (2017b). *How emotions are made: The secret life of the brain*. New York, NY: Houghton Mifflin Harcourt.

Barrett, L. F. (2017c). Screening of passengers by observation technique. Retrieved from https://how-emotions-are-made.com/notes/Screening_of_Passengers_by_Observation_Techniques

Barrett, L. F., & Finlay, B. L. (2018). Concepts, goals and the control of survival-related behaviors. *Current Opinion in the Behavioral Sciences*, *24*, 172–179.

Barrett, L. F., Lindquist, K., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., & Brennan, L. (2007). Of mice and men: Natural kinds of emotion in the mammalian brain? *Perspectives on Psychological Science*, *2*, 297–312.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*, 286–290.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, *58*, 373–403.

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, *37*, 2049–2058.

Baum, S. (Creator) & Grazer, B. (Producer). (2009, January 21). *Lie to me* [Television series]. Los Angeles, CA: Fox.

Baylor, A. L., & Kim, S. (2008, September). *The effects of agent nonverbal communication on procedural and attitudinal learning outcomes*. Paper presented at the International Conference on Intelligent Virtual Agents, Tokyo, Japan. doi:10.1007/978-3-540-85483-8_21

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217–239.

Bedny, M., & Saxe, R. (2012). Insights into the origins of knowledge from the cognitive neuroscience of blindness. *Cognitive Neuropsychology*, *29*, 56–84.

Benitez-Quiroz, C. F., Srinivasan, R., Feng, Q., Wang, Y., & Martinez, A. M. (2017). *EmotioNet Challenge: Recognition of facial expressions of emotion in the wild*. arXiv preprint arXiv:1703.01210.

Benitez-Quiroz, C. F., Srinivasan, R., & Martinez, A. M. (2016, June). Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5562–5570). New York, NY: IEEE. doi:10.1109/CVPR.2016.600

Benitez-Quiroz, C. F., Wang, Y., & Martinez, A. M. (2017, October). Recognition of action units in the wild with deep nets and a new global-local loss. In *Proceedings of the 16th IEEE International Conference on Computer Vision* (pp. 3990–3999). New York, NY: IEEE. doi:10.1109/ICCV.2017.428

Bennett, D. S., Bendersky, M., & Lewis, M. (2002). Facial expressivity at 4 months: A context by expression analysis. *Infancy*, *3*, 97–113.

Berggren, S., Fletcher-Watson, S., Milenkovic, N., Marschik, P. B., Bölte, S., & Jonsson, U. (2018). Emotion recognition training in autism spectrum disorder: A systematic review of challenges related to generalizability. *Developmental Neurorehabilitation*, *21*, 141–154.

Bick, J., & Nelson, C. A. (2016). Early adverse experiences and the developing brain. *Neuropsychopharmacology*, *41*, 177–196.

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*, 103–124.

Bolzani Dinehart, L. H., Messinger, D. S., Acosta, S. I., Cassel, T., Ambadar, Z., & Cohn, J. (2005). Adult perceptions of positive and negative infant emotional expressions. *Infancy*, *8*, 279–303.

Boucher, J. D., & Carlson, G. E. (1980). Recognition of facial expression in three cultures. *Journal of Cross-Cultural Psychology*, *11*, 263–280.

Bridges, C. B. (1932). The suppressors of purple. *Zeitschrift für induktive Abstammungs- und Vererbungslehre*, *60*, 207–218.

Briggs-Gowan, M. J., Pollak, S. D., Grasso, D., Voss, J., Mian, N. D., Zobel, E., . . . Pine, D. S. (2015). Attention bias and anxiety in young children exposed to family violence. *Journal of Child Psychology and Psychiatry*, *56*, 1194–1201.

Brodny, G., Kolakowska, A., Landowska, A., Szwoch, M., Szwoch, W., & Wróbel, M. R. (2016, July). *Comparison of selected off-the-shelf solutions for emotion recognition based on facial expressions*. Paper presented at the 9th International Conference on Human System Interactions, Portsmouth, England. doi:10.1109/HSI.2016.7529664

Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, *8*, 135–148.

Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., . . . Chavez, B. (2016). Detecting affiliation in colaughter across 24 societies. *Proceedings of the National Academy of Sciences, USA*, *113*, 4682–4687.

Buck, R. (1984). *The communication of emotion*. New York, NY: Guilford Press.

Bui, D., Heylen, D., Poel, M., & Nijholt, A. (2004, June). Combination of facial movements on a 3D talking head. In *Proceedings of the 21st Computer Graphics International Conference* (pp. 284–291). New York, NY: IEEE. doi:10.1109/CGI.2004.1309223

Cacioppo, J. T., Berntson, G. G., Larsen, J. H., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In R. Lewis & J. M. Haviland-Jones (Eds.), *The handbook of emotions* (2nd ed., pp. 173–191). New York, NY: Guilford Press.

Cain, J. (2000). *The way I feel*. Seattle, WA: Parenting Press.

Caldara, R. (2017). Culture reveals a flexible system for face processing. *Current Directions in Psychological Science*, *26*, 249–255. doi:10.1177/0963721417710036

Camras, L. A. (1992). Expressive development and basic emotions. *Cognition & Emotion*, *6*, 269–283.

Camras, L. A., & Allison, K. (1985). Children's understanding of emotional facial expressions and verbal labels. *Journal of Nonverbal Behavior*, *9*, 84–94.

Camras, L. A., Castro, V. L., Halberstadt, A. G., & Shuster, M. M. (2017). Spontaneously produced facial expressions in infants and children. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 279–296). New York, NY: Oxford University Press.

Camras, L. A., Fatani, S. S., Fraumeni, B. R., & Shuster, M. M. (2016). The development of facial expressions. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (4th ed., pp. 255–271). New York, NY: Guilford Press.

Camras, L. A., Oster, H., Ujiie, T., Campos, J. J., Bakeman, R., & Meng, Z. (2007). Do infants show distinct negative facial expressions for fear and anger? Emotional expression in 11-month-old European American, Chinese, and Japanese infants. *Infancy*, *11*, 131–155.

Camras, L. A., Perlman, S. B., Fries, A. B. W., & Pollak, S. D. (2006). Post-institutionalized Chinese and Eastern European children: Heterogeneity in the development of emotion understanding. *International Journal of Behavioral Development*, *30*, 193–199.

Camras, L. A., & Shutter, J. M. (2010). Emotional facial expressions in infancy. *Emotion Review*, *2*, 120–129.

Caron, R. F., Caron, A. J., & Myers, R. A. (1985). Do infants see emotional expressions in static faces? *Child Development*, *56*, 1552–1560.

Carrera-Levillain, P., & Fernández-Dols, J. M. (1994). Neutral faces in context: Their emotional meaning and their function. *Journal of Nonverbal Behavior*, *18*, 281–299.

Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, *70*, 205–218.

