Supplementary On-Line Materials

Box 1: Google Search Evidence Demonstrating the Existence of a Common View of Emotional Expressions

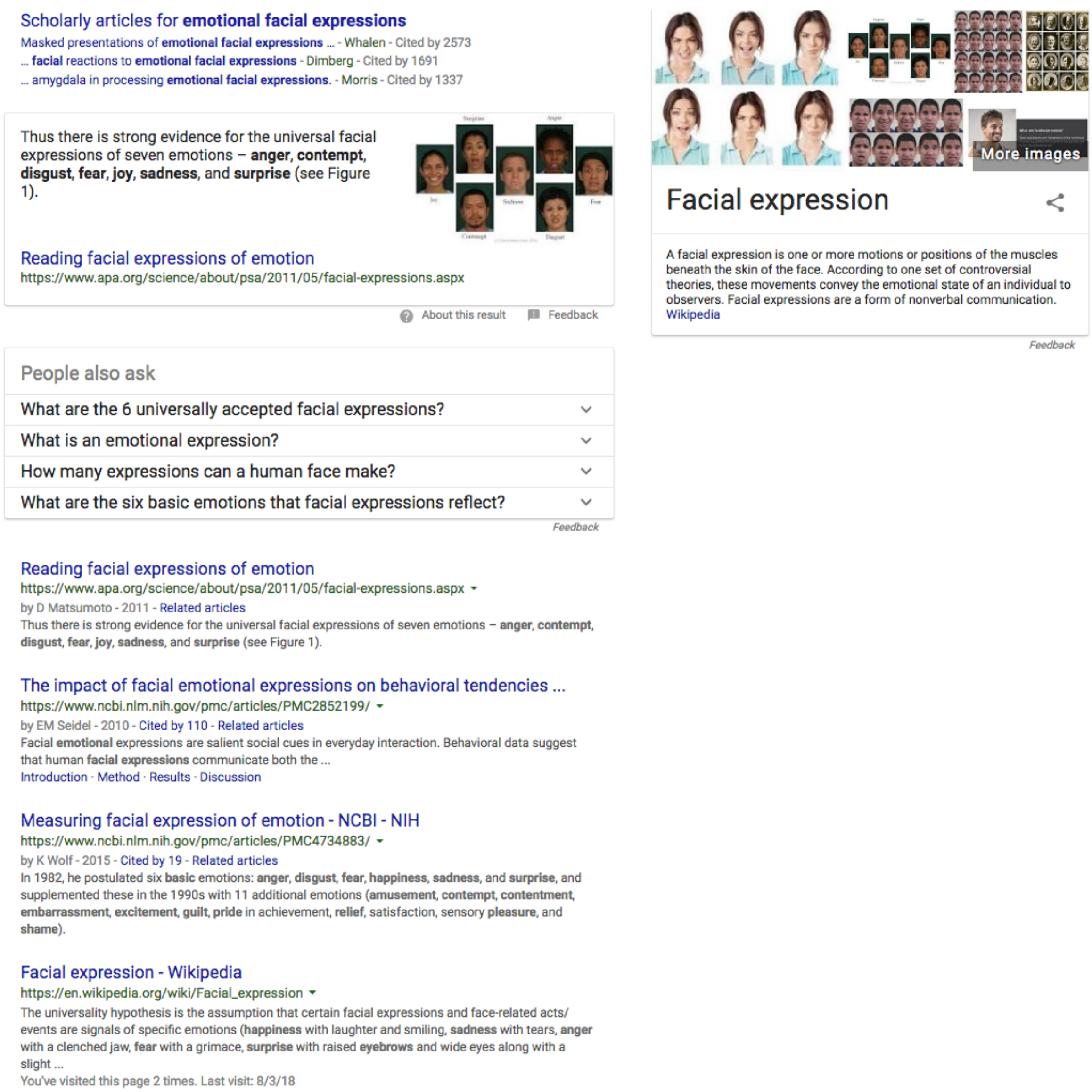
This box provides perhaps the simplest characterization of the "common view" of emotional expressions: how it is represented in the internet. It is important to note that we are not suggesting that the common view is necessarily the view that individual scientists personally hold, nor the view that all laypeople personally hold. Instead, it is the view that most laypeople think science supports, and the view that many scientists, intentionally or unintentionally, have perpetuated in the literature. It is the modal view that computer scientists, educators, and psychiatrists draw from, in the assumption that it is the view currently best supported by scientific evidence; and this is the assumption we evaluate in our paper.

The items retrieved from a Google search represent the consensus of our culture in many ways, and certainly represent what most people take as a summary of the conclusions best supported by current science, broadly conceived. Since different search histories and browsers may produce somewhat different results, we asked 23 students and post-doctoral fellows to search the term "emotional facial expressions," and send us a screenshot of their first page of results.

This search produced web pages very similar to the small reproduction below. In all pages that included images (21/23; like the one at the upper right in the example), the first and most prominent image shows the 6 "basic" facial expressions of emotion; in several this was the only image produced. In all searches that included the box "People also ask" (7/23), the first item listed was "What are the 6 universally accepted facial expressions?"

In 22/23 searches, the very first hit, titled "Reading facial expressions of emotion," brings up an APA article (https://www.apa.org/science/about/psa/2011/05/facial-expressions.aspx) that leads with a section on the universality of facial expressions; Figure 1 in that article is a picture of the facial configurations that are hypothesized to express anger, disgust, fear, happiness, sadness, surprise and contempt and the APA article concludes with "Because facial expressions of emotion are part of our evolutionary history and are a biologically innate ability, we all have the ability to read them." The second common hit (15/23; this and subsequent hits get truncated because not all searches had screens that permitted a long list) is to a website (https://thoughtcatalog.com/january-nelson/2018/06/list-of-emotions/) that lists the 6 basic emotions together with their facial expressions (the source of the most common image retrieved by the search). The next most common hit in the list (14/23, usually third or fourth entry) is a 2018 news from Science Daily titled, "How emotions in facial expressions are understood." It mentions by name only the following emotions: fear, happiness, surprise, anger, sadness. It covers a research paper that showed participants "images of faces expressing the six emotions and one neutral expression". The next most common hit is typically a research paper; in our sample it was https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5152920/ (10/23 hits) which is an article investigating emotion perception from faces, using as stimuli only photos of the six basic emotions. Another common research article here was https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2852199/, whose very first sentence in the abstract is, "Emotional faces communicate both the emotional state and behavioral intentions of an individual."

These searches are complemented by the Wikipedia entry for "facial expressions," as well as by later search hits that paint a more nuanced picture and nearly always explicitly note that there is considerable controversy about how reliably we can judge emotion from faces, what categories of emotional expression there are, and whether emotional facial expressions are culturally universal. Our conclusion from this google search is twofold. First, there is overwhelming evidence that the modal view that is available on the internet with a casual search represents the common view: that specific facial expressions correspond to specific emotions (usually about 6), and that these can be reliably perceived. Second, closer scrutiny notes that the common view has strong critiques. These two facts motivated our paper and the aim to produce the most thorough review of the evidence.



Box Figure 1-1. Screenshot of the first page of a google search for “emotional facial expressions.”

Box 2: Theoretical Frameworks and Their Relation to Emotion Categories

An instance of emotion can be described by a variety of features: *physical features* (such as patterns of expressive facial movements, vocal acoustics, autonomic nervous system changes, and neural activity), *affective features* that capture what the instance feels like (e.g., how pleasant or unpleasant the episode feels, how arousing it feels; Barrett & Bliss-Moreau, 2009; Russell & Barrett, 1999), *appraisal features* that refer to how the situation is experienced (e.g., whether the situation is experienced as novel or familiar, as conducive to one’s immediate goals or not, and so on; Barrett, Mesquita et al., 2007; Clore & Ortony, 2008, 2013; Scherer et al., 2017) and *functional features* that refer to the goals that a person is attempting to meet (e.g., to avoid a predator, to get closer to someone, to win a competition, etc.; e.g., Adolphs, 2017; Lazarus, 1993). An emotion category is a grouping of emotional episodes that share some feature or set of features in common.

Many debates about the nature of emotion boil down to disagreements about *the nature of the similarities* shared by instances of the same emotion category and the *degree of variation* in the relevant features, as well as potential similarity and differences in features across emotion categories. These debates can be summarized by two dimensions, as depicted in Figure 1: The horizontal dimension represents hypotheses about the surface similarities across instances of the same emotion category, such as being expressed by similar patterns of facial movements. And the vertical dimension represents hypotheses about the deep similarities across instances of the same emotion category, such as sharing the same neural circuitry.

Similarities and Variation in Surface Features

In this paper, we are primarily concerned with the horizontal dimension representing the similarity and variation in surface features, such as the facial movements that express instances of emotion. At one end of the "surface" continuum, basic emotion approaches (e.g., Ekman, 1972; Tracy & Randles, 2011) and the discrete emotion approach (e.g., Izard, 2007, 2010; Izard, Woodburn & Finlon, 2010), in their original form, propose that anger, sadness, fear, disgust, happiness and surprise are each spontaneously expressed with universal configurations of unique facial movements (as depicted in Figure 1). These approaches acknowledge some variation in the facial movements that express the instances of each category, but they are hypothesized to result from processes that are external to the processes that caused the emotional instance, such as display rules, cultural learning, emotion regulation, and so on. Therefore, instances within the same emotion category would be expected to be relatively similar in their expressions (high reliability) across contexts, people and time, and instances of different emotion categories are expected to be relatively unique in their expressions (high specificity). Updated basic emotion theory allows for more variation in expressions within a category, assuming that instances of an emotion category vary around one supposed "best instance" (the prototype) that is either most typical or most frequent in nature.[43] Nonetheless, each instance of an emotion category is hypothesized to share enough of a characteristic facial expression that is *consistently present and recognizably different* from the patterns found in other emotion categories (for specific quotations, see Ekman, 1992, p. 550; Ekman & Cordaro, 2001, p. 364; Levenson, 2011, p. 379; Scarantino & Griffiths, pp. 448-449).

Intermediate along the horizontal dimension is the component process model of emotion which proposes that sequences of evaluations, referred to as appraisal checks, drive the dynamics of facial movements that express emotion (Scherer, Mortillaro & Mehu, 2017; Scherer et al., 2018). This approach belongs to a collection of causal appraisal approaches, which characterize appraisals as more than descriptive features of an emotional instance – they are considered to be actual causal cognitive mechanisms that produce instances of emotion, including how a situation is experienced and the associated facial expressions as components of the instance. Causal appraisal approaches acknowledge the possibility that, in principle, different temporally extended patterns of appraisals can cause instances of the same emotion category, and therefore is consistent with a larger variety of facial configurations that can express the emotion (for discussion, see Barrett, 2017c; Gross & Barrett, 2011; Scherer et al., 2017). In practice, however, this possible variation remains largely unexplored in scientific experiments.

At the other end of the continuum are approaches that explicitly predict substantial, situated variation within categories and similarity between categories, e.g., the theory of constructed emotion (Barrett, 2012, 2013, 2017a, b); core affect theory (Russell, 2003); functional approaches to emotion (relevant references, see Adolphs, 2017; Anderson & Adolphs, 2018; Campos et al., 1994); and descriptive appraisal approaches (e.g., Ortony & Clore, 2013; for a discussion, see Barrett, 2017c; Gross & Barrett, 2011). That is, they predict, in advance, that instances of the same emotion category are highly variable in their expressions (a specific set of facial movements have low reliability as expressions across people, situations and cultures) and instances of different emotion categories or even non-emotion categories are similar in their expressions (a given set of facial movements have low specificity as expressions of a single emotion category). For example, the theory of constructed emotion proposes that facial expressions of emotion are intrinsically constructed in a context-dependent way that has been learned in a particular culture.

Similarities and Differences in Deep Features

At the top of the "deep" continuum, basic emotion approaches propose that anger, sadness, fear, disgust, happiness and surprise are each caused by their own set of dedicated neural circuits (Tracy & Randles, 2011), a hypothetical affect program (Ekman & Cordaro, 2011), or a set of computations (Bach & Dayan, 2017). Some approaches allow for considerable variation in the physical causes of emotion and instead propose that each category shares a universal, functional similarity across situations (e.g., fear is the desire to escape from a predator; Adolphs, 2017; Anderson & Adolphs, 2018; Campos et al., 1994). At the bottom end of the continuum, theoretical approaches hypothesize that an emotion category does not exist in the brain as a fixed neural circuit, a fixed computation, or a fixed function. Instead, the human brain is thought to construct an emotion category on the fly, as needed for a situation-specific goal, with the help of the emotion concepts that person has acquired using the language they speak (Barrett, 2017a, b). This type of ad hoc category is called a goal-based category, because the similarity of its instances is based on the goal that the instances serve in a particular situation at a particular moment in time (Barsalou, 1983, 1985; Barsalou et al., 2003).

Importantly, most of these approaches explicitly anchor their hypotheses in evolutionary considerations, so it is misleading to refer to any one approach as an "evolutionary" approach. Most assume that emotion categories are psychological as well as biological categories. And both draw inspiration from Charles Darwin (albeit from different books, making very different assumptions about the nature of biological categories; for discussion, see Barrett 2017a, b).

Emotion Concepts

An emotion concept is a mental representation of an emotion category. Theoretical hypotheses about the proposed degree of variation in surface and deep features of an emotion category strongly relate to the proposed nature of emotion concepts. Approaches that hypothesize strong reliability and specificity in emotional expressions and other physical features shared by the instances of an emotion category, along with a deep causal similarity, propose that emotion categories are natural kind or *Aristotelian* categories (the red zone in Figure 1); correspondingly, instances of different categories can be distinguished by their physical features, like the facial movements expressing emotion. If an emotion category has a classical structure, then its corresponding concept reads like a dictionary definition that is stored in memory, describing its necessary and sufficient features.