Carver, L. J., & Vaccaro, B. G. (2007). 12-month-old infants allocate increased neural resources to stimuli associated with negative adult emotion. *Developmental Psychology*, *43*, 54–69.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. F. (Eds.). (2000). *Embodied conversational agents*. Cambridge, MA: MIT Press.

Cassia, V. M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science*, *15*, 379–383.

Castro, V. L., Camras, L. A., Halberstadt, A. G., & Shuster, M. (2018). Children's prototypic facial expressions during emotion-eliciting conversations with their mothers. *Emotion*, *18*, 260–276. doi:10.1037/emo0000354

Cattaneo, L., & Pavesi, G. (2014). "The facial motor system." *Neuroscience & Biobehavioral Reviews*, *38*, 135–159.

Cecchini, M., Baroni, E., Di Vito, C., & Lai, C. (2011). Smiling in newborns during communicative wake and active sleep. *Infant Behavior & Development*, *34*, 417–423.

Chapman, H. A., & Anderson, A. K. (2013). Things rank and gross in nature: A review and synthesis of moral disgust. *Psychological Bulletin*, *139*, 300–327. doi:10.1037/a0030964

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, *323*, 1222–1226.

Chiesa, S., Galati, D., & Schmidt, S. (2015). Communicative interactions between visually impaired mothers and their sighted children: Analysis of gaze, facial expressions, voice and physical contacts. *Child: Care, Health and Development*, *41*, 1040–1046.

Chu, W. S., De la Torre, F., & Cohn, J. F. (2017). Selective transfer machine for personalized facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*, 529–545.

Cicchetti, D., & Curtis, W. J. (2005). An event-related potential study of the processing of affective facial expressions in young children who experienced maltreatment during the first year of life. *Development and Psychopathology*, *17*, 641–677.

Clark-Polner, E., Johnson, T., & Barrett, L. F. (2017). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*, *27*, 1944–1948.

Cleese, J. (Writer), Erskine, J., & Stewart, D. (Directors). (2001, March 7). *The human face* [Television series]. London, England: BBC.

Clore, G. L., & Ortony, A. (1991). What more is there to emotion concepts than prototypes? *Journal of Personality and Social Psychology*, *60*, 48–50.

Clore, G. L., & Ortony, A. (2008). Appraisal theories: How cognition shapes affect into emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 628–642). New York, NY: Guilford Press.

Clore, G. L., & Ortony, A. (2013). Psychological construction in the OCC model of emotion. *Emotion Review*, *5*, 335–343. doi:10.1177/1754073913489751

Cole, P. M. (1986). Children's spontaneous control of facial expression. *Child Development*, *57*, 1309–1321.

Cole, P. M., Jenkins, P. A., & Shott, C. T. (1989). Spontaneous expressive control in blind and sighted children. *Child Development*, *60*, 683–688.

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion*, *18*, 75–93. doi:10.1037/emo0000302

Corneanu, C. A., Simon, M. O., Cohn, J. F., & Guerrero, S. E. (2016). Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*, 1548–1568.

Crittenden, A. N., & Marlowe, F. W. (2008). Allomaternal care among the Hadza of Tanzania. *Human Nature*, *19*, 249–262.

Crivelli, C., Carrera, P., & Fernández-Dols, J.-M. (2015). Are smiles a sign of happiness? Spontaneous expressions of judo winners. *Evolution & Human Behavior*, *36*, 52–58.

Crivelli, C., & Gendron, M. (2017). Facial expressions and emotions in indigenous societies. In J. M. Fernandez-Dols & J. A. Russell (Eds). *The science of facial expression* (pp. 497–515). New York, NY: Oxford.

Crivelli, C., Jarillo, S., & Fridlund, A. J. (2016). A multidisciplinary approach to research in small-scale societies: Studying emotions and facial expressions in the field. *Frontiers in Psychology*, *7*, Article 1073. doi:10.3389/fpsyg.2016.01073

Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J. M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, *145*, 830–843. doi:10.1037/xge0000172

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J. M. (2016). The fear gasping face as a threat display in a Melanesian society. *Proceedings of the National Academy of Sciences, USA*, *113*, 12403–12407. doi:10.1073/PNAS.1611622113

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J. M. (2017). Recognizing spontaneous facial expressions of emotion in a small scale society of Papua New Guinea. *Emotion*, *17*, 337–347.

Cunningham, D. W., & Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *Journal of Vision*, *9*(13), Article 7. doi:10.1167/9.13.7

Danziger, K. (2006). Universalism and indigenization in the history of modern psychology. In A. C. Brock (Ed.), *Internationalizing the history of psychology* (pp. 208–225). New York, NY: New York University Press.

Darwin, C. (1965). *The expression of the emotions in man and animals*. Chicago, IL: University of Chicago Press. (Original work published 1872)

da Silva Ferreira, G. C., Crippa, J. A., & Osório, F. L. (2014). Facial emotion processing and recognition among maltreated children: A systematic literature review. *Frontiers in Psychology*, *5*, Article 1460. doi:10.3389/fpsyg.2014.01460

DeCarlo, L. T. (2012). On a signal detection approach to *m*-alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation. *Journal of Mathematical Psychology*, *56*(3), 196–207. doi:10.1016/j.jmp.2012.02.004

de Melo, C., Carnevale, P., Read, S., & Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, *106*, 73–88.

de Mendoza, A. H., Fernández-Dols, J. M., Parrott, W. G., & Carrera, P. (2010). Emotion terms, category structure,

and the problem of translation: The case of shame and vergüenza. *Cognition & Emotion*, *24*, 661–680.

Detego Group. (2018). Paul Ekman PhD. Retrieved from http://www.detegogroup.eu/paul-ekman-introduction/?lang=en

DiGirolamo, M. A., & Russell, J. A. (2017). The emotion seen in a face can be a methodological artifact: The process of elimination hypothesis. *Emotion*, *17*, 538–546. doi:10.1037/emo0000247

Ding, Y., Prepin, K., Huang, J., Pelachaud, C., & Artières, T. (2014). Laughter animation synthesis. In *Proceedings of AAMAS '14 International Conference on Autonomous Agents and Multiagent Systems* (pp. 773–780). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Docter, P., & Del Carmen, R. (Directors), LeFauve, M., & Cooley, J. (Writers), & Lasseter, J. (Producer). (2015). *Inside out* [Motion picture]. Emeryville, CA: Pixar.

Dondi, M., Gervasi, M. T., Valente, A., Vacca, T., Bogana, G., De Bellis, I., . . . Oster, H. (2012). Spontaneous facial expressions of distress in fetuses. In C. De Sousa & A. M. Oliveira (Eds.), *Proceedings of the 14th European Conference on Facial Expression: New Challenges for Research* (pp. 16–18). Coimbra, Portugal: University of Coimbra.

Dondi, M., Messinger, D., Colle, M., Tabasso, A., Simion, F., Barba, B. D., & Fogel, A. (2007). A new perspective on neonatal smiling: Differences between the judgments of expert coders and naive observers. *Infancy*, *12*, 235–255.

Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General*, *147*, 62–73. doi:10.1037/xge0000361

Duchenne, G.-B. (1990). *The mechanism of human facial expression* (R. A. Cutherbertson, Trans.). London, England: Cambridge University Press. (Original work published 1862).

Duran, J. I., & Fernández-Dols, J.-M. (2018). *Do emotions result in their predicted facial expressions? A meta-analysis of studies on the link between expression and emotion*. PsyArXiv preprint. Retrieved from https://psyarxiv.com/65qp7

Duran, J. I., Reisenzein, R., & Fernández-Dols, J.-M. (2017). Coherence between emotions and facial expressions: A research synthesis. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 107–129). New York, NY: Oxford University Press.

Eibl-Eibesfeldt, I. (1972). Similarities and differences between cultures in expressive movements. In R. A. Hinde (Ed.), *Non-verbal communication* (pp. 297–314). Oxford, England: Cambridge University Press.

Eibl-Eibesfeldt, I. (1989). *Human ethology* (P. Wiessner-Larsen & A. Heunemann, Trans.). New York, NY: Aldine deGruyter.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.), *Nebraska Symposium on Motivation, 1971* (pp. 207–283). Lincoln: University of Nebraska Press.

Ekman, P. (1980). *The face of man*. New York, NY: Garland Publishing, Inc.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*, 169–200.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, *115*, 268–287.

Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on Psychological Science*, *11*, 31–34. doi:10.1177/1745691615596992

Ekman, P. (2017). Facial expressions. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 39–56). New York, NY: Oxford University Press.

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

Ekman, P., Friesen, W., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings*. Elmsford, NY: Pergamon Press.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–129.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System: The manual On CD ROM*. Salt Lake City, UT: The Human Face.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*, 1208–1210.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86–88.

Elfenbein, H., Beaupre, M., Levesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, *7*, 131–146.

Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review*, *5*, 90–96.

Elfenbein, H. A. (2017). Emotional dialects in the language of emotion. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 479–496). New York, NY: Oxford University Press.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203–235.

Emde, R. N., & Koenig, K. L. (1969). Neonatal smiling and rapid eye movement states. *Journal of the American Academy of Child Psychiatry*, *8*(1), 57–67.

Emojipedia.org. (2019). Emoji people and smileys meanings. Retrieved from https://emojipedia.org/people/

Essa, I. A., & Pentland, A. P. (1997). Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*, 757–763.

Feldman, R. (2016). The neurobiology of mammalian parenting and the biosocial context of human caregiving. *Hormones and Behavior*, *77*, 3–17.

Feng, D., Jeong, D., Krämer, N., Miller, L., & Marsella, S. (2017). "Is it just me?" Evaluating attribution of negative

feedback as a function of virtual instructor's gender and proxemics. In *Proceedings of AAMAS '16 International Conference on Autonomous Agents and Multiagent Systems* (pp. 810–818). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Fernández-Dols, J.-M. (2017). Natural facial expression: A view from psychological construction and pragmatics. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 457–478). New York, NY: Oxford University Press.

Fernández-Dols, J.-M., & Crivelli, C. (2013). Emotion and expression: Naturalistic studies. *Emotion Review*, *5*, 24–29. doi:10.1177/1754073912457229

Fernández-Dols, J.-M., & Ruiz-Belda, M.-A. (1995). Are smiles a sign of happiness? Gold medal winners at the Olympic Games. *Journal of Personality and Social Psychology*, *69*, 1113–1119.

Fernández-Dols, J.-M., Sanchez, F., Carrera, P., & Ruiz-Belda, M.-A. (1997). Are spontaneous expressions and emotions linked? An experimental test of coherence. *Journal of Nonverbal Behavior*, *21*, 163–177.

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision*, *11*(3), Article 17. doi:10.1167/11.3.17

Firestein, S. (2012). *Ignorance: How it drives science*. Oxford, England: Oxford University Press.

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*, 238–252.

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology*, *80*, 75–85.

Fridlund, A. J. (2017). The behavioral ecology view of facial displays, 25 years later. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 77–92). New York, NY: Oxford University Press.

Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology*, *60*, 229–240.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.

Galati, D., Miceli, R., & Sini, B. (2001). Judging and coding facial expression of emotions in congenitally blind children. *International Journal of Behavioral Development*, *25*, 268–278. doi:10.1080/01650250042000393

Galati, D., Scherer, K. R., & Ricci-Bitti, P. E. (1997). Voluntary facial expression of emotion: Comparing congenitally blind with normally sighted encoders. *Journal of Personality and Social Psychology*, *73*, 1363–1379.

Galati, D., Sini, B., Schmidt, S., & Tinti, C. (2003). Spontaneous facial expressions in congenitally blind and sighted children aged 8-11. *Journal of Visual Impairment & Blindness*, *97*, 418–428.

Gandhi, T. K., Singh, A. K., Swami, P., Ganesh, S., & Sinha, P. (2017). Emergence of categorical face perception after extended early-onset blindness. *Proceedings of the National Academy of Sciences, USA*, *114*, 6139–6143.

Gendron, M., & Barrett, L. F. (2009). Reconstructing the past: A century of ideas about emotion in psychology. *Emotion Review*, *1*, 316–339.

Gendron, M., & Barrett, L. F. (2017). Facing the past: A history of the face in psychological research on emotion perception. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 15–36). New York, NY: Oxford University Press.

Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality reconsidered: Diversity in meaning making of facial expressions. *Current Directions in Psychological Science*, *27*, 211–219. doi:10.1177/0963721417746794

Gendron, M., Hoemann, K., Crittenden, A. N., Msafiri, S., Ruark, G., & Barrett, L. F. (2018). *Emotion perception in the Hadza hunter-gatherers*. PsyArXiv. doi:10.31234/osf.io/pf2q3

Gendron, M., Lindquist, K., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, *12*, 314–325.

Gendron, M., Mesquita, B., & Barrett, L. F. (2013). Emotion perception: Putting a face in context. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 539–556). New York, NY: Oxford University Press.

Gendron, M., Roberson, D., & Barrett, L. F. (2015). Cultural variation in emotion perception is real: A response to Sauter et al. *Psychological Science*, *26*, 357–359.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, *25*, 911–920.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, *14*, 251–262. doi:10.1037/a0036052

Gewald, J.-B. (2010). Remote but in contact with history and the world. *Proceedings of the National Academy of Sciences, USA*, *107*(18), Article E75. doi:10.1073/pnas.1001284107

Gibbons, A. (2018). Farmers, tourists, and cattle threaten to wipe out some of the world's last hunter-gatherers. *Science*. doi:10.1126/science.aau2032

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 89–150). New York, NY: McGraw-Hill.

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., . . . Srinath, A. (2013). The efficiency of dynamic and static facial expression recognition. *Journal of Vision*, *13*(5), Article 23. doi:10.1167/13.5.23

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, *26*, 381–386.

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, *31*, 169–180.