A variety of theoretical approaches propose that emotion categories are prototype categories (the yellow zone in Figure 1), whose instances share some family resemblance. Of all the features that might describe the category, each instance might contain only a sample (resulting is more within-category variation and more between-category similarity than is true for Aristotelian categories). The corresponding emotion concept (its prototype) might be the most frequent instance found in the category, or its most typical instance (i.e., it is the instance that has all or most of the category's distinguishing features). Or the prototype might be a theory that describes the most typical instance (e.g., Clore & Ortony, 1991). The hypothesis that emotion categories are structured as prototypes is consistent with a variety of theoretical approaches in the science of emotion, including basic emotion approaches (e.g., Cowen & Keltner, 2017; Ekman & Cordaro, 2011), appraisal approaches (e.g., Shaver et al., 1987) psychological construction approach (e.g., Russell, 2003) and functional approaches (e.g., Campos et al., 1994; Campos, Campos & Barrett, 1989). The assumption, however, is that there is a single representation – a single prototype – for each category.

More recently, it has been proposed that emotion categories are goal-based, conceptual categories (green zone, Figure 1). The instances of a given emotion category are thought to share a common set of features *within a specific situation*, but these features (including the goal or function of the category) will change from situation to situation (Barrett, 2006, 2012, 2013, 2017a, b; Barrett, Wilson-Mendenhall, & Barsalou, 2015; LeBois et al., 2018; Wilson-Mendenhall et al., 2011). The hypothesis is that emotion categories, like all abstract categories, do not have conceptual cores (Barrett, Wilson-Mendenhall, & Barsalou, 2015; Wilson-Mendenhall et al., 2011), meaning that emotion concepts are constructed on the spot, as needed (i.e., they are ad hoc concepts).

Unlike the typological thinking that supports classical and prototype categories, conceptual categories are rooted in population thinking, or the idea that a biological category is populated with context-dependent, variable instances, so that any summary of the category (like a prototype) is an abstraction. The proposal that emotion categories are goal-based, conceptual categories, derives from Darwin's use of population thinking in On the Origin of Species (1859/2001; see Mayr, 2004), as well as Barsalou's research on grounded concepts (Barsalou, 1983, 1985, 2008; Barsalou et al., 2003), where the prototype of a category is context-dependent, and represents the ideal instance that best suits the function or goal in a specific situation, whether or not it actually exists in nature (e.g., Barsalou, 1993; Voorspoels, Vanpaemel, & Storms, 2011). Correspondingly, the hypothesis is that the similarity in an emotion category is not fixed or static – it varies from situation to situation because the similarity of its instances is based on the goal that the instances serve in a particular situation at a particular moment in time.

References

Adolphs, R. (2017). How should neuroscience study emotions? by distinguishing emotion states, concepts, and experiences. *Soc Cogn Affect Neurosci, 12*(1), 24-31. doi:10.1093/scan/nsw153

Anderson & Adolphs, 2018;

Bach, D. R., & Dayan, P. (2017). Algorithms for survival: a comparative perspective on emotions. *Nature Reviews Neuroscience, 18*, 311. doi:10.1038/nrn.2017.35

Barrett, L.F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*, 20-46.

Barrett, L. F. (2012). Emotions are real. *Emotion, 12*, 413-429

Barrett, L. F. (2013). Psychological construction: A Darwinian approach to the science of emotion. *Emotion Review, 5*, 379-389.

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

Barrett, L. F. (2017b). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12,* 1-23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F. (2017c). Functionalism cannot save the classical view of emotion. *Social Cognitive and Affective Neuroscience*, *12*, 34-36. https://doi.org/10.1093/scan/nsw156

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in Experimental Social Psychology*, *41*, 167-218.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology, 58*, 373-403. PMC1934613

Barrett, L. F., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). The conceptual act theory: A road map. Chapter in L. F. Barrett and J. A. Russell (Eds.), *The psychological construction of emotion* (p. 83-110). New York: Guilford.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11*(3), 211-227.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *J Exp Psychol Learn Mem Cogn, 11*(4), 629-654.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences, 7(*2), 84-91.

Campos, J. J., Campos, R. G., & Barrett, K. C. (1989). Emergent themes in the study of emotional development and emotion regulation. *Developmental Psychology, 25*(3), 394.

Campos, J. J., Mumme, D., Kermoian, R., & Campos, R. G. (1994). A functionalist perspective on the nature of emotion. *Japanese Journal of Research on Emotions, 2*(1), 1-20.

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 GL, Ortony A. (2013). Psychological Construction in the OCC Model of Emotion. *Emotion Review,* 5(4):335-343. doi:10.1177/1754073913489751.

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.

Cowen, A. S. & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, www.pnas.org/cgi/doi/10.1073/pnas.1702247114
Darwin's

Darwin, C. (1859/1964). *On the Origin of Species*. (A facsimile of the first edition. ed.). Cambridge, MA.: Harvard University Press.

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, Nebraska: University of Nebraska Press.

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

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

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. PMC3072688

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

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review, 2(4),* 363-370.

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

Lazarus, R. (1991). *Emotion and Adaptation*. NY, Oxford University Press.

Lebois, L. A. M., Wilson-Mendenhall, C. D., Simmons, W. K., Barrett, L. F., & Barsalou, L.W. (2018). Learning situated emotions. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2018.01.008

Levenson, R. W. (2011). Basic emotion questions. *Emotion Review, 3(4)*, 379-386.

Mayr, E. (2004). *What makes biology unique? Considerations on the autonomy of a scientific discipline*. New York: Cambridge University Press.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110(1)*, 145-172.

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.

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review, 3(4)*, 444-454.

Scherer, K. R., Mortillaro, M., & Mehu, M. (2017). Facial expression is driven by appraisal and generates appraisal inferences. In J-M. Fernandez-Dols & J. A. Russell (Eds.), *The Science of Facial Expression* (p. 353-374). New York: Oxford U Press.

Scherer, K. R., Mortillaro, M., Rotondi, I., Sergi, I., & Trznadel, S. (2018). Appraisal-driven facial actions as building blocks for emotion inference. *Journal of Personality and Social Psychology, 114*(3), 358-379.

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.

Tracy, J., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3(4)*, 397-405.

Voorspoels, W., Vanpaemel, W., & Storms, G. (2011). A formal idea-based account of typicality. *Psychon Bull Rev, 18,* 1006-1014.

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia, 49,* 1105-1127.

[43] The word "theory" is used here in its everyday definition to mean "a group of ideas" rather than its scientific definition to mean "a comprehensive explanation of some aspect of nature that is supported by a vast body of evidence" (National Academies of Science, http://www.nas.edu/evolution/TheoryOrFact.html).

Box 3: The Current Study of Context Effects in Emotion Perception

There is growing evidence that both the facial movements that express emotions and the emotional meaning inferred for those movements are strongly influenced by the contexts in which they occur. In the scientific study of how people infer emotions in the facial movements of other people, three types of context have been increasingly studied: situationally-based context, in which a face is physically presented with other sensory input that has some informational value, such as a body posture or a tone of voice, person-based context, in which the processes inside the brain or body the person inferring emotion in the facial movements of other people, and broader cultural contexts (for summaries, see Aviezer & Hassin, 2017; Barrett, Mesquita, & Gendron, 2011; de Gelder, 2016; Gendron, Mesquita & Barrett, 2013; Hess & Hareli, 2017; Wieser & Brosch, 2012).[1] In the majority of studies, other cues dominate emotion-related inferences, such that the emotional meaning of face is interpreted in line with its context, rather than its hypothesized emotional meaning as portrayed in Figure 4 (e.g., Aviezer et al., 2008, 2012; van den Stock et al., 2007; Wallbott, 1988; Wood, Martin, Alibali, & Niedenthal, 2018). Studies of how the context influence the activation of facial muscle movements are less frequent, but still exist (but for several clever examples, see Fernandez-Dols & Ruiz-Belda, 1997; Fridlund, 1991; Ruiz-Belda et al., 2003). Nonetheless, when context is properly acknowledged and assessed, scientific findings run contrary to the notion that facial muscle movements are universal expressions of emotion containing all of the information that is necessary and sufficient to communicate emotional states. They are consistent with functionalist, constructionist and behavioral ecology theories of emotion (Box 2). Contextual influences here are consistent with

[1] Gendron et al. (2013) was first written and submitted for publication in 2010.

evidence showing that context is intrinsically involved in the most basic aspects of movement and object perception.

The classic demonstration of how context impacts the interpretation of facial movements is the Kuleshov effect. In the early 20th century, the Soviet filmmaker Lev Kuleshov demonstrated that that the emotional meaning of an actor's neutral face changes depending on what was viewed immediately before: videos designed to evoke pleasantness (a little girl playing with a doll) or lust (a woman on a divan), unpleasantness (a dead woman in a coffin), and hungry (a bowl of soup) (Barratt et al., 2016; Mobbs et al., 2006; see Calbi et al., 2017 for a recent study as well as older references). These findings are echoed in more recent studies by the US psychologist Jim Russell and may explain the recent emergence of "resting bitch face" and "backpfeifengesicht" (literally, a face in need of a punch) in social media.

When scientists ask participants to pose an emotional expression in the absence of context (or in a singular, impoverished; e.g., Cordaro et al., 2017) or to infer emotion in a facial configuration that is absent any context except that provided by the experimental task, they are typically taking a reductionist approach to discover the emotional meaning of facial movements alone. Participants may instead be communicating a culturally established expectation, stereotype, or meme that may or may not hold in everyday contexts that are richer in the multimodal information that they carry (Srinivasan & Martinez, 2018). We focus our attention on one another's faces when we communicate, but this is not evidence that emotional information is carried solely, or even primarily, in the face alone (Aviezer et al., 2012). In real life, faces don't appear in isolation. Instead, they appear in a multi-sensory context that includes a body, a broader situational arrangement, and often a voice, smells, and so on. These additional sources of information are only considered "context" when the starting assumption is that the face is primary in communication emotion. In our view, this "context" in all its forms must be explicitly measured and modeled to achieve an ethology of emotional communication in the wild (Martinez, 2017b).

References

Aviezer, H., & Hassin, R. R. (2017). Inherently ambiguous: An argument for contextualized emotion perception. In J-M. Fernandez-Dols & J. A. Russell (Eds.), *The science of facial expression* (p. 333-349). NY: Oxford U Press.

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*(7), 724-732.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science, 338*(6111), 1225-1229. doi:10.1126/science.1224313

Barratt, D., Rédei, A. C., Innes-Ker, A., & van de Weijer, J. (2016). Does the Kuleshov effect really exist? Revisiting a classic film experiment on facial expressions and emotional contexts. *Perception, 45*, 847–874.

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

Calbi, M., Heimann, K., Barratt, D., Siri, F., Umilta, M. A., & Gallese, V. (2017). How Context Influences Our Perception of Emotional Faces: A Behavioral Study on the Kuleshov Effect. *Front Psychol, 8*, 1684. doi:10.3389/fpsyg.2017.01684

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2017). Universals and Cultural Variations in 22 Emotional Expressions Across Five Cultures. *Emotion*. doi:10.1037/emo0000302

de Gelder, B. (2016). *Emotions and the body*. New York: Oxford University Press.

Fernandez-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.

Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology, 60(2),* 229-240.

Gendron, M., Mesquita, B., & Barrett, L. F. (2013). Emotion perception: Putting a face in context. Chapter in D. Reisberg (Ed.), *Oxford Handbook of Cognitive Psychology* (p. 539-556). New York: Oxford University Press.

Hess, U., & Hareli, S. (2017). The social signal value of emotions: The role of contextual factors in social inferences drawn from emotion displays. In J-M. Fernandez-Dols & J. A. Russell (Eds.), The science of facial expression (p. 375-392). NY: Oxford U Press

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

Mobbs, D., Weiskopf, N., Lau, H., Featherstone, E., Dolan, R., & Frith, C. (2006). The Kuleshov Effect: The influence of contextual framing on emotional attributions. *Social Cognitive and Affective Neuroscience, 1*, 95–106.

Ruiz-Belda, M. A., Fernandez-Dols, J-M., Carrera, P., & Barchard, K. (2003). Spontaneous facial expressions of happy bowlers and soccer fans. *Cognition and Emotion, 17*, 315-326.

Srinivasan, R., and 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. ArXiv preprint, available at https://arxiv.org/pdf/1808.04399.pdf

Van den Stock, J., Righart, R., & De Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion, 7(3),* 487.

Wallbott, H. G. (1988). In and out of context: Influences of facial expression and context information on emotion attributions. British Journal of Social Psychology, 27, 357-369.

Wieser, M. J., & Brosch, T. (2012). Faces in context: a review and systematization of contextual influences on affective face processing. *Front Psychol*, *3*, 471. doi:10.3389/fpsyg.2012.00471

Wood, A., Martin, J. D., Alibali, M. W., & Niedenthal, P. M. (2018). A sad thumbs up: incongruent gestures and disrupted sensorimotor activity both slow processing of facial expressions. *Cognition and Emotion*, 1-14.