Granchrow, J. R., Steiner, J. E., & Daher, M. (1983). Neonatal facial expressions in response to different qualities and intensities of gustatory stimulation. *Infant Behavior & Development*, *6*, 153–157.

Gross, J. J., & Barrett, L. F. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review*, *3*, 8–16.

Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, *28*, 219–236.

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological Bulletin*, *141*, 1266–1287. doi:10.1037/bul0000002

Guillory, S. A., & Bujarski, K. A. (2014). Exploring emotion using invasive methods: Review of 60 years of human intracranial electrophysiology. *Social Cognitive and Affective Neuroscience*, *9*, 1880–1889. doi:10.1093/scan/nsu002

Gunnery, S. D., & Hall, J. A. (2014). The Duchenne smile and persuasion. *Journal of Nonverbal Behavior*, *38*, 181–194. doi:10.1007/s10919-014-0177-1

Gunnery, S. D., Hall, J. A., & Ruben, M. A. (2013). The deliberate Duchenne smile: Individual differences in expressive control. *Journal of Nonverbal Behavior*, *37*, 29–41. doi:10.1007/s10919-012-0139-4

Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition & Emotion*, *13*, 225–266.

Hao, L., Wang, S., Peng, G., & Ji, Q. (2018). Facial action unit recognition augmented by their dependencies. *Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition* (pp. 187–194). New York, NY: IEEE. doi:10.1109/FG.2018.00036

Hata, T., Hanaoka, U., Mashima, M., Ishimura, M., Marumo, G., & Kanenishi, K. (2013). Four-dimensional HDlive rendering image of fetal facial expression: A pictorial essay. *Journal of Medical Ultrasonics*, *40*, 437–441.

Haviland, J. M., & Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. *Developmental Psychology*, *23*, 97–104.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? [Target article and commentaries]. *Behavioral and Brain Sciences*, *33*, 61–135. doi:10.1017/S0140525X0999152X

Hertenstein, M. J., & Campos, J. J. (2004). The retention effects of an adult's emotional displays on infant behavior. *Child Development*, *75*, 595–613.

Hock, R. R. (2009). *Forty studies that changed psychology: Explorations into the history of psychological research* (6th ed.). Upper Saddle River, NJ: Pearson Education.

Hjortsjö, C. H. (1969). *Man's face and mimic language*. Lund, Sweden: Studentlitteratur.

Hoehl, S., & Striano, T. (2008). Neural processing of eye gaze and threat-related emotional facial expressions in infancy. *Child Development*, *79*, 1752–1760.

Hoemann, K., Crittenden, A. N., Msafiri, S., Liu, Q., Li, C., Roberson, D., . . . Barrett, L. F. (2018). Context facilitates the cross-cultural perception of emotion. *Emotion*. Advance online publication. doi:10.1037/emo0000501

Holodynski, M., & Friedlmeier, W. (2006). *Development of emotions and emotion regulation* (Vol. 8). Berlin, Germany: Springer Science & Business Media.

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*, 256–281.

Hoyt, C., Blascovich, J., & Swinth, K. (2003). Social inhibition in immersive virtual environments. *Presence*, *12*, 183–195.

Hutto, J. R., & Vattoth, S. (2015). A practical review of the muscles of facial mimicry with special emphasis on the superficial musculoaponeurotic system. *American Journal of Radiology*, *204*, W19–W26.

Izard, C. E. (1971). *The face of emotion*. East Norwalk, CT: Appleton-Century-Crofts.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*, 260–280.

Izard, C. E., Fantauzzo, C. A., Castle, J. M., Haynes, O. M., Rayias, M. F., & Putnam, P. H. (1995). The ontogeny and significance of infants' facial expressions in the first 9 months of life. *Developmental Psychology*, *31*, 997–1013.

Izard, C. E., Hembree, E., Dougherty, L., & Spirrizi, C. (1983). Changes in 2-to 19-month-old infants' facial expressions following acute pain. *Developmental Psychology*, *19*, 418–426.

Izard, C. E., Hembree, E., & Huebner, R. (1987). Infants' emotional expressions to acute pain: Developmental changes and stability of individual differences. *Developmental Psychology*, *23*, 105–113.

Izard, C. E., Woodburn, E. M., & Finlon, K. J. (2010). Extending emotion science to the study of discrete emotions in infants. *Emotion Review*, *2*, 134–136.

Jack, R. E., Crivelli, C., & Wheatley, T. (2018). Data-driven methods to diversify knowledge of human psychology. *Trends in Cognitive Science*, *22*, 1–5. doi:10.1016/j.tics.2017.10.002

Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual Review of Psychology*, *68*, 269–297. doi:10.1146/annurev-psych-010416-044242

Jack, R. E., Sun, W., Delis, I., Garrod, O. G., & Schyns, P. G. (2016). Four not six: Revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General*, *145*, 708–730. doi:10.1037/xge0000162

James, W. (1894). The physical basis of emotion. *Psychological Review*, *1*, 516–529.

James, W. (2007). *The principles of psychology* (Vol. 1). New York, NY: Dover. (Original work published 1890)

Jeni, L., Cohn, J. F., & De la Torre, F. (2013). Facing imbalanced data recommendations for the use of performance metrics. In *ACII '13 Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 245–251). New York, NY: IEEE. doi:10.1109/ACII.2013.3

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*, 313–323. doi:10.1016/j.cognition.2011.08.001

Jones, N. B. (2016). *Demography and evolutionary ecology of Hadza hunter-gatherers* (Vol. 71). Cambridge, England: Cambridge University Press.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception*, *30*, 875–887.

Kayyal, M. H., & Russell, J. A. (2013). Americans and Palestinians judge spontaneous facial expressions of emotion. *Emotion*, *13*, 891–904. doi:10.1037/a0033244

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, *68*, 441–454.

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct forms and appeasements functions. *Psychological Bulletin*, *122*, 250–270.

Keltner, D., & Cordaro, D. T. (2017). Understanding multimodal emotional expressions: Recent advances in Basic Emotion Theory. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 57–75). New York, NY: Oxford University Press.

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior*. Advance online publication. doi:10.1007/s10919-019-00293-3

Kim, Y., Baylor, A. L., & Shen, E. (2007). Pedagogical agents as learning companions: The impact of agent affect and gender. *Journal of Computer Assisted Learning*, *23*, 220–532.

Kobiella, A., Grossmann, T., Reid, V. M., & Striano, T. (2008). The discrimination of angry and fearful facial expressions in 7-month-old infants: An event-related potential study. *Cognition & Emotion*, *22*, 134–146.