Box 4: The Origin of the Proposed Facial Expressions

The belief that certain configurations of facial movements express emotion and therefore display a person's inner state of mind reached prominence in the paintings by Rembrandt in the 17th century. Different sets of universal expressive forms were proposed by various artists, such as in the drawings of Charles Le Brun, the 17th century French painter for King Louis XIV of France (Montagu, 1994), in the facial movements found in Hindu Navarasa dance (https://www.youtube.com/watch?v=1uIan6u3UrQ), in the heads depicted by the 18th century German sculptor Franz Xaver Messerschmidt (https://www.pinterest.com/tkearneyg/franz-xaver-messerschmidt/), in moving photographs by the 19th century Czech neuroscientist Jan Evangelista Purkinje (Wade, 2016), in the prescriptions of the 19th century French dramatist François Delsarte (Stebbins & Delsarte, 1887/2013), and in the photographs taken by the French physician Guillaume-Benjamin-Amand Duchenne, which were highlighted in the writings of Charles Darwin (Darwin, 1872). These varying proposals were united by the assumption that each emotion category was expressed by one specific facial configuration, making it possible to infer one (emotional state) from the other (the configuration of facial movements).

How did the science of emotion end up focusing almost exclusively on the facial configurations displayed in Figure 4? The story of their origin has been discussed in Ekman, Friesen & Ellsworth (1972), Gendron & Barrett (2017) and Russell (1994). Darwin did not discover the facial configurations through careful observation in the same way that such observations led him to discover the idea of natural selection. Instead, he stipulated them based on drawing by Bell (1806) and photographs by Duchenne (1990/1862), continuing the tradition of others who, before him, stipulated other facial movements as the expressions of emotions (see introduction in main paper).

Darwin conducted two informal studies of emotional expressions. In *The Expression of the Emotions in Man and Animals*, he described an informal survey about these facial movements that he believed expressed emotions in a way that is shared with other animal species and therefore in a specific and universal way around the world. He provided his colleagues with verbal descriptions of specific expressive forms along with the emotion category he thought they expressed, and he asked his colleagues (living in various parts of the world) whether they believed his hypotheses were true. Darwin also conducted a second informal study that he described in a letter to a colleague, in which he presented 24 participants with 11 static photographs of facial configurations elicited by electrical stimulation of facial muscles (detailed in Snyder, Kaufman, Harrison, & Maruff, 2010). These photographs were taken by Duchenne, who believed that facial muscles produce expressions that reveal a person's inner state. Duchenne created over 60 photographs of facial configurations that are often referred to as "induced emotional expressions," but actually the photographs capture exaggerated facial muscle contractions elicited by external electrical stimulation. Darwin's description of the study suggests he asked his participants to freely describe the emotion presenting in each photo (i.e., he used a free-labeling response method).

Research that proceeded in the early 20th century attempted to replicate and extend Darwin's second study using photographs of faces posed in exaggerated configurations. Others went about the task of detailing the specific facial movements that constitute the configurations that were believed to be emotional expressions. Few research studies evaluated whether emotional expressions made in everyday life actually conform to these portrayals. The emotion perception studies during this period generally showed that perceivers were highly variable in the in the emotional causes they inferred for each configuration, with little reliability and specificity.

(The one exception was studies using a choice-from-array approach involving brief vignettes describing each emotion category, referred to as the Dashiell method). This larger body of research gave rise to a scientific era that was guided by the hypothesis emotions and their expressions were socially constructed and culturally variable.

In the early 1960's the U.S. psychologist Paul Ekman and his colleagues resurrected a research paradigm to test Darwin's original ideas about emotional expressions. They took a large set of photographs developed by Sylvan Tomkins, in which actors posed what they believed to be the expression of anger, contempt, disgust, fear, happiness, interest, sadness, shame and surprise categories. Participants viewed the candidate poses and based on their reliability in choosing the expected emotion word to label each photo, a final set was chosen. The focus on six categories -- anger, disgust, fear, happiness, sadness, and surprise was accidental, not based on theoretical considerations – portrayals of contempt, interest and shame expressions were not labeled reliably and so were not initially studied. Ekman and colleagues also incorporated work by the Swedish anatomist Carl-Herman Hjortsjö, who catalogued Duchenne's facial configurations and stipulated them to be emotional expressions (Hjortsjö, 1969).

This is the origin of the facial configurations in Figure 4, which are used in the majority of experiments of emotion perception and that correspondingly constitute commonsense beliefs about how certain emotion categories are specifically expressed. Many different sets of facial poses haven been developed over the years, a selection of which can be found here:

https://rystoli.github.io/FSTC.html

http://cbcsl.ece.ohio-state.edu/downloads.html

http://cseweb.ucsd.edu/~gary/CAFE/

https://ieeexplore.ieee.org/document/8013713/

References

Bell C. (1806). *Essays on the Anatomy of Expression in Painting*. London: Longman, Hurst,Rees & Orme.

Darwin, C. (1872/1965). *The Expression of the Emotions in Man and Animals* (Chicago: University of Chicago Press).

Duchenne CB (1990/1862). *The Mechanism of Human Facial Expression* (Renard, Paris); reprinted (1990) (Cambridge Univ Press, London).

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

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

Hjortsjö, C. H. (1969). *Man's face and mimic language*. Studen litteratur.

Montagu, J. (1994). *The expression of the passions: The origin and influence of Charles Le Brun's lecture on general and particular expression*. London: Yale University Press.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin, 115(1)*, 102-141.

Snyder, P. J., Kaufman, R., Harrison, J., & Maruff, P. (2010). Charles Darwin's emotional expression "experiment" and his contribution to modern neuropharmacology. *Journal of the History of the Neurosciences*, 19:2, 158-170, doi: 10.1080/09647040903506679

Stebbins, G. & Delsarte, F. (1887/2013). Delsarte system of expression. E. S. Werner.

Wade, N. J. (2016). Faces and photography in 19th-century visual science. *Perception, 45*, 1008-1035.

Box 5: The Face: Anatomy of Facial Muscles and Facial Electromyography

The Anatomy of Facial Muscles

To understand how people make certain patterns of facial movements, it's helpful to consider how facial muscles develop and are structured. Healthy humans have a common set of 17 facial muscles on each side of the face that contract and relax in patterns (Rinn, 1984) [2]. Facial muscles develop in utero. By 36 weeks, almost all of the muscles used to produce facial movements are formed (Gasser, 1967), although their morphology is not identical to an adult human. The muscles are controlled by cranial nerves V and VII which become functional a bit earlier, by 11 weeks in utero (Reissland et al., 2011). Facial muscle movements support many functions, such as sucking movements necessary for feeding and tongue movements necessary for speech (Haywood & Getchell, 2014). As a consequence, the development of facial muscles influences a wide range of behaviors and capacities over and above expressing emotion.

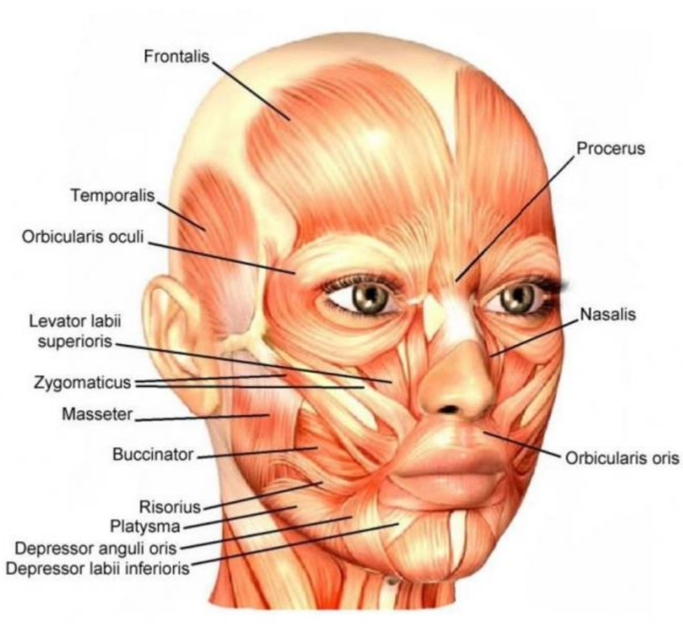
To create facial movements that are visible to the naked eye, facial muscles contract, moving skin into folds and wrinkles on an underlying skeletal structure. These are the movements people observe in one another, that are captured in photographs and videos, and that express emotions. These visible facial articulations are called Action Units (AUs). Clear and consistent data are currently not available to indicate when facial expressions are first formed. While fetuses in utero definitely make some facial movements, it is currently entirely unknown if these have anything to do with emotion-like states. Some scientists claim that young infants produce facial movements to pain that resemble those of adults (Izard et al., 1987), even though the underlying anatomy is not morphologically identical (Camras & Shutter, 2010).

---

[2] There are somewhat different numbers often given, because a single muscle may comprise more than one functional unit.

Individual differences in facial anatomy, as well as in the brain's control of facial muscles, cause variation in the details of how facial movements are executed at the muscular level and how they look to the naked eye. People vary in the underlying bone structure of the face and details of the skin, the structure and strength of their facial muscles (Pessa et al., 1988), the dynamics of facial muscle movements, and consequently they vary in how their facial movements look to a human observer (e.g., Farahvash et al., 2010; Shim et al., 2008; Shimada & Gasser, 1989). In addition, some people have strong asymmetries for one side of the face or the other, and some people lack certain smaller muscles altogether (Waller, 2008). In fact, if you inserted your exact facial muscles into a different face (someone with a different bone structure or someone much older or younger than you, or whose face is thinner or fatter), the resulting muscle movements would look different than they do on your face. And even when facial movements look the same to the naked eye, there may be differences in their execution under the skin. As human perceivers, we see stable facial behaviors (i.e., a frown) when in reality, under the skin, there is more variation that meets the eye. A facial behavior, like frowning, or scowling, is, in fact, a category of variable instances. When you watch a frown unfold in the same person on two different occasions, the exact muscle contractions that curl the upper lip and turn down the corners of the mouth can subtly vary from one instance to another. What to the naked eye looks like the same frown in two different people can result from different patterns of underlying muscle contractions.

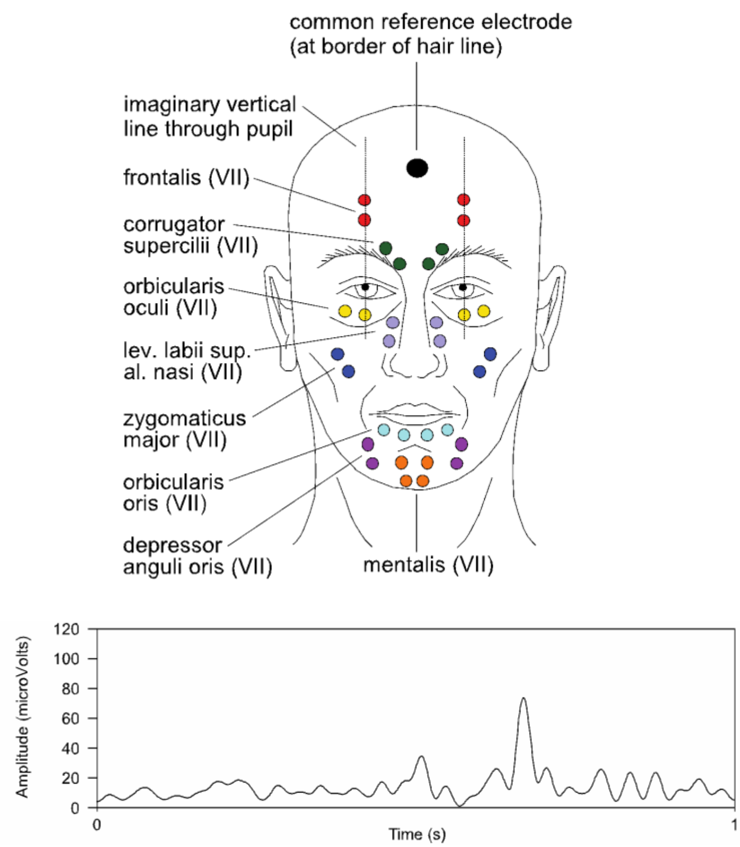
Put simply: something as seemingly simple as a single facial movement is best understood as a conceptual category, resulting from a variable set of more basic, variable physical changes. The same is true for all motor movements, for smells and for sounds. For example, the sound of a "b" is acoustically different when heard in the words "bad" and "bed", yet human brains wire themselves to hear both as the sound of "b" (i.e., the sound of a "b" is a category; Barsalou 1992).



Box Figure 5-1. The muscles of the face. From www.hdanatomy.com. Permission pending.

Measuring Facial Movements with Facial Electromyography (fEMG)

The most sensitive and direct way to measure facial muscle movements is to record the electrical activity of facial muscles as they contract. This is called facial electromyography, or facial EMG, and is done by placing electrodes in or on the muscles of the face. Actually inserting thin electrodes into the muscles provides the most sensitive and specific electrical recordings, but this approach is rarely used since it is more invasive. It is more common to use surface recordings from electrodes placed on the skin that measure the changes in electrical potential of muscle depolarization. EMG recordings of facial muscle movements were first made in the late 1950s and were later used for studying emotion and affect in the 1970s (Tassinary et al., 2007). The specificity with which individual muscles can be distinguished from one another depends on how many electrodes are put on the face. Typically, only three to six electrodes are put on the face, although many more than that can be uncomfortable for a test subject. This means that many facial muscle movements are not measured in most studies that use facial EMG. Current research suggests that specific patterns of facial EMG activity reliably distinguish between pleasant vs. unpleasant states, as well as the intensity of the states along with how social the situation is, but that they do not reliably distinguish between different individual categories of emotion (Cacioppo et al., 2000).



Box Figure 5-2. Measuring facial electromyography (fEMG). Left: Common electrode locations, showing some of the facial muscles whose activity they can measure. VII indicates that all the muscles are controlled by the facial nerve (7th cranial nerve). Right: example of a recorded EMG signal. The trace plots the absolute value of the change in electrical activity (in microvolts) versus time (in seconds). Reproduced with permission from van Boxtel (2010). See Figure 4 for a summary of how combinations of these muscles contribute to facial action units in FACS coding.