Koster-Hale, J., Bedny, M., & Saxe, R. (2014). Thinking about seeing: Perceptual sources of knowledge are encoded in the theory of mind brain regions of sighted and blind adults. *Cognition*, *133*, 65–78. doi:10.1016/j.cognition.2014.04.006

Kosti, R., Alvarez, J. M., Recasens, A., & Lapedriza, A. (2017). EMOTIC: Emotions in Context dataset. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Los Alamitos, CA: IEEE. Retrieved from http://openaccess.thecvf.com/content_cvpr_2017/papers/Kosti_Emotion_Recognition_in_CVPR_2017_paper.pdf

Kouo, J. L., & Egel, A. L. (2016). The effectiveness of interventions in teaching emotion recognition to children with autism spectrum disorder. *Review Journal of Autism and Developmental Disorders*, *3*, 254–265. doi:10.1007/s40489-016-0081-1

Kragel, P. A., & LaBar, K. S. (2013). Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. *Emotion*, *13*, 681–690.

Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, *10*, 1437–1448.

Krumhuber, E., & Kappas, A. (2005). Moving smiles: The role of dynamic components for the perception of the genuineness of smiles. *Journal of Nonverbal Behavior*, *29*, 3–24.

Krumhuber, E. G., & Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion*, *9*, 807–820. doi:10.1037/a0017844

Krumhuber, E., Manstead, A., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*, 730–735.

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., & Rosin, P. L. (2009). Effects of dynamic attributes of smiles in human and synthetic faces: A simulated job interview setting. *Journal of Nonverbal Behavior*, *33*, 1–15.

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, *5*, 41–46.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Leitzke, B. T., & Pollak, S. D. (2016). Developmental changes in the primacy of facial cues for emotion recognition. *Developmental Psychology*, *52*, 572–581.

Leppänen, J. M., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child Development*, *78*, 232–245.

Leppänen, J. M., & Nelson, C. A. (2009). Tuning the developing brain to social signals of emotions. *Nature Reviews Neuroscience*, *10*, 37–47.

Leppänen, J. M., Richmond, J., Vogel-Farley, V. K., Moulson, M. C., & Nelson, C. A. (2009). Categorical representation of facial expressions in the infant brain. *Infancy*, *14*, 346–362.

Levari, D. E., Gilbert, D. T., Wilson, T. D., Sievers, B., Amodio, D. M., & Wheatley, T. (2018). Prevalence-induced concept change in human judgment. *Science*, *360*, 1465–1467.

Levenson, R. W. (2011). Basic emotion questions. *Emotion Review*, *3*, 379–386.

Lewis, M., Ramsay, D. S., & Sullivan, M. W. (2006). The relation of ANS and HPA activation to infant anger and sadness response to goal blockage. *Developmental Psychobiology*, *48*, 397–405.

Lewis, M., Sullivan, M. W., & Kim, H. M. S. (2015). Infant approach and withdrawal in response to a goal blockage: Its antecedent causes and its effect on toddler persistence. *Developmental Psychology*, *51*, 1553–1563.

Lewis, M., & Sullivan, M. W. (Eds.). (2014). *Emotional development in atypical children*. New York, NY: Psychology Press.

Lindquist, K. A., & Barrett, L. F. (2008). Emotional complexity. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 513–530). New York, NY: Guilford Press.

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception but not affect perception is impaired with semantic memory loss. *Emotion*, *14*, 375–387.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral & Brain Sciences*, *35*, 121–143.

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological Science*, *25*, 1663–1672. doi:10.1177/0956797614541991

Marsella, S., & Gratch, J. (2016). Computational models of emotion as psychological tools. In L. F. Barrett, M. Lewis, & J. Haviland-Jones (Eds.), *Handbook of emotions* (4th ed., pp 113–132). New York, NY: Guilford Press.

Marsella, S. C., Johnson, W. L., & LaBore, C. (2000). Interactive pedagogical drama. In *Proceedings of the Fourth International Conference on Autonomous Agents (AGENTS '00)* (pp. 301–308). New York, NY: ACM. doi:10.1145/336595.337507

Martin, J., Rychlowska, M., Wood, A., & Niedenthal, P. (2017). Smiles as multipurpose social signals. *Trends in Cognitive Sciences*, *21*, 864–877.

Martin, N. G., Maza, L., McGrath, S. J., & Phelps, A. E. (2014). An examination of referential and affect specificity with five emotions in infancy. *Infant Behavior & Development*, *37*, 286–297.

Martinez, A., & Du, S. (2012). A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *Journal of Machine Learning Research*, *13*, 1589–1608.

Martinez, A. M. (2017a). Computational models of face perception. *Current Directions in Psychological Science*, *26*, 263–269.

Martinez, A. M. (2017b). Visual perception of facial expressions of emotion. *Current Opinion in Psychology*, *17*, 27–33.

Matias, R., & Cohn, J. F. (1993). Are max-specified infant facial expressions during face-to-face interaction consistent with differential emotions theory? *Developmental Psychology*, *29*, 524–531

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion*, *14*, 195–214.

Matsumoto, D., Keltner, D., Shiota, M., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 211–234). New York, NY: Macmillan.

Matsumoto, D., & Willingham, B. (2006). The thrill of victory and the agony of defeat: Spontaneous expressions of medal winners of the 2004 Athens Olympic Games. *Journal of Personality and Social Psychology*, *91*, 568–581.

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology*, *96*, 1–10.

McCall, C., Blascovich, J., Young, A., & Persky, S. (2009). Proxemic behaviors as predictors of aggression towards Black (but not White) males in an immersive virtual environment. *Social Influence*, *4*, 138–154.

McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 573–599

Medin, D. L., Ojalehto, B., Marin, A., & Bang, M. (2017). Systems of (non-)diversity. *Nature Human Behaviour*, *1*, Article 0088.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, *2*, 179–204.

Messinger, D. S. (2002). Positive and negative: Infant facial expressions and emotions. *Current Directions in Psychological Science*, *11*, 1–6.

Michel, G. F., Camras, L. A., & Sullivan, J. (1992). Infant interest expressions as coordinative motor structures. *Infant Behavior & Development*, *15*, 347–358.

Microsoft Azure. (2018). Cognitive services. Retrieved from https://azure.microsoft.com/en-us/services/cognitive-services/face/

Miles, L., & Johnston, L. (2007). Detecting happiness: Perceiver sensitivity to enjoyment and non-enjoyment smiles. *Journal of Nonverbal Behavior*, *31*, 259–275.

Miniland Group. (2019). Emotional journey. Retrieved from https://www.minilandgroup.com/us/usa/inteligencias-multiples/emotional-journey/

Mollahosseini, A., Hassani, B., Salvador, M. J., Abdollah, H., Chan, D., & Mahoor, M. N. (2016). *Facial expression recognition from world wild web*. arXiv. Retrieved from https://arxiv.org/abs/*1605*.03639

Montague, D. P., & Walker-Andrews, A. S. (2001). Peekaboo: A new look at infants' perception of emotion expressions. *Developmental Psychology*, *37*(6), 826.

Morenoff, D. (Director). (2014, October 14). Emotions [Video file]. Retrieved from https://vimeo.com/108524970

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*, 718–735.

Motley, M. T., & Camden, C. T. (1988). Facial expressions of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication*, *52*, 1–22.

Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development*, *67*, 3219–3237.

Müri, R. M. (2015). Cortical control of facial expression. *The Journal of Comparative Neurology*, *524*, 1578–1585.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion*, *7*, 736–744.

Namba, S., Makihara, S., Kabir, R. S., Miyatani, M., & Nakao, T. (2016). Spontaneous facial expressions are different from posed facial expressions: Morphological properties and dynamic sequences. *Current Psychology*, *36*, 593–605.

Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, *110*, 52–61.

Nelson, N. L., & Russell, J. A. (2016). Building emotion categories: Children use a process of elimination when they encounter novel expressions. *Journal of Experimental Child Psychology*, *151*, 120–130.

Neth, D., & Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *Journal of Vision*, *9*(1), Article 5. doi:10.1167/9.1.5

Ngo, N., & Isaacowitz, D. M. (2015). Use of context in emotion perception: The role of top-down control, cue type, and perceiver's age. *Emotion*, *15*, 292–302.

Niewiadomski, R., Ding, Y., Mancini, M., Pelachaud, C., Volpe, G., & Camurri, A. (2015). Perception of intensity incongruence in synthesized multimodal expressions of laughter. In *2015 International Conference on Affective Computing and Intelligent Interaction* (pp. 684–690). New York, NY: IEEE. doi:10.1109/ACII.2015.7344643

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Norenzayan, A., & Heine, S. J. (2005). Psychological universals: What are they and how can we know? *Psychological Bulletin*, *13*, 763–784.

Oakes, L. M., & Ellis, A. E. (2013). An eye-tracking investigation of developmental changes in infants' exploration of upright and inverted human faces. *Infancy*, *18*, 134–148.

Ochs, M., Niewiadomski, R., & Pelachaud, C. (2010). How a virtual agent should smile? Morphological and dynamic characteristics of virtual agent's smiles. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA '10)* (pp. 427–440). Berlin, Germany: Springer-Verlag.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*, 315–331.

Osgood, C. E., May, W. H., & Miron, M. S. (1975). *Cross-cultural universals of affective meaning*. Urbana: University of Illinois Press.

Oster, H. (2007). *BabyFACS: Facial action coding system for infants and young children*. Unpublished manuscript, New York University, New York, NY.

Oster, H. (2005). The repertoire of infant facial expressions: An ontogenetic perspective. In J. Nadel & D. Muir (Eds.), *Emotional development: Recent research advances* (pp. 261–292). New York, NY: Oxford University Press.

Oster, H., Hegley, D., & Nagel, L. (1992). Adult judgments and fine-grained analysis of infant facial expressions: Testing the validity of a priori coding formulas. *Developmental Psychology*, *28*, 1115–1131.

Parkinson, B. (1997). Untangling the appraisal- emotion connection. *Personality and Social Psychology Review*, *1*, 62–79.

Parr, L. A., Waller, B. M., Burrows, A. M., Gothard, K. M., & Vick, S.-J. (2010). MaqFACS: A muscle-based facial movement coding system for the rhesus Macaque. *American Journal of Physical Anthropology*, *143*, 625–630.

Parr, T. (2005). *The feelings book*. New York, NY: Little, Brown Books for Young Readers.

Pavlenko, A. (2014). *The bilingual mind: And what it tells us about language and thought*. New York, NY: Cambridge University Press.

Peltola, M. J., Leppänen, J. M., Palokangas, T., & Hietanen, J. K. (2008). Fearful faces modulate looking duration and attention disengagement in 7-month-old infants. *Developmental Science*, *11*, 60–68.

Perlman, S. B., Kalish, C. W., & Pollak, S. D. (2008). The role of maltreatment experience in children's understanding of the antecedents of emotion. *Cognition & Emotion*, *22*, 651–670.

Pinker, S. (1997). *How the mind works*. New York, NY: W.W. Norton.

Plate, R. C., Fulvio, J. M., Shutts, K., Green, C. S., & Pollak, S. D. (2018). Probability learning: Changes in behavior across time and development. *Child Development*, *89*, 205–218.

Plate, R. C., Wood, A., Woodard, K., & Pollak, S. D. (2018). Probabilistic learning of emotion categories. *Journal of Experimental Psychology: General*. Advance online publication. doi:10.1037/xge0000529

Pliskin, A. (Associate Producer). (2015, January 15). Episode 4515 [Television series episode]. In *Sesame street*. New York, New York: Sesame Workshop. https://www.youtube.com/watch?v=y28GH2GoIyc

Pollak, S. D. (2015). Multilevel developmental approaches to understanding the effects of child maltreatment: Recent advances and future challenges. *Development and Psychopathology*, *27*(4, Pt. 2), 1387–1397.

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.

Pollak, S. D., & Kistler, D. J. (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences, USA*, *99*, 9072–9076.

Pollak, S. D., Messner, M., Kistler, D. J., & Cohn, J. F. (2009). Development of perceptual expertise in emotion recognition. *Cognition*, *110*, 242–247.

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

Pollak, S. D., & Tolley-Schell, S. A. (2003). Selective attention to facial emotion in physically abused children. *Journal of Abnormal Psychology*, *112*, 323–338.

Pollak, S. D., Vardi, S., Bechner, A. M. P., & Curtin, J. J. (2005). Physically abused children's regulation of attention in response to hostility. *Child Development*, *76*, 968–977.

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review Psychology*, *28*, 363–392.

Rehm, M., & André, E. (2005). Catch me if you can: Exploring lying agents in social settings. In *Proceedings of AAMAS '05: Fourth International Joint Conference on Autonomous Agents and Multiagent Systems* (pp. 937–944). New York, NY: Association for Computing Machinery. doi:10.1145/1082473.1082615

Reissland, N., Francis, B., & Mason, J. (2013). Can healthy fetuses show facial expressions of "pain" or "distress"? *PLOS ONE*, *8*(6), Article e65530. doi:10.1371/journal.pone.0065530

Reissland, N., Francis, B., Mason, J., & Lincoln, K. (2011). Do facial expressions develop before birth? *PLOS ONE*, *6*(8), Article e24081. doi:10.1371/journal.pone.0024081

Rickel, R., Marsella, S., Gratch, J., Hill, R., Traum, D., & Swartout, B. (2002). Towards a new generation of virtual humans for interactive experiences. *In IEEE Intelligent Systems*, *17*(4), 32–38. doi:10.1109/MIS.2002.1024750

Riggins v. Nevada. 504 U.S. 127 (1992).

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, *95*, 52–77.

Robbins, J., & Rumsey, A. (2008). Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropological Quarterly*, *81*, 407–420.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*, 934–960.

Roch-Levecq, A.-C. (2006). Production of basic emotions by children with congenital blindness: Evidence for the embodiment of theory of mind. *British Journal of Developmental Psychology*, *24*, 507–528.

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68–91). New York, NY: Oxford University Press.

Rosenberg, E. (2018). FACS: what is FACS? Retrieved from http://erikarosenberg.com/facs/

Rosenberg, E. L., & Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition & Emotion*, *8*, 201–229.