References

Barsalou, L.W. (1992). Frames, concepts, and conceptual fields. In: Frames, Fields, and Contrasts: new essays in semantic and lexical organization, pp. 21-74. Lehrer, A. and Kittay, E.F., eds. Hillsdale, NJ: Lawrence Erlbaum.

Camras, L., and Shutter, J.M. (2010). Emotional facial expressions in infancy. Emotion Review 2, 120-129.

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

Gasser, R.F. (1967). The development of the facial muscles in man. Developmental Dynamics 120, 357-375.

Farahvash, M.R., Abianeh, S. H., Farahvash, B., Farahvash, Y., Yagoobi, A., & Nazparvar, B. (2010). Anatomic variations of midfacial muscles and nasolabial crease: A survey on 52 hemifacial dissections in fresh Persian cadavers. *Aesthetic Surgery Journal*, 30(1) 17–21.

Haywood, K.M., and Getchell, N. (2014). Life Span Motor Development, 6th Edition, (eBook: Human Kinetics).

Izard, C.E., Hembree, E.A., and Huebner, R.R. (1987). Infants' emotion expressions to acute pain: developmental change and stability of individual differences. Developmental Psychology 23, 105-113.

Pessa, J.E., Zadoo, V.P., Adrian, E.K., Yuan, C.H., Aydelotte, J., and J.R, G. (1988). Variability of the midfacial muscles: analysis of 50 hemifacial cadaver dissections. Plastic and Reconstructive Surgery 102, 1888-1893.

Reissland, N., Francis, B., Mason, J., and Lincoln, K. (2011). Do facial expressions develop before birth? PLoS One 6, e24081.

Rinn, W.E. (1984). The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. Psychological Bullettin 95, 52-77.

Shim, K-S., Hu, K-S., Kwak, H-H., Koh, K-S., Fontaine, C., & Kim, H-J. (2008). An anatomical study of the insertion of the zygomaticus major muscle in humans focused on the muscle arrangement at the corner of the mouth. *Plastic and Reconstructive Surgery, 121*, 466-473.

Shimada, K., and Gasser, R.F. (1989). Variations in the facial muscles at the angle of the mouth. Clinical Anatomy 2: 129-134.

Tassinary, L.G: Chapter in: Cacioppo, J.T., Tassinary, L.G., and Berntson, G.G. (2007). Handbook of Psychophysiology, 3rd edition, (Cambridge, England: Cambridge University Press).

Van Boxtel, A. (2010). Facial EMG as a tool for inferring affective states. Proceedings of Measuring Behavior 2010. Eds. Spink, Grieco, Krips, Loijens, Noldus, Zimmerman. Eindhoven, the Netherlands (August 24-27, 2010).

Waller, B.M. (2008). Selection for universal facial emotion. Emotion 8, 435-439.

Box 6: A Summary of Computer Vision Algorithms for Automatically Detecting Facial Actions

The goal of computer vision algorithms is to identify a functional mapping between to variables x and y, where x is the image of a face and y the list of active action units (AUs) observed in the face in x, i.e., AUs=f(face image). There are two main solutions to the problem of identifying f(.). One is to manually derive functions that we believe do a good job at identifying AUs in images. The other approach is to learn f(.) from the data – from pairs of samples x and y. This second approach is called **machine learning**.

Manually-Derived Functions

There are three types of manually-derived functions for solving AUs=f(face image).

1. *Optical flow*: Optical flow is the apparent motion of landmark points on an object, such as a face (Kroeger et al., 2016), e.g., how does the landmark point defining the left corner of the mouth appear to move when we activate AU 12 (lip corner puller). A dense optical flow approach is when most of the pixels defining a face are landmark points. A computer vision algorithm can use dense optical flow to identify AUs on a face by computing the apparent movement of a set of facial landmark points between a generic neutral face and the given image (Donato et al., 1999; Yacoob & Davis, 1996; Martinez, 2003). One way to improve the performance of this algorithm is to use a person-specific approach, where the neutral face used to compute the optical flow is an image of the same individual as that in the image we need to analyze. <u>Advantages</u>: This approach generally worked well in lab conditions (>80%) and where the face is seen frontally and is not occluded. The algorithm's performance improves (>90%) when we use a person-specific approach. <u>Disadvantages</u>: Both a person's

neutral face (at rest) and the image of the apex of a facial configuration must first be known. Detection of the apex is a problem that has received scarce attention in computer vision. This, combined with the need for non-occluded, frontal images, has limited the use of this approach to images filmed in controlled, lab conditions. The method is also limited to people for whom we have a neutral face available or one can be detected in a video sequence.

2. *Linear filters*: Another solution is to use a linear filter h(.). A linear filter (such as a Gabor filter) is a functional mapping that acts locally on an image and has the linear constraints. The functional mapping is given by a convolution of the image with the filter h(.). A convolution can be intuitively understood as the multiplication of the pixels of the filter with a local patch of the image of the same size as the filter, with this multiplication being applied to every possible placing of the filter on the image. The response of these local multiplications is generally different when the AU is active in a face than when it is not. Because of this, local filters have been extensively used to identify AUs in images of faces (Benitez-Quiroz et al., 2016; De la Torre & Cohn, 2011; Liu & Wechsler, 2002; Lyons et al., 1999; Tian et al., 2002). <u>Advantages</u>: Convolutions are fast operators. This makes for fast (i.e., processing at >30 frames per second) and accurate algorithms (>90%) detection of AUs in images filmed in the lab as well as in the wild. To date, this is one of the most successful approaches to identifying some (but not all) AUs (Benitez-Quiroz et al., 2017a). <u>Disadvantages</u>: Local filters only work when the texture in the patch of the image is well defined. When the local patch is smooth (e.g., middle of the cheeks), the response of a filter is similar for active versus inactive AUs.

3. *Shape descriptors*: The shape of an object is defined as the geometric properties of the object when all information related to the object's position, scale and rotation have been eliminated

(Hamsici & Martinez, 2009). Shape is especially useful to define AUs that deform major components of the face, e.g., the lips, eyelids, eyebrows, nose, and jaw line. Shape can also be used to define the wrinkles seen in the forehead when activating AUs 1 (inner brow raiser) and 2 (outer brow raiser) as well as those around the nose caused by AUs 9 (nose wrinkler) and 10 (upper lid raiser). These descriptors have been successfully used to detect these and other AUs in images filmed in the lab and in the wild (Benitez-Quiroz et al., 2016; Martinez & Du, 2012; Neth & Martinez, 2010; Kotsia et al., 2008). Advantages: Works in local patches with little texture and contrast. Can be applied to low-resolution images and under varying pose. Disadvantages: Not all image changes caused by AUs define an easily detectable shape change, e.g., AUs 11 (nasolabial deepener) and 14 (dimpler).

Machine Learning-Derived Functions

There are two types of machine-learning functions for solving AUs=f(face image).

1. *Probabilistic algorithms*: Statistical learning theory and statistical pattern recognition are algorithms that learn from sample pairs: $S=\{(x_1,y_1),\ldots,(x_n,y_n)\}$, where $y_i=f(x_i)$, and f(.) is generally a probability density function (pdf) or a mixture of pdfs (McLachlan & Peel, 2004). Intuitively, a pdf is a function that give the relative likelihood of a value of x_i to belong to a value of y_i. Given the training set S, we can estimate these likelihoods and, hence, the pdf. This approach works best when combined with the computer vision features defined above (optical flow, linear filters, and shape) (Benitez-Quiroz et al., 2016; Zafeiriou et al., 2016; Corneanu et al., 2016). Advantages: To date this is the most successful approach to the recognition of AUs in the lab and in the wild (Benitez-Quiroz et al., 2016, 2017a, 2019). Disadvantages: The training set S needs to include images under a variety of image conditions (illumination, pose) as well as people of diverse

ethnicities, races and skin colors. The lack of diversity in these datasets has resulted in technology that is based to minorities (Buolamwini, 2018).

2. *Deep learning*: An alternative to the above approach is to learn the function f(.) using regression analysis. Regression analysis identifies relationships between predictor variables and dependent variables, much like probabilistic algorithms. The difference here is that we need not use a probabilistic model (Belkin & Niyogi, 2003). Rather, we assume the functional mapping provides a direct relationship between x and y (see, for example, Martinez, 2017a, for some examples). In this approach, the goal is to estimate the parameters of the manifold (function). If the function is linear, we can use linear least-squares. If the function is non-linear (which is typically the case), we need to use a non-linear optimization approach. One famous solution is gradient descent (Rumelhart et al., 1988). Additionally, when the number of parameters to be estimated is very large, this approach is called *deep learning*. Deep learning has been extremely successful in many computer vision applications. Missing from the list of successful applications, however, is the recognition of AUs. The reason for this is simple: to estimate a very large number of parameters in f(.), we require an equally large number of sample pairs (x_i,y_i), $i=1,\ldots,n$, with n large. As we discussed in the paper though, manually annotating a large number of images (x_i) with their corresponding AUs (y_i) would require years and millions of dollars to complete. A solution that is being attempted is to use automatic annotations instead (Benitez-Quiroz et al., 2017b; Zhao et al., 2016b). That is, we can use current algorithms to automatically annotate AUs in a large number of images. One then uses these annotations to train a deep learning algorithm that is robust to errors in the training data. We are awaiting additional experiments to determine if this approach will succeed. One

solution may be to use deep learning to generate realistically-looking samples, typically called deep fakes. A recent paper (Pumarola et al., 2018) shows this should be possible.

Improving Automatic Annotations

Researchers in computer vision are studying ways of improving the performance of current algorithms to map facial movements to AU codes (Corneanu et al., 2016). One promising approach identifies dependencies in facial movements (Zhao et al., 2016). Facial movements typically do not occur in isolation. Try this: Face a mirror and try to move the outer corners of your eyebrows (AU2) while keeping every other facial muscle at rest. You will notice this is a difficult task. Algorithms can learn these dependencies to improve the automatic annotation of AUs of any of the algorithms described above (Benitez-Quiroz et al., 2017b). For example, if we uncover that AUs 1 and 2 are co-articulated quite often and that AU 1 almost never co-occurs with AU 23, we can use this knowledge to improve the annotations of AU 1 once we know whether a face has AU 2 and/or 23 active. Box Figure 1 shows the dependencies obtained on a small set of posed facial configurations taken in controlled, lab conditions. Learning AU dependencies is important to understand the evolution of AU production from birth to adulthood. For example, if these dependencies are small in babies and large in adults, it suggests a learning or developmental process in the production of facial configurations; no changes may indicate an innate system that is anatomically constrained and possibly available at birth. In addition, if these dependencies vary across cultures, it would point to a cultural influence in the malleability of facial movements; no cultural variation would suggest that there are universal (possibly anatomical or neural) constraints on the facial configurations that a human face can produce. These hypotheses about AU dependencies await future research.

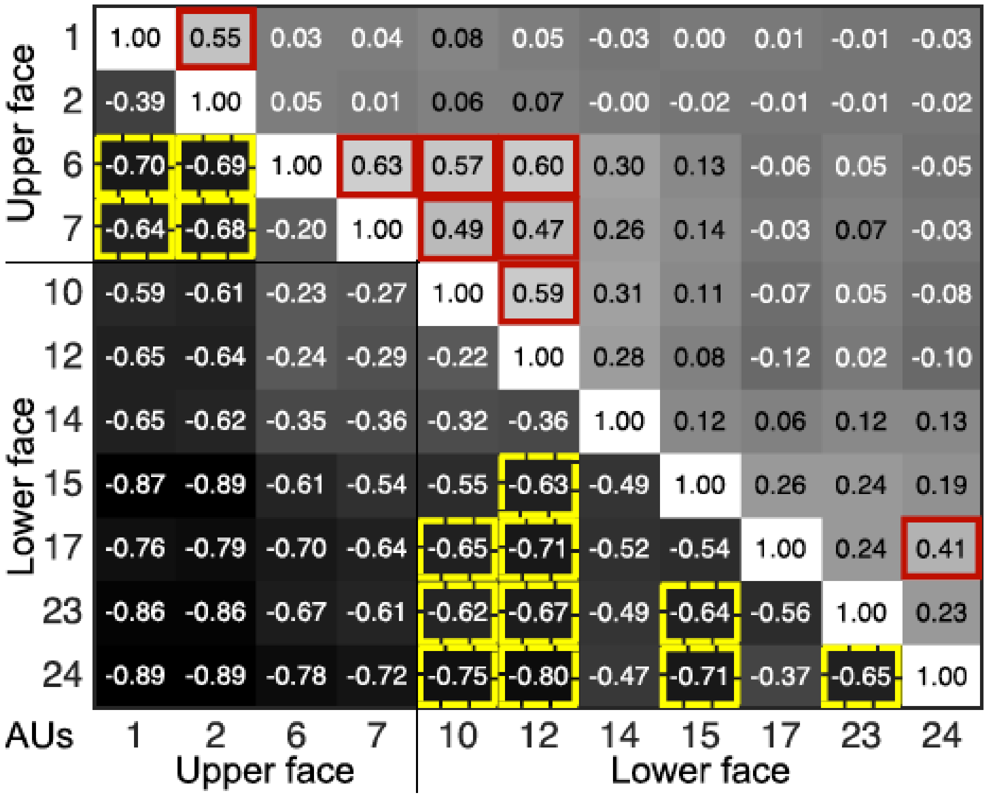
Research to Validate Automated Coding of AUs Relies on Supervised Learning

For the above algorithms to work, we need the training set S defined above. That is, we need a set of images whose AUs have been coded by an expert FACS coder (i.e., the images are manually annotated). Many databases have been collected of posed and spontaneous facial configurations in lab conditions (for example, see Lucey et al., 2010; Mavadati et al., 2013). These databases typically have between a few to several hundred training images. Recently, a large database of facial expressions in the wild was collected and made available to researchers (Benitez-Quiroz et al., 2016). This large dataset was used in the EmotioNet Challenge[3] (Benitez-Quiroz et al., 2017a), as discussed in the paper. To evaluate the accuracy of AU annotation of these computer vision algorithms, part of the dataset is used for training an algorithm and an independent set of images is used for testing the accuracy of the algorithm. Accuracy of annotations as well as true and false positive and true and false negative are usually reported.

Temporal Information

Temporal features have also been used to identify AUs (Bartlett et al., 2014). Unfortunately, these algorithms are not based on a natural model of facial movement dynamics (Cohn & Schmidt, 2004; Zuckerman et al., 1976). Instead, these algorithms attempt to identify correlations between temporal *image* features with the presence of AUs. Ideally, we would like to have algorithms that use the same temporal features employed by humans. Unfortunately, these temporal features are, for the most part, unknown at present. Hence, there is a need to better understand the dynamics of facial movements and how these dynamics aid detection and contribute to their inferred psychological meaning. This is an area of future research for both psychologists and computer scientists.

[3] http://cbcsl.ece.ohio-state.edu/EmotionNetChallenge/index.html



Box Figure 6-1. *Co-articulation of action units*. Positive correlations (red) and negative correlations (yellow) between facial actions, estimated from a set of 350,000 frames of facial movements. Adapted from Zhao et al. (2016). This approach uses dependencies to predict the presence or absence of AUs before they are detected, improving the accuracy of the algorithms to detect facial movements and map them to AUs (Benitez-Quiroz et al., 2017b). These algorithms have also been recently tested in less constrained conditions (i.e., in the wild; Zafeiriou et al., 2016).

References

Bartlett, M. S., Littlewort, G. C., Frank, M. G., & Lee, K. (2014). Automatic decoding of facial movements reveals deceptive pain expressions. Current Biology, 24(7), 738-743.

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation, 15(6), 1373-1396.

Benitez-Quiroz, C. F., Srinivasan, R., & Martinez, A. M. (2016). Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5562-5570).

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

Benitez-Quiroz, C. F., Wang, Y., & Martinez, A. M. (2017b). Recognition of Action Units in the Wild with Deep Nets and a New Global-Local Loss. In Proc. International Conference on Computer Vision (pp. 3990-3999).

Benitez-Quiroz, F., Srinivasan, R., & Martinez, A. M. (2019). Discriminant Functional Learning of Color Features for the Recognition of Facial Action Units and their Intensities. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability and Transparency (pp. 77-91).

Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. International Journal of Wavelets, Multiresolution and Information Processing, 2, 121-132.

Corneanu, C. A., Simón, 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(8), 1548-1568.

De la Torre, F., & Cohn, J. F. (2011). Facial expression analysis. In Visual analysis of humans (pp. 377-409). Springer, London.

Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Classifying facial actions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10), 974.

Hamsici, O. C., & Martinez, A. M. (2009). Rotation invariant kernels and their application to shape analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(11), 1985-1999.

Kotsia, I., Zafeiriou, S., & Pitas, I. (2008). Texture and shape information fusion for facial expression and facial action unit recognition. Pattern Recognition, 41(3), 833-851.

Liu, C., & Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Transactions on Image processing, 11(4), 467-476.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, June). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on (pp. 94-101). IEEE.

Lyons, M. J., Budynek, J., & Akamatsu, S. (1999). Automatic classification of single facial images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(12), 1357-1362.

Martinez, A. M. (2003). Matching expression variant faces. Vision Research, 43(9), 1047-1060.

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(May), 1589-1608.

Martinez, A. M. (2017a). Computational models of face perception. Current Directions in Psychological Science, 26(3), 263-269.

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

Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). Disfa: A spontaneous facial action intensity database. IEEE Transactions on Affective Computing, 4(2), 151-160.

McLachlan, G., & Peel, D. (2000). Finite mixture models. Willey, series in probability and statistics.

Neth, D., & Martinez, A. M. (2010). A computational shape-based model of anger and sadness justifies a configural representation of faces. Vision Research, 50(17), 1693-1711.

Pumarola, A., Agudo, A., Martinez, A. M., Sanfeliu, A., & Moreno-Noguer, F. (2018, July). Ganimation: Anatomically-aware facial animation from a single image. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 818-833).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, pp. 533-536, 533.

Tian, Y. L., Kanade, T., & Cohn, J. F. (2002, May). Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. In Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on (pp. 229-234). IEEE.

Yacoob, Y., & Davis, L. S. (1996). Recognizing human facial expressions from long image sequences using optical flow. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(6), 636-642.

Zafeiriou, S., Papaioannou, A., Kotsia, I., Nicolaou, M., & Zhao, G. (2016). Facial Affect``In-The-Wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Workshops (pp. 36-47).

Zhao, K., Chu, W. S., De la Torre, F., Cohn, J. F., & Zhang, H. (2016). Joint patch and multi-label learning for facial action unit and holistic expression recognition. IEEE Transactions on Image Processing, 25(8), 3931-3946.

Zhao, K., Chu, W. S., & Martinez, A. M. (2018). Learning Facial Action Units From Web Images With Scalable Weakly Supervised Clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2090-2099).

Zuckerman, M., Hall, J. A., DeFrank, R. S., & Rosenthal, R. (1976). Encoding and decoding of spontaneous and posed facial expressions. Journal of Personality and Social Psychology, 34(5), 966.

Box 7: Variations in Facial Movements

In principle, a human face can make a multitude of movement patterns: 16 million different combinations are possible, in principle, assuming each of the 24 AUs corresponding to muscle movements could move independently (ignoring temporal dynamics; Martinez, 2017b). In addition, facial muscles can contract with different intensities and varying time to peak contraction (Jack & Schyns, 2017), further increasing the number of movement patterns a face can generate, in principle. In practice, however, a much smaller subset of combinations is likely because of the way that brain controls the face as well as various anatomical constraints (e.g., some muscles are more or less likely to move together because of their relative position to one another, how they are attached to facial bones, or how they are innervated by nerves).

Several obstacles make it challenging to scientifically observe which configurations are made by healthy people and with what frequency. The most serious obstacle is one of naming: *many published studies equate facial movements with emotional expressions rather than treating their correspondence as a hypothesis to be tested*. For example, facial AU 4, 5, 7 and 23 are referred to as an expression of anger rather than as a lowered brow (AU 4), raised upper lid (AU 5), tightened lid (AU 7) and tightened lip (AU 23). This conflation is sometimes built into an experiment itself, particularly if FACS coding is not used (for a good example, see Calvo et al., 2014). Some studies measure the presence or absence of facial movements with a less systematic coding system called the EMFACS (Emotion FACS), in which coders decide whether a pre-specified group of AUs (i.e., the stipulated expression for each emotion category) are active en masse (i.e., coders identify the presence or absence of the entire configuration), rather than *independently* identify the presence or absence of each facial AU. Hence, EMFACS is less reliable than the FACS [4] because it encourages coders to make mental inferences about the meaning of the muscle movements while they are describing the muscle movements, thereby conflating scientific measurements (which facial muscles moved) with their interpretation (which emotions were expressed).

Approach

To date, no research has systematically or fully cataloged the number of configurations that are biologically possible nor which are routinely made in the wild and with what frequency, but one study makes a start, and also offers a clear example of combining automated-human FACS coding (Srinivasan & Martinez, 2018). Over seven million images were mined from the internet by first identifying all the nouns, verbs, adjectives and adverbs, along with their semantic and lexical relations in the English dictionary that many people think of as emotions.[5] These words were then translated into Spanish, Mandarin Chinese, Farsi, Arabic and Russian. The words were then used to identify and download images of human faces using a variety of online search engines. The configuration of AUs in each image was automatically FACS coded using a computer vision algorithm (Benitez-Quiroz et al., 2016; Benitez-Quiroz et al., 2017b; Benitez-Quiroz et al., 2019). Human FACS coders then manually verified the accuracy of the results provided by the automated analysis for a subset of images.

---

[4] EMFACS identifies a prescribed set of AUs that are thought to express emotion rather than coding the presence or absence of each AU one at a time. Anyone who is trained to use FACS can also use EMFACS; there is no special training. It is important to take note of which coding is used, because EMFACS is less reliable (Rosenberg, Erika. Frequently Asked Questions. erikarosenberg.com/faq) and potentially more prone to bias.

[5] WordNet (Miller, 1995) defines a structure—a graph—that identifies synonyms as well as superordinate and subordinate concepts for each word. A subordinate is a word with a more specific meaning (e.g., *despair* a subordinate of *sadness*). A superordinate is a word with more general meaning (e.g., *emotion* is a superordinate of *sadness*).

Facial Configurations Discovered

Thirty-five configurations of AUs were observed as common to the images mined in all six languages (see Box Table 7-1). This amounts to 22% of the facial configurations that were identified in the seven million images (1.87% of the seven million images contained these 35 configurations). Only eight additional facial configurations were identified that were common in the images mined in one or more, but not all, of these languages.[6]

The results provide the very first attempt at cross-cultural assessment of facial configurations observed in the wild (rather than deliberately posed) and invite a variety of interpretations. One possibility is that these findings support the hypothesis of a small number of facial configurations that are available for emotional expression within and across cultures (although admittedly significantly larger than the six proposed configurations in Figure 4, e.g., Martinez, 2017a). This interpretation is cautioned by several considerations, however.

First, the scientific approach taken by Srinivasan & Martinez (2018) is likely to have missed some facial configurations. Current algorithms only identify 16 of the 24 possible AUs.[7] Human coders only verified the AUs detected by the algorithm, meaning that AUs that were actually present in the images but went undetected by the algorithm were missed entirely. More generally, as illustrated by the EmotioNet Challenge (http://cbcsl.ece.ohio-state.edu/EmotionNetChallenge/), current algorithms do not have high accuracy for detecting facial movements in the wild. In addition, English words were used to mine for images, but the other languages sampled (Spanish, Mandarin Chinese, Farsi, Arabic and Russian) contain

[6] Of these, 2 were observed in a single language, 2 in two languages, 3 in three languages and 1 in four languages.
[7] These are the most frequently observed AUs. Those AUs that are not currently detected are infrequent.

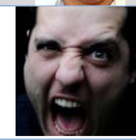
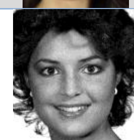
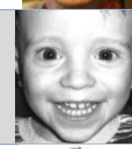
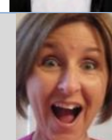
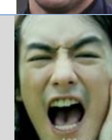
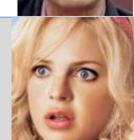
indigenous words for emotions that do not correspond easily to single English words (Smith, 2016), and therefore may be associated with AU configurations that were not sampled in the first place.

Second, the images used by Srinivasan & Martinez (2018) were derived from the internet. Internet images, while better than posed faces, do not substitute for scientific observations of facial movements in the real world. The internet is a curated version of reality. Some common facial configurations are likely missing because they are rarely uploaded to the internet, and some configurations commonly found on the internet may not be commonly observed in the real world.

For these reasons, and also because the cultures sampled have some contact with practices and norms of the U.S., the Srinivasan & Martinez (2018) study does not, on its own, confirm that the 35 identified facial configurations are, in fact, universal. Instead, the study suggest that these facial configurations are commonly used to express instances of emotion across a number of languages in industrialized nations. Furthermore, even if future studies reveal that some or all of these 35 facial configurations are indeed universally made, this does not automatically mean that each is an *innate* expression with a unique emotional meaning; the universality of these expressions may be a result of cultural norms people learn as children (see Box 10).

Box Table 7-1: Thirty-five unique combinations of facial actions observed in cultures whose primary languages are English, Spanish, Mandarin Chinese, Farsi, and Russian (from Srinivasan & Martinez, 2018).

ID	AUs	Examples	ID	AUs	Examples	ID	AUs	Examples

1	4		13	4, 7, 9, 10, 17		25	1, 2, 5, 25, 26	
2	5		14	9, 10, 15		26	4, 7, 9, 25, 26	
3	2, 4		15	12, 25		27	12, 25, 26	
4	4, 7		16	6, 12, 25		28	2, 12, 25, 26	
5	12		17	2, 5, 12, 25		29	5, 12, 25, 26	
6	2, 12		18	1, 2, 5, 12, 25		30	2, 5, 12, 25, 26	
7	5, 12		19	6, 12, 25		31	1, 2, 5, 12, 25, 26	
8	1, 2, 5, 12		20	10, 12, 25, 26		32	6, 12, 25, 26	
9	6, 12		21	1, 2, 25, 26		33	7, 9, 20, 25, 26	
10	4, 15		22	1, 4, 25, 26		34	1, 2, 5, 20, 25, 26	
11	1, 4, 15		23	5, 25, 26		35	1, 4, 5, 20, 25, 26	
12	4, 7, 17		24	2, 5, 25, 26				

Note. ID is the unique identification number given to each facial configuration. An example of each is shown. AUs are the active facial action units that describe each configuration. IDs 1, 2, 10, 11, and 21 were most frequently labeled by participants as expressions of sadness. IDs 5 through 9, 15 through 19, and 26 through 32 were most frequently labeled as expressions of happiness. IDs 20 and 22 through 24 were most frequently labeled as expressions of surprise. IDs 4, 12, 14, 25, and 35 were most frequently labeled as expressions of anger. IDs 33 and 34 were most frequently labeled as expressions of fear. ID 13 was most frequently labeled as an expression of disgust. No consistent labels were offered for ID 3. No configuration exactly matches the AU configurations proposed by Darwin or documented in prior research (for AU comparisons, see Cordaro et al., 2017; Ekman et al., 1968; Matsumoto et al., 2008).