Rosenstein, D., & Oster, H. (1988). Differential facial responses to four basic tastes in newborns. *Child Development*, *59*, 1555–1568.

Rozin, P., Hammer, L., Oster, H., Horowitz, T., & Marmora, V. (1986). The child's conception of food: Differentiation of categories of rejected substances in the 16 months to 5 year age range. *Appetite*, *7*, 141–151.

Ruba, A. L., Johnson, K. M., Harris, L. T., & Wilbourn, M. P. (2017). Developmental changes in infants' categorization of anger and disgust facial expressions. *Developmental Psychology*, *53*, 1826–1832.

Rudovic, O., Lee, J., Dai, M., Schuller, B., & Picard, R. W. (2018). Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, *3*(19), Article eaao6760. doi:10.1126/scirobotics.aao6760

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin*, *110*, 426–450.

Russell, J. A. (1993). Forced-choice response format in the study of facial expressions. *Motivation and Emotion*, *17*, 41–51.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, *115*, 102–141.

Russell, J. A. (1995). Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin*, *118*, 379–391.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*, 805–819.

Rychlowska, M., Jack, R. E., Garrod, O. G. B., Schyns, P. G., Martin, J. D., & Niedenthal, P. M. (2017). Functional smiles: Tools for love, sympathy, and war. *Psychological Science*, *28*, 1259–1270.

Rychlowska, M., Miyamoto, Y., Matsumoto, D., Hess, U., Gilboa-Schechtman, E., Kamble, S., . . . Niedenthal, P. M. (2015). Heterogeneity of long-history migration explains cultural differences in reports of emotional expressivity and the functions of smiles. *Proceedings of the National Academy of Sciences, USA*, *112*, E2429–E2436.

Saarimäki, H., Gotsopoulos, A., Jaaskelainen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., . . . Nummenmaa, L. (2016). Discrete neural signatures of basic emotions. *Cerebral Cortex*, *26*, 2563–2573.

Saarni, C., Campos, J. J., Camras, L. A., & Witherington, D. (2006). Emotional development: Action, communication, and understanding. In W. Damon, R. M. Lerner, & N. Eisenber (Eds.), *Handbook of child psychology, social, emotional, and personality development* (Vol. 3, pp. 226–299). New York, NY: John Wiley & Sons.

Sato, W., & Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cognition & Emotion*, *18*, 701–710.

Scherer, K. R., Mortillaro, M., & Mehu, M. (2017). Facial expression is driven by appraisal and generates appraisal inferences. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 353–374). New York, NY: Oxford University Press.

Schwartz, G. M., Izard, C. E., & Ansul, S. E. (1985). The 5-month-old's ability to discriminate facial expressions of emotion. *Infant Behavior & Development*, *8*, 65–77.

Sears, M. S., Repetti, R. L., Reynolds, B. M., & Sperling, J. B. (2014). A naturalistic observational study of children's expressions of anger in the family context. *Emotion*, *14*, 272–283.

Serrano, J. M., Iglesias, J., & Loeches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear, and surprise in 4-to 6-month-old infants. *Developmental Psychobiology*, *25*, 411–425.

Shackman, J. E., Fatani, S., Camras, L. A., Berkowitz, M. J., Bachorowski, J. A., & Pollak, S. D. (2010). Emotion expression among abusive mothers is associated with their children's emotion processing and problem behaviours. *Cognition & Emotion*, *24*, 1421–1430.

Shackman, J. E., & Pollak, S. D. (2014). Impact of physical maltreatment on the regulation of negative affect and aggression. *Development and Psychopathology*, *26*, 1021–1033.

Shackman, J. E., Shackman, A. J., & Pollak, S. D. (2007). Physical abuse amplifies attention to threat and increases anxiety in children. *Emotion*, *7*, 838–852.

Shariff, A. F., & Tracy, J. (2011). What are emotion expressions for? *Current Directions in Psychological Science*, *20*, 395–399.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*, 1061–1086.

Shepard, R. N., & Cooper, L. A. (1992). Representation of colors in the blind, color-blind, and normally sighted. *Psychological Science*, *3*, 97–104.

Shiota, M. N., Campos, B., & Keltner, D. (2003). The faces of positive emotion: Prototype displays of awe, amusement, and pride. *Annals of the New York Academy of Sciences*, *1000*, 296–299.

Shuster, M. M., Camras, L. A., Grabell, A., & Perlman, S. B. (2018). Faces in the wild: A naturalistic study of children's facial expressions in response to a YouTube video prank. *PsyArXiv*. doi:10.31234/osf.io/7vm5w

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., . . . Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic

investigation of autonomic features of emotion categories. *Psychological Bulletin*, *144*, 343–393.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108–131.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*, P325–P336. doi:10.1016/j.tics.2018.02.004

Smith, T. W. (2016). *The book of human emotion*. New York, NY: Little, Brown.

Soken, N. H., & Pick, A. D. (1999). Infants' perception of dynamic affective expressions: Do infants distinguish specific expressions? *Child Development*, *70*, 1275–1282.

Sorenson, E. R. (1975). Culture and the expression of emotion. In T. R. Williams (Ed.), *Psychological Anthropology* (pp. 361–372). Chicago, IL: Aldine.

Srinivasan, R., & Martinez, A. M. (2018). Cross-cultural and cultural-specific production and perception of facial expressions of emotion in the wild. *IEEE Transactions on Affective Computing*. Advance online publication. doi:10.1109/TAFFC.2018.2887267

Sroufe, A. L. (1996). *Emotional development: The organization of emotional life in the early years*. New York, NY: Cambridge University Press.

Stebbins, G., & Delsarte, F. (1887). *Delsarte system of expression*. New York, NY: E. S. Werner.

Stein, M., Ottenberg, P., & Roulet, N. (1958). A study of the development of olfactory preferences. *Archives of Neurology & Psychiatry*, *80*, 264–266.

Stemmler, G. (2004). Physiological processes during emotion. In P. Philippot & R. S. Feldman (Eds.), *The regulation of emotion* (pp. 33–70). Mahwah, NJ: Erlbaum.

Stenberg, C. R., Campos, J. J., & Emde, R. N. (1983). The facial expression of anger in seven-month-old infants. *Child Development*, *54*, 178–184.

Stephens, C. L., Christie, I. C., & Friedman, B. H. (2010). Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis. *Biological Psychology*, *84*, 463–473.

Sullivan, M. W., & Lewis, M. (2003). Contextual determinants of anger and other negative expressions in young infants. *Developmental Psychology*, *39*, 693–705.

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, *11*, 843–850.

Tassinary, L. G., & Cacioppo, J. T. (1992). Unobservable facial actions and emotion. *Psychological Science*, *3*, 28–33.

Tinbergen, N. (1953). *The herring gull's world*. London, England: Collins.

Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton, NJ: Princeton University Press.

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, *11*, 375–424.

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *The handbook of emotions* (3rd ed., pp. 114–137). New York, NY: Guilford Press.

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences, USA*, *105*, 11655–11660.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*, 397–405.