References

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

Benitez-Quiroz, F., Srinivasan, R., & Martinez, A. M. (2019). Discriminant Functional Learning of Color Features for the Recognition of Facial Action Units and their Intensities. *IEEE transactions on pattern analysis and machine intelligence*.

Benitez-Quiroz, C. F., Wang, Y., & Martinez, A. M. (2017b). *Recognition of action units in the wild with deep nets and a new global-Local loss*. In Proceedings of the International Conference on Computer Vision.

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

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

Srinivasan, R., and 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. ArXiv preprint, available at https://arxiv.org/pdf/1808.04399.pdf

Box 8: Meta-Analytic Evidence of Autonomic Nervous System Changes During Emotion

There have been four statistical summaries of scientific findings (called meta-analyses) from experiments designed to measure autonomic nervous system (ANS) changes during emotional episodes (Cacioppo et al., 2000; Siegel et al., 2018; Stemmler, 2004; for a discussion, see Quigley & Barrett, 2014). All of them, including the most comprehensive meta-analysis published to date covering over 200 published experiments involving more than 20,000 participants (Siegel et al., 2018) replicate the same results: ANS changes are neither consistent nor specific for any emotion category. Combining different measurements into a pattern performed no better at distinguishing one emotion category from another than did individual measures (Siegel et al., 2018). Some review articles have tried to make a case for the existence of emotion-specific ANS fingerprints (e.g., Friedman, 2010; Kreibig, 2010), but the meta-analyses are consistent in their findings that different emotion categories cannot be distinguished from one another by changes in heart rate, respiration rate, skin conductance, or any other measure of the autonomic nervous system, alone or in combination; said another way, ANS measures individually, or in combination, are neither consistent nor specific for emotion categories.

The variety in emotion-related ANS changes is consistent with the writings of William James (1890), population views of emotion (Box 2), with Darwin's articulation of *population thinking* in *On the Origin of Species* (1859/2001) (for a discussion, see Barrett, 2017) and with the US physiologist Paul Obrist's findings that peripheral physiological changes are tied to the metabolic demands associated with action (e.g., cardiosomatic coupling; Obrist, Webb, Sutterer, & Howard, 1970) or anticipated action (e.g., supra-metabolic activity; Obrist, 1981; Sterling, 2012; Turner & Carroll, 1985). Because all animals (including humans) behave in a variety of context-sensitive ways, crying, shouting, smiling, freezing and laughing in anger will each be supported by a distinct pattern of ANS change. In this view, ANS variation is not a bug to be explained away as error or designated as epiphenomenal to the nature of emotion. Substantial variation in ANS patterns within an emotion category is a feature that should be expected because it confers evolutionary advantage (for a discussion on how evolution selects for variation in emotion, see Barrett, 2017a).

The variation in emotion categories observed for ANS changes is consistent with the evidence for the brain basis of emotion in humans (Barrett, 2017b; Clark-Polner et al., 2016) and in non-human animals (for a discussion, see Barrett, 2017b; Barrett & Satpute, 2017). For example, even studies of "fear" learning in rodents find evidence of variability in ANS responses and neural circuitry (e.g., Barrett & Finlay, in press; Gross & Canteras, 2012; Iwata & LeDoux, 1988; Tovote et al., 2015). Behavioral experiments also clearly show that the variation within emotion categories is meaningfully tied to context and situational factors and is not merely due to variability in the experimental method. A growing number of studies of emotion are designed to explicitly model and capture heterogeneity within emotion categories both within individuals and across cultures (e.g., Ceulemans, Kuppens, & Van Mechelen, 2012; Gendron, Roberson, van der Vyver, & Barrett, 2014; 2014; Hortensius, Schutter, & Harmon-Jones, 2011; Kuppens, Van Mechelen, & Rijmen, 2008; Kuppens et al., 2007; Nezlek, Vansteelandt, Van Mechelen, & Kuppens, 2008; Stemmler, Aue, & Wacker, 2007; Wilson-Mendenhall, Barrett, & Barsalou, 2013; 2015; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011).

References

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

Barrett, L. F. (2017b). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12,* 1-23. https://doi.org/10.1093/scan/nsw154

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

Barrett, L. F., & Satpute, A. B. (2017). Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience Letters*.

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 emotion* (2 ed., pp. 173-191). New York: Guilford Press.

Ceulemans, E., Kuppens, P., & Mechelen, I. V. (2012). Capturing the Structure of Distinct Types of Individual Differences in the Situation-specific Experience of Emotions: The Case of Anger. *European Journal of Personality, 26*(5), 484-495.

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.

Darwin, C. (1859/1964). *On the Origin of Species*. (A facsimile of the first edition. ed.). Cambridge, MA.: Harvard University Press

Friedman, B. H. (2010). Feelings and the body: the Jamesian perspective on autonomic specificity of emotion. *Biological Psychology, 84(3)*, 383-393.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion, 14(2)*, 251-262. doi:10.1037/a0036052

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

Gross, C. T., & Canteras, N. S. (2012). The many paths to fear. *Nature Reviews Neuroscience, 13(9)*, 651-658.

Hortensius, R., Schutter, D. J., & Harmon-Jones, E. (2012). When anger leads to aggression: induction of relative left frontal cortical activity with transcranial direct current stimulation increases the anger–aggression relationship. *Soc Cogn Affect Neurosci, 7(3)*, 342-347.

Iwata, J., & LeDoux, J. E. (1988). Dissociation of associative and nonassociative concomitants of classical fear conditioning in the freely behaving rat. *Behavioral Neuroscience, 102(1)*, 66-76.

James, W. (1890/2007). *The Principles of Psychology* (Vol. 1). New York: Dover.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biological Psychology, 84(3)*, 394-421. doi:10.1016/j.biopsycho.2010.03.010

Kuppens, P., Van Mechelen, I., & Rijmen, F. (2008). Toward disentangling sources of individual differences in appraisal and anger. *Journal of Personality, 76(4)*, 969-1000.

Kuppens, P., Van Mechelen, I., Smits, D. J. M., De Boeck, P., & Ceulemans, E. (2007). Individual differences in patterns of appraisal and anger experience. *Cognition & Emotion, 21(4)*, 689-713. doi:Doi 10.1080/02699930600859219

Nezlek, J. B., Vansteelandt, K., Van Mechelen, I., & Kuppens, P. (2008). Appraisal-emotion relationships in daily life. *Emotion, 8(1)*, 145-150.

Obrist, P. A. (1981). *Cardiovascular psychophysiology: A perspective.* New York: Plenum.

Obrist, P. A., Webb, R. A., Sutterer, J. R., & Howard, J. L. (1970). The cardiac-somatic relationship: some reformulations. *Psychophysiology, 6(5)*, 569-587.

Quigley, K. S., & Barrett, L. F. (2014). Is there reliability and specificity of autonomic changes during emotional episodes? Guidance from the conceptual act theory and psychophysiology. *Biological Psychology, 98*,82-94.

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K. S., & Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin, 144(4)*, 343-393.

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

Stemmler, G., Aue, T., & Wacker, J. (2007). Anger and fear: Separable effects of emotion and motivational direction on somatovisceral responses. *International Journal of Psychophysiology, 66(2)*, 141-153.

Sterling, P. (2012). Allostasis: a model of predictive regulation. *Physiology & Behavior, 106(1)*, 5-15.

Tovote, P., Fadok, J. P., & Lüthi, A. (2015). Neuronal circuits for fear and anxiety. *Nature Reviews Neuroscience, 16(6)*, 317-331.

Turner, J. R., & Carroll, D. (1985). Heart rate and oxygen consumption during mental arithmetic, a video game, and graded exercise: Further evidence of metabolically exaggerated cardiac adjustments? *Psychophysiology, 22*, 261–267.

Wilson-Mendenhall, C., Barrett, L. F., & Barsalou, L. W. (2013). Neural evidence that human emotions share core affective properties. *Psychological Science, 24*, 947-956. PMC4015729

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.

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia, 49*, 1105-1127.

Box 9: Emotional Episodes and Their Affective Features

In English, the word ‘‘affect’’ means ‘‘to produce a change.’’ To be affected by something is to be influenced by it. In science, and particularly in psychology, ‘‘affect’’ refers to a special kind of influence—something’s ability to influence your mind in a way that is linked to changes in your body. Sometimes “affect” is used as a cautious term, to mean anything emotional. It allows people to refer to emotions in general terms, without specifying exactly what an emotion is or how it should be defined. Sometimes “affect” is used to refer specifically to emotional experiences – to be affected is to feel something (e.g., Panksepp, 1998). In modern psychological usage, ‘‘affect’’ refers to the mental counterpart of internal bodily sensations, whether or not those sensations are associated with emotions. Historically, “affect” referred to simple feelings that are part of every waking moment of your life (Wundt, 1998b/1897; for a discussion, see Barrett & Bliss-Moreau, 2009). This allows us to clarify some persistent confusions that muddle the scientific study of emotions and their expressions.

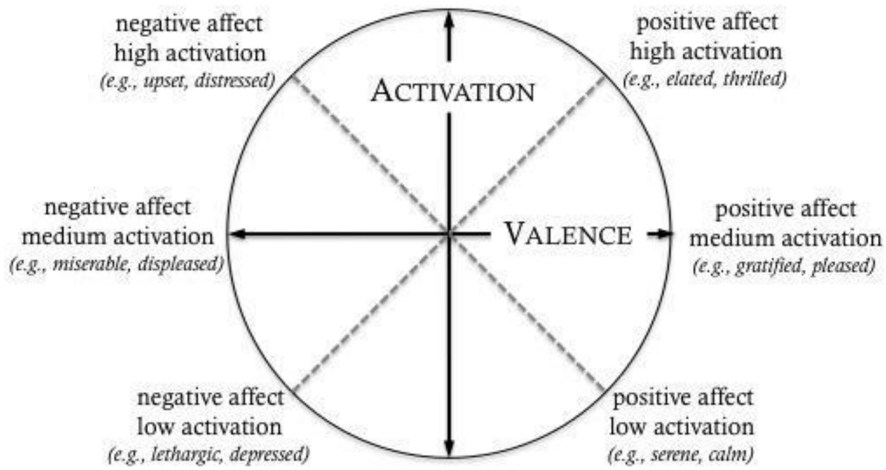
Affective Features

Affect is a property of consciousness (e.g., Edelman & Tononi, 2000; James, 1890; Searle, 1992, 2004) and is not perfectly synonymous with emotion. More than a century of research has revealed that affect, whether part of an emotional episode or otherwise, can be described as a single point in a space with at least two features: valence (pleasure to displeasure), arousal (high to low), although other features are sometimes discussed (interpersonal closeness to distance, dominance to submissiveness, etc.; Russell & Barrett, 1999). Valence and arousal are not ingredients of affect – they are descriptions of feeling in a given moment. They are descriptive features of emotional instances (Box 2), but they are not specific to emotion. Nor are causal processes or mechanisms that cause anything. These two features of affect form a low-dimensional, circular space that describes how a person feels at any moment in time, as in Box Figure 9-1. Valence and arousal are not independent from one another (as one feature changes, so does the other; for a discussion, see Barrett & Bliss-Moreau, 2009; Kuppens et al., 2013).

Thus far, what we understand about the brain-basis of affect (e.g., Lindquist et al., 2016) is consistent with the hypothesis that affective feelings are associated with a wide range of psychological phenomena (reviewed in Barrett & Bliss-Moreau, 2009). Valence and arousal are likely low dimensional representations of internal bodily sensations (referred to as interoception; Craig, 2015) that result from the constantly changing state of the body's internal systems, such as the autonomic nervous system, the immune system, the endocrine system, and so on (referred to as allostasis; Sterling, 2012). Somehow, bodily sensations, which are physical, are transformed into affective feelings, which are mental. Scientists don't yet know how this transformation happens, but many studies suggest that it does. For the moment, it remains one of the mysteries of consciousness.

Emotional Episodes

Affect provides a quick summary of the physiological state of the body, like a barometer, without much detail (Barrett & Bliss-Moreau, 2009). Therefore, affect alone does not indicate what to do next, or how to act, other than to approach something or avoid it (Davidson, 1992; Lang et al., 1993). Many scientists propose that emotional episodes, on the other hand, are specific instances of affect that involve very specific intentions to act. Their specific relation to affect depends on how emotion is defined (various theoretical proposals are presented in Box 2).



Box Figure 9-1. The affective circumplex. Hedonic valence is represented on the horizontal axis and arousal on the vertical axis. Reprinted with permission from Barrett & Bliss-Moreau (2009).

References

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in Experimental Social Psychology*, *41*, 167-218.

Craig, A. D. (2015). *How do you feel?: an interoceptive moment with your neurobiological self*. Princeton University Press.

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. Brain Cogn, 20(1), 125-151.

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York: Basic books.

James, W. (1890/2007). *The Principles of Psychology* (Vol. 1). New York: Dover.

Kuppens, P., Tuerlinckx, F., Russell, J. A and Barrett, L. F. (2013). The relationship between valence and arousal in subjective experience. *Psychological Bulletin, 139*, 917-940.

Lang, P.J., Greenwald, M.K., Bradley, M.M., & Hamm, A.O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*, 261–273.

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The brain basis of positive and negative affect: Evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex*. 26, 1910-22.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

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.

Searle, J. R. (1992). The rediscovery of the mind: MIT press.

Searle, J. R. (2004). Mind: a brief introduction: Oxford University Press.

Sterling, P. (2012). Allostasis: a model of predictive regulation. *Physiology & Behavior*, *106(1)*, 5-15.

Wundt, W. (1998/1897). *Outlines of Psychology* In: Judd, C.H. (Trans.). Leipzig: Wilhelm Engelmann.

Box 10: Learning to Express and Perceive Emotions

Expressing Emotion

Increasingly, scientists have been riveted by discoveries that very young human infants are powerful learners who quickly learn and make use of cues from their environments. When children begin to produce facial movements during emotional episodes with some regularity, they appear to reflect the influences of learning about cultural and social expectations (for review, see Kärtner, Holodynski, & Wörmann, 2013). For example, social smiling is often described as a purely maturational phenomenon, but recent cross-cultural studies suggest its development is influenced by sociocultural factors (see Wörmann, Holodynski, Kärtner, & Keller, 2012). Children are sensitive to variations in the distribution of multimodal sensory changes in their environments, and this variation becomes reflected in the perceptual categories children form for facial movements (Pollak & Kistler, 2002). Infants generalize from small samples to larger populations and gauge their inferences depending upon whether sampling strategies are strong or weak (Denison, Reed & Xu, 2013; Gweon, Tenenbaum & Schulz, 2010). Young children are also able detect complex probabilistic information in their environments, as well as implicitly learn and reproduce the underlying statistical distributions (Plate et al., 2017). These studies suggest that the patterns of facial communication are readily learned by children from their social environments, likely beginning shortly after birth. As noted by Sullivan and Lewis (2003), how specific patterns of facial movements become associated with specific contexts is still unresolved.

There is some evidence that infants learn to express instances of emotion with facial movements as they begin to acquire the emotion concepts of their cultural context. For example, recent studies suggest the facial configuration proposed as the expression of disgust (see Figure

4) is learned in middle childhood as children learn their culture's concept for disgust (Stevenson, Oaten, Case, Repacholi & Wagland, 2010; Widen & Russell, 2013).[8] Further details remain unknown however, due to a gap in the scientific literature. Published studies that carefully describe a child's facial muscle movements during emotional episodes have not yet directly measured individual differences in emotion concept learning, nor variation in parenting or family environment that serves as the context to learn emotion concepts. Studies that have examined the link between how children moves their faces during emotional episodes and how their parents act during emotional situations unfortunately do not precisely quantify the facial muscle movements.

Emotion Perception

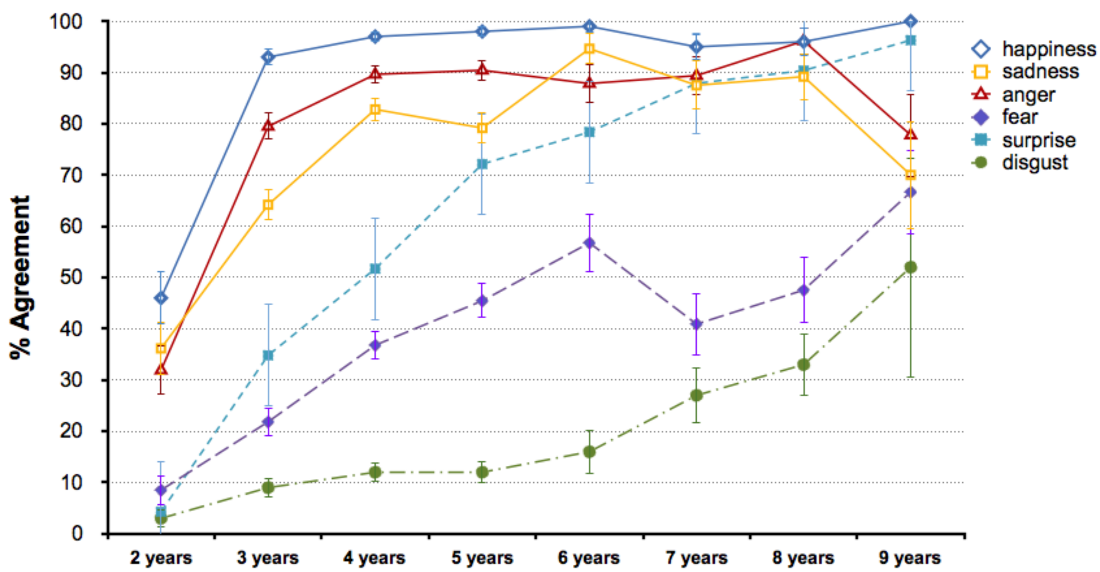
Infants begin by implicitly perceiving broad affective distinctions (e.g., positive-negative; approach-avoid; pleasant-unpleasant) as discussed in the main text of the paper. Additional research is needed to understand the developmental trajectory of infants' implicit emotion perceptions. More is known about the development of children's explicit emotion perception capacities. Numerous experiments now suggest that very young children do not explicitly understand the emotional meaning of facial movements, but that this is a skill they learn over the course of development (Pollak & Kistler, 2002). That is, newborns appear to learn to detect faces and facial movements in the first few days of life, but their capacity to associate these movements with the emotional meaning (i.e., to see facial movements as emotional expressions) is learned during childhood.

---

[8] These studies focus on learning the association between emotion words and facial movements. However, as we describe below, other theorists (e.g., Röttger-Rössler et al. 2015) have proposed that infants begin learning associations between faces and other nonverbal emotion components during early interactions with their caregivers. In this manner, caregivers might shape infants' nonverbal emotion concepts. For example, adults might respond to infants' diffuse responses to emotion eliciting situations by producing specific facial movements and behavioral responses of their own. Thus, probabilistic learning from the environment could result in the infant learning emotion categories, which might be considered "concepts" as well (Pollak, 2009). Still, there is little direct evidence with respect to distinguishing among different negative emotions concepts in infancy.

One careful line of research provides robust, consistent evidence for this hypothesis (reviewed in Widen, 2016). In this research, young children are shown the proposed expressive forms in Figure 4 and are asked to freely label them as emotional expressions by nominating emotion words. Almost twenty experiments now suggest that North American toddlers make broad affective distinctions (e.g., feels good-feels bad) when explicitly labeling the facial configurations, but these perceptions narrow and become more emotionally differentiated (more **emotionally granular**) as children's emotion concepts develop (again, see Widen, 2016). Summarizing across eleven different published studies observing over 1,000 children ranging in age from two to nine years old, it is possible to discern an average developmental trajectory from affect perception to emotion perception (see Box Figure 10-1). Children's perceptions begin to differentiate around the age of two years of age and continue becoming more specific until the middle school years when they look more adult-like (for a review, see Widen, 2016). Even elementary school-aged children commonly label disgust as "anger", and when asked to select disgust from standardized sets of facial configurations, improvement with age is very gradual, with only about 50% of nine year olds offering the word "disgusted" (or a close synonym) for the proposed disgust expression (Widen & Russell 2013). There is growing evidence that

Nonetheless, some scientists argue that more specific aspects of emotion perception are operating much earlier in development than the above studies seem to suggest (e.g., Izard et al., 2010; Leppänen & Nelson, 2009). For example, Grossman (2010) claims that by age three months infants can distinguish proposed expressive configurations for happiness, surprise, and anger, and that by seven months they can discriminate the proposed expressive configurations for fear, sadness, and interest. These claims suggest the interesting possibility that emotion perception might be observed much earlier than two years of age using methods that do not require children to overtly label faces depicting emotional expressions. They are also consistent with other evidence suggesting that infants and young children may implicitly perceive emotion in facial movements when assessed with the habituation task or a perceptual matching task. Left uncertain is whether, beyond discriminating between different facial movements, infants understand the emotional meaning that is typically inferred from these cues. Once again, we see that the choice of stimuli and task design can yield substantially different conclusions.



Box Figure 10-1. Developmental trajectories for emotion concept acquisition. N = 1065. Adapted from Widen (2016) with permission From an early age, children used the expected emotion labels (with standard errors) for the smiling facial configuration ("happiness"), the scowling configuration ("anger"), and the frowning configuration ("sadness") but the expected emotion labels for the other facial configurations gradually increased with age. Data from 11 studies were aggregated (for details, see Widen, 2016). The N for each age group was: two years (N=94), three years (N=229), four years (N=299), five years (N=209), six years (N=74), seven years (N=66), eight years (N=61), and nine years (n=33).

References

Denison, S., Reed, C. & Xu, F. (2013) The emergence of probabilistic reasoning in very young infants: evidence from 4.5- and 6-month-old infants. *Developmental Psychology, 49(2)*, 243-249.

Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restor Neurol Neurosci, 28(2):*219-36

Gweon, H., Tenenbaum, J., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences, 107(20)*, 9066-9071

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

Kärtner, J., Holodynski, M., & Wörmann, V. (2013). Parental ethnotheories, social practice and the culture-specific development of social smiling in infants. *Mind, Culture, and Activity, 20(1),* 79-95.

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

Plate, RC. Fulvio, JM.; Shutts, K; Green, CS; Pollak, SD. (2018). Probability learning: Changes in behavior across time and development. *Child Development, 89*, 205-218.

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, 99(13),* 9072-9076.

Stevenson, R. J., Oaten, M. J., Case, T. I., Repacholi, B. M., & Wagland, P. (2010). Children's response to adult disgust elicitors: Development and acquisition. *Developmental Psychology, 46(1),* 165.

Sullivan, M. W., & Lewis, M. (2003). Contextual determinants of anger and other negative expressions in young infants. *Developmental Psychology, 39(4)*, 693.

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

Widen, S. C. (2016). The development of children's concepts of emotion. In L. F. Barrett, M. Lewis, J. M., & Haviland-Jones (Eds.), *Handbook of emotions* (4th ed., pp. 307-318). New York: Guilford Press.

Wörmann, V., Holodynski, M., Kärtner, J., & Keller, H. (2012). A cross-cultural comparison of the development of the social smile: A longitudinal study of maternal and infant imitation in 6-and 12-week-old infants. *Infant Behavior and Development, 35(3),* 335-347.

Box 11: Research with Virtual Humans

Virtual humans (Rickel et al., 2002) (or Embodied Conversational Agents, Cassell et al., 2000) are software-based artifacts that not only look like people but are also engineered to interact with us using the same verbal and nonverbal behavior that we use to interact with each other. To support interaction, virtual humans are designed with a range of technologies that allow them to interpret a human's behavior, maintain beliefs about the interaction, reason about how to achieve its goals given those beliefs and express itself to achieve those goals. This requires the researcher to engineer virtual humans to simulate perception, beliefs, goals and behavior, bringing together theory and scientific insights spanning traditional scientific disciplines. For example, to model emotion expression and perception in a virtual human, a designer may consider the interactions between the simulations of emotion with those for perception, decision-making, behavior and consequently social interaction (Rickel et al., 2002; Becker-Asano & Wachsmuth, 2008).

In the process of crafting a virtual human, a designer is exploring challenges from an engineering perspective that are closely tied to fundamental psychological questions about the mechanisms and processes that underlie mental events, such as how should a virtual human's internal states such as emotions and communicative intentions map to its expressive behavior in the context of a social interaction? A designer must also grapple with questions about the mechanisms and processes that infer emotional meaning from facial movements (i.e., emotion perception), such as how should a virtual human interpret its perceptions of another's facial movements and what role should its prior beliefs, goals, emotions and the context of the interaction play in that interpretation? Virtual human designs vary considerably in the details of the mechanisms and processes used to perform these tasks. They may seek to simulate how these inference processes work, for example how do prior beliefs bias interpretations. They may use machine learning approaches to acquire a model from human behavioral data. At another extreme, the designer may more simply use a canned or fixed approach that makes specific inferences in specific situations.

There are some unique challenges when considering how to engineer emotion expressions and perceptions, because often the goal is not to faithfully render human behavior as it occurs in everyday life, but to implement them in a virtual human to achieve some other scientific or applied goal. For example, the design of a child's virtual tutor (Lester et al. 1997) may employ exaggerated, unambiguous or supernormal nonverbal behavior to motivate the student. Nonetheless, in other cases, psychological realism in the virtual human's nonverbal behavior, as well as accurate inferences about the human's nonverbal behavior, is often desirable, such as in systems that are designed to train social skills, where inferring the psychological meaning of movements, like facial movements, is crucial (Kron et al., 2017). In some cases, psychological realism is paramount, as when virtual humans are designed to elicit behavior in a human interaction partner that allows the diagnosis of depression (Devault et al, 2015).

Mapping the virtual human's simulated internal states to its movements

Often the realization of expressive behavior in a virtual human involves two connected computational models. There is the model of the emotion elicitation that determines the virtual human's internal, mental states, in essence simulating how the virtual human's experiences elicit the internal emotional states it "feels." In addition, there is a model that maps those states to the behaviors or physical movements such as facial actions designed to express those states. Some studies also endow virtual humans with the capacity to regulate their emotion simulations

(Marsella & Gratch, 2009), allowing the face to dynamically portray episodes of both authentically "felt" emotion and strategically intended emotional expressions.