Tracy, J. L., & Robins, R. W. (2008). The nonverbal expression of pride: Evidence for cross-cultural recognition. *Journal of Personality and Social Psychology*, *94*, 516–530.

Turati, C. (2004). Why faces are not special to newborns: An alternative account of the face preference. *Current Directions in Psychological Science*, *13*, 5–8.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Vaillant-Molina, M., & Bahrick, L. E. (2012). The role of intersensory redundancy in the emergence of social referencing in 51/2-month-old infants. *Developmental Psychology*, *48*, 1–9.

Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science*, *7*, 261–269.

Valente, D., Theurel, A., & Gentaz, E. (2018). The role of visual experience in the production of emotional facial expressions by blind people: A review. *Psychonomic Bulletin & Review*, *25*, 483–497. doi:10.3758/s13423-017-1338-0

Valentine, E. (Writer), & Lehmann, B. (Producer). (2015, January 13). Episode 4517 [Television series episode]. In *Sesame street*. New York, NY: Sesame Workshop. https://www.youtube.com/watch?v=ZxfJicfyCdg

Vallacher, R. R., & Wegner, D. M. (1987). What do people think their doing? Action identification and human behavior. *Psychological Review*, *94*, 3–15.

Valstar, M., Zafeiriou, S., & Pantic, M. (2017). Facial actions as social signals. In J. K. Burgoon, N. Magnenat-Thalmann, M. Pantic, & A. Vinciarelli (Eds.), *Social signal processing* (pp. 123–154). Cambridge, England: Cambridge University Press.

Vick, S.-J., Waller, B. M., Parr, L. A., Smith Pasqualini, M., & Bard, K. A. (2007). A cross-species comparison of facial morphology and movement in humans and chimpanzees using FACS. *Journal of Nonverbal Behavior*, *31*, 1–20.

Viviani, P., Binda, P., & Borsato, T. (2007). Categorical perception of newly learned faces. *Visual Cognition*, *15*, 420–467.

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLOS Computational Biology*, *11*(4), Article e100s4066. doi:10.1371/journal.pcbi.1004066

Walker-Andrews, A. S. (2005). Perceiving social affordances: The development of emotion understanding. In B. D. Horner & C. S. Tamis LeMonda (Eds.), *The development of social cognition and communication* (pp. 93–116). New York, NY: Psychology Press.

Walker-Andrews, A. S., & Dickson, L. R. (1997). Infants' understanding of affect. In S. Hala (Ed.), *Studies in developmental psychology: The development of social cognition* (pp. 161–186). Hove, England: Psychology Press.

Walle, E. A., & Campos, J. J. (2014). The development of infant detection of inauthentic emotion. *Emotion*, *14*, 488–503.

Wang, N., Marsella, S., & Hawkins, T. (2008). Individual differences in expressive response: A challenge for ECA design. In *AAMAS '08 Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems* (Vol. 3, pp. 1289–1292). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Wang, X., Peelen, M. V., Han, Z., He, C., Caramazza, A., & Bi, Y. (2015). How visual is the visual cortex? Comparing connectional and functional fingerprints between congenitally blind and sighted individuals. *Journal of Neuroscience*, *35*, 12545–12559. doi:10.1523/JNEUROSCI.3914-14.2015

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, *78*, 105–119.

Weiss, B., & Nurcombe, B. (1992). Age, clinical severity, and the differentiation of depressive psychopathology: A test of the orthogenetic hypothesis. *Development and Psychopathology*, *4*, 113–124.

Widen, S. C., & Russell, J. A. (2013). Children's recognition of disgust in others. *Psychological Bulletin*, *139*, 271–299.

Wierzbicka, A. (1986). Human emotions: Universal or culture-specific? *American Anthropologist*, *88*, 584–594.

Wierzbicka, A. (2014). *Imprisoned in English: The hazards of English as a default language*. Oxford, England: Oxford University Press.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*, 1325–1331.

Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. *Social Psychological & Personality Science*, *7*, 331–338. doi:10.1177/1948550615624142

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*, 387–402.

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Neural evidence that human emotions share core affective properties. *Psychological Science*, *24*, 947–956. doi:10.1177/0956797612464242.

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2015). Variety in emotional life: Within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Social Cognitive and Affective Neuroscience*, *10*, 62–71.

Witherington, D. C., Campos, J. J., Harriger, J. A., Bryan, C., & Margett, T. E. (2010). Emotion and its development in infancy. In J. G. Bremner & T. D. Wachs (Eds.), *Wiley-Blackwell handbook of infant development* (Vol. 1, 2nd ed., pp. 568–591). Chichester, England: Wiley-Blackwell.

Witherington, D. C., Campos, J. J., & Hertenstein, M. J. (2001). Principles of emotion and its development in infancy. In *Blackwell handbook of infant development* (1st ed., pp. 427–464). Oxford, England: Blackwell.

Wolff, P. H. (1987). *The development of behavioral states and the expression of emotions in early infancy: New proposals for investigation*. Chicago, IL: University of Chicago Press.

Wood, A., Rychlowska, M., & Niedenthal, P. (2016). Heterogeneity of long-history migration predictions emotion recognition accuracy. *Emotion*, *16*, 413–420.

Yan, F., Dai, S. Y., Akther, N., Kuno, A., Yanagihara, T., & Hata, T. (2006). Four-dimensional sonographic assessment of fetal facial expression early in the third trimester. *International Journal of Gynecology & Obstetrics*, *94*, 108–113.

Yik, M., & Russell, J. A. (1999). Interpretation of faces: A cross-cultural study of a prediction from Fridlund's theory. *Cognition & Emotion*, *13*, 93–104.

Yin, Y., Nabian, M., Ostadabbas, S., Fan, M., Chou, C., & Gendron, M. (2018). Facial expression and peripheral physiology fusion to decode individualized affective experiences. *ArXive.org*. Retrieved from https://arxiv.org/abs/1811.07392v1

Yitzhak, N., Giladi, N., Gurevich, T., Messinger, D. S., Prince, E. B., Martin, K., & Aviezer, H. (2017). Gently does it: Humans outperform a software classifier in recognizing subtle, nonstereotypical facial expressions. *Emotion*, *8*, 1187–1198.

Young, A. W., & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Sciences*, *22*, 100–110.

Yu, H., Garrod, O. G. B., & Schyns, P. G. (2012). Perception-driven facial expression synthesis. *Computers & Graphics*, *36*, 152–162.

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Boulder, CO: Westview Press.

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science*, *26*, 237–242.

Zhang, Q., Chen, L., & Yang, Q. (2018). The effects of facial features on judicial decision making. *Advances in Psychological Science*, *26*, 698–709.

Zoll, C., Enz, S. H. S., Aylett, R., & Paiva, A. (2006, April). Fighting bullying with the help of autonomous agents in a virtual school environment. Paper presented at the *7th International Conference on Cognitive Modelling (ICCM-06)*, Trieste, Italy. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.1956&rep=rep1&type=pdf