*Simulating emotional states***.** The design of the computational models of emotion elicitation used in these systems (Dias & Paiva, 2005; Marsella & Gratch, 2009; Becker-Asano & Wachsmuth, 2008) have been most heavily influenced by appraisal theories (See Box 2). Appraisal theories are popular in computational models in virtual human research because each appraisal (e.g., pleasantness, novelty, etc.) maps readily to formalisms used in Artificial Intelligence to model internal states such as beliefs, desires and intentions.[9]

The computational models of emotion elicitation used in virtual humans have not been empirically validated against human data, except for a few notable exceptions. For example, one study compared human self-reports of their changing emotional experience over the course of playing the two-person competitive game of Battleship to the predictions of alternative computational models of emotion (Gratch et al., 2009). A related study (Marsella et al, 2009) evaluated a computational model, EMA, (EMotion and Adaptation, inspired by the appraisal model developed by Lazarus, 1991) that could simulate an emotional episode and regulate the simulation. Again, the game of Battleship was used. In both cases, the subjective reports of emotional experience at key points in the unfolding game were compared to the model's predictions for those experiences (i.e., what the model predicted the real human would feel and how the person would regulate those feelings). In both cases evidence was found supporting aspects of the various models while also identifying some discrepancies between subjects' subjective report and model predictions. For example, EMA correctly predicted the changes in

[9] Interestingly, many of the computational models have been based on appraisal models, such as Ortony, Clore & Collins (1988). However, the computational work has treated it as a causal model, so that the appraisals are the sole cause of emotions, even though the work of Ortony et al was designed as a descriptive model.

subjective reports of emotion intensity, as well as changes in subject's reports of their effort to win and the importance of winning, in the service of regulating their emotions. However, it incorrectly predicted that subjects would engage in wishful thinking, increasing their expectations of eventually winning, in the face of evidence that they were actually losing. In fact, when faced with losing, subjects tended to lower their expectations of winning, perhaps again as a means to down regulate the eventual negative emotions that would result from losing.

*Mapping states to movements***.** Once an emotion simulation is implemented, another model is used to express the simulated emotions in the virtual human's movements, including facial actions. Expression stereotypes (see Figure 4) have frequently been used to map the simulated emotions provided by these computational models to facial actions made by virtual humans. The component process model of emotion (Scherer et al., 2017) has also been explored by tying the virtual human's facial actions to the simulated appraisals generated by its emotion model (Paleari et al, 2007; Malatesta et al, 2009).

Data-driven techniques provide an alternative to theory driven approaches. One approach to using data is to depend on other people's beliefs about which facial movements express emotion, rather than to rely on scientists. For example, posed expressions from actors can be digitally scanned, creating visually realistic faces with posed facial actions (Alexander et al., 2009) that in appearance can be indistinguishable from humans.[10] However, the goal here is typically grounded in perception, as opposed to physical realism. Specifically, the facial movements and their dynamics, as expressions of emotion, are based on beliefs, stereotypes or artistic interpretation, with an underlying expectation that a human user will infer emotional

[10] Digital scanning enables film makers to fool audiences so that, for example, they can bring an actor back to life, twenty years after his death, as was done in the recent Hollywood film, Rogue One, when Peter Cushing (who played a general in the Empire) made an appearance.

meanings from those movements. The main weakness in these approaches is ecological – as we discuss in the main text of the paper, people's beliefs about emotional expressions are often better thought of as stereotypes and may not accurately capture how they actually move their faces when expressing emotion in real life. An alternative data-driven approach, one with greater potential for **ecological validity**, is to record people's facial movements while they are engaged in situations that are assumed to evoke emotion, either in the lab or in everyday life, but this requires inferring which emotional episodes are being created (for a discussion on the difficulties in measuring an emotional episode, see discussion in the main paper).

To date, there is no strong empirical evidence that compares these alternative models of expressing emotions in terms of the psychological realism of the facial movements they generate.

An additional challenge is the mismatch between abstract details provided by a psychological hypothesis and the far more specific engineering details about the appearance of the character's face, skin, as well as realizing realistic dynamics for the individual facial actions required to animate a virtual human's face. Those dynamics include not only timing of each action unit, but also its duration and the rates at which it moves. Furthermore, the dynamics of the action units must influence other elements, such as the bunching and wrinkling of skin. Beyond realizing one configuration of facial movements, there is also sequencing of the movements over time, such as when initial surprise may transition into an instance of fear or anger. Such dynamics encompass not only the facial movements themselves but also the dynamics of the underlying mental states that are being expressed. These factors are all critical to getting a realistic effect. That is, basic emotion theory and the componential model of emotion describe their hypotheses in terms of abstract (non-physical) processes, not the physical mechanisms by which movements or simulations are realized. This requires designers to make inferences as they build computational models. As a consequence, in a given study, it is difficult to infer whether the results pertain to the engineering choices made to realize a psychological theory or more generally to the theory itself.

More broadly, mapping a virtual human's simulated internal states to its movements is, in fact, more complex than simply connecting the emotion simulation (as an output of an emotion elicitation model) to physical movements (as an output of an expression model). Virtual humans are designed to engage us in face-to-face interaction. As in human-to-human social interactions, facial movements in a virtual human-to-human interaction serve a variety of functions, such as establishing rapport by mirroring the human participants' behavior (Huang et al., 2011; Tickle-Degnen & Rosenthal, 1990), emphasizing something with shared attention (Ekman, 1979), regulating turn-taking in the interaction, greetings, communicating attitudes, and so on (Cafaro et al., 2017).

Endowing a virtual human with the capacity to infer emotions from facial movements

Virtual humans often are designed to infer psychological meaning in the nonverbal movement of real humans and use those inferences to inform how they interact. This involves two connected technologies: technologies to sense facial movements, and a model of emotion perception that determines the virtual human's emotion inferences about what those states mean.

In drawing emotional inferences about a human's facial movements, virtual human designs have typically exploited standard technologies that rely on data-driven machine learning techniques discussed earlier. For example, the SimSensei virtual human system was designed to interview patients suffering from depression and post-traumatic stress syndrome (PTSD), as well as nonpatients, asking people questions such as "What is your dream job?", "What do you do now?" and "Tell me about something you did recently that you really enjoyed?" (Devault et al., 2015). A range of sensing technologies were used to measure, for example, smile intensity/duration (OKAO[11]), facial actions & emotion (FACET[12]) and acoustic/vocal qualities[13].

Wörtwein & Scherer (2017) reports on another study that used the SimSensei virtual human. The focus of the study was to discover which questions asked by the virtual human and which nonverbal actions emitted by the participants were most diagnostic of whether or not a given test-subject was suffering from PTSD. Questions were ranked in terms of information gained[14] to discover the most diagnostic nonverbal actions associated with that information. For example, depending on the question, reduced variability of movement (they moved the facial feature less), or reduced average displacement for facial actions AU1, AU2, AU9, AU17, AU18 and AU26 was found to be correlated with PTSD. Interestingly, which facial actions were more diagnostic depended on the question, leading the researchers to suggest that context matters in the sense that the specific context of the question being asked determined what facial actions were most diagnostic of PTSD.

More recent work has begun to explore how a virtual human's model of the human can help bias the inferences of the human's mental states. For example, to the extent the virtual human has a model of the goals and beliefs of the human, it can use a theory of emotion[15] to predict how the human may emotionally react and use those predictions to refine its interpretation of the human's facial movements as well as use the movements to refine its model of the human's beliefs and goals (Alfonso et al., 2015). In addition, the virtual human can

---

[11] //www.omron.com/r d/coretech/vision/okao.html
[12] http : //www.emotient.com/
[13] http : //www.cogitocorp.com/
[14] See the discussion of information theory in Box 16.
[15] In the two works we cite here (Alfonso et al, 2015; Yongsatianchot & Marsella, 2016), variants of appraisal theories were used to provide the virtual agent with a folk theory of emotion. Alfonso et al (2015) used a model inspired by Smith & Lazarus (1990) and Roseman (2001) while Yongsatianchot & Marsella (2016) used a model inspired by Ortony et al., (1988)

acquire and refine over the course of an interaction how a specific human tends to move his face during a situation thought to evoke emotions (Yongsatianchot & Marsella, 2016).

Should we care about the impact of virtual human technology?

Virtual humans provide an important tool for studying the human perception of emotion in the laboratory. However, they are likely to have an additional, broader social impact. As these systems increasingly play a role in our day-to-day lives, they are likely to have a significant impact on our culture. For example, major corporations such as Microsoft and Amazon are exploring the use of virtual humans in a range of business and home applications, suggesting that virtual humans increasingly will integrate with our everyday life. A valuable lesson is to be learned from the study of how television watching hinders children's ability to infer emotion in facial movements in natural settings (Coats et al., 1999). We might similarly expect the design of virtual humans to impact how we perceive emotion in one another. One might even expect the impact to be more profound and immediate than passive watching of television since virtual humans are designed to engage us in an interaction.

More specifically, virtual humans are being designed to train social skills that have significant real-world consequences, such as medical student's bedside manner (Kron et al, 2017) or a soldier's ability to cross-culturally negotiate (Kim et al, 2009; Traum et al., 2003). These systems either implicitly or explicitly seek to teach how to express emotion and infer emotions from expressions, in highly critical situations such as informing patients that they have a life-threatening disease. As part of the learning process, the system can try to assess and seek to improve the learner's ability to infer emotion from the virtual human's expression as well as to express their emotion in ways the system design deems is appropriate. Clearly, the models of expression production and recognition that are incorporated into the system will impact what is learned.

References

Alexander, O., Rogers, M., Lambeth, W., Chiang, M. and Debevec, P. (2009) Creating a Photoreal Digital Actor: The Digital Emily Project, In IEEE European Conference on Visual Media Production (CVMP).

Alfonso, B., Pynadath, D., Lhommet, M. and Marsella, S. (2015), "Emotional Perception for Updating Agents' Beliefs", in proceedings of the *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, IEEE, pp. 201-207.

Becker-Asano, C. and Wachsmuth, I. Affect simulation with primary and secondary emotions. In Proceedings of the Eighth International Conference on Intelligent Virtual Agents (Tokyo). Springer, 2008, 15–28.

Cafaro, A., Pelachaud, C. & Marsella. S. (2017) Nonverbal behavior in multimodal performances, editor Phil Cohen, Springer.

Cassell, J., Sullivan, J., Prevost, S. & Churchill, E., Eds. (2000). Embodied Conversational Agents. Cambridge, MA, MIT Press.

Coats, E. J., Feldman, R. S., & Philippot , P. (1999). The influence of television on children's nonverbal behavior. In P. Philippot & R. S. Feldman(Eds.), The social context of nonverbal behavior (pp. 156–181). New York: Cambridge University Press.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, GM., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, Suri, A., Traum, D., Wood, R., Xu, Y. Rizzo, S. and Morency, LP. (2014) "SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support", in The 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS).

Dias, J. and Paiva, A. (2005) Feeling and reasoning: A computational model for emotional agents. In Proceedings of the 12th Portuguese Conference on Artificial Intelligence (Covilhã, Portugal). Springer, 127–140.

Ekman, P. (1979). About Brows: emotional and conversational signals. In von Cranach, Foppa, Lepenies & Ploog (eds) Human ethology, Cambridge University Press.

Gratch, J., Marsella. S., Wang, N. and Stankovic, B. (2009) Assessing the validity of appraisal-based models of emotion, in Proceedings of Affective Computing and Intelligent Interaction, ACII, (Amsterdam, Netherlands.

Hoyt, C., Blascovich, J., & Swinth, K. (2003). Social inhibition in immersive virtual environments. *Presence*, *12*(2), 183-195.

Huang, L., Morency, L.P. and Gratch. J. (2011). Virtual rapport 2.0. In Proceedings of the 10th international conference on Intelligent virtual agents (IVA'11), Hannes Högni Vilhjálmsson, Stefan Kopp, Stacy Marsella, and Kristinn R. Thórisson (Eds.). Springer-Verlag, Berlin, Heidelberg, 68-79.

Kim, J., R. Hill, P. Durlach, H. Lane, E. Forbell, M. Core, S. Marsella, D. Pynadath & Hart, J. (2009). "BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. Special Issue on Ill-Defined Domains", International Journal of AI in Education, vol. 19, no. 3, pp. 289-308

Kron, et al. (2017) Using a computer simulation for teaching communication skills: A blinded multisite mixed methods randomized controlled trial. Patient Education and Counseling 100 748–759.

Lester, J., Converse, S., Kahler, S., Barlow, S.T., Stone, B. & Bhogal, R. (1997). The persona effect: affective impact of animated pedagogical agents. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97). ACM, New York, NY, USA, 359-366.

Malatesta, L., Raouzaiou, A., Karpouzis, K., Kollias, S.D. (2009). Towards modeling embodied conversational agent character profiles using appraisal theory predictions in expression synthesis. Applied Intelligence. 30(1), 58-64.

Marsella, S., Johnson, W.L., and LaBore, C. 2000. Interactive pedagogical drama. In Proceedings of the fourth international conference on Autonomous agents (AGENTS '00). ACM, New York, NY, USA, 301-308.

Marsella, S. and Gratch, J. , Computational Models of Emotion as Psychological Tools. In Feldman Barrett, Lewis and Haviliand-Jones (eds) Handbook of Emotions, Guilford, 2016.

Marsella, S. and Gratch, J. EMA: A process model of appraisal dynamics. Journal of Cognitive Systems Research 10, 1 (Mar. 2009), 70–90.

Marsella, S., Gratch, J., Wang, N. and Stankovic, B. (2009) Assessing the validity of a computational model of emotional coping, in Proceedings of Affective Computing and Intelligent Interaction, ACII, (Amsterdam, Netherlands).

Ortony, A., Clore, G. & Collins. (1988). The Cognitive Structure of Emotion. Cambridge University Press.

Paleari, M., Grizard, A., Lisetti, C.L. (2007). Adapting psychologically grounded facial emotional expressions to different anthropomorphic embodiment platforms. In: 20th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2007), May 7-9, 2007, Key West, USA.

Rickel, R., Marsella, S., Gratch, J., Hill, R., Traum, D. and Swartout, B. (2002) "Towards a New Generation of Virtual Humans for Interactive Experiences," in IEEE Intelligent Systems July/August, pp. 32-38.

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.), *Series in affective science. Appraisal processes in emotion: Theory, methods, research* (pp. 68-91). New York, NY, US: Oxford University Press.

Scherer, K. R., Mortillaro, M., & Mehu, M. (2017). Facial expression is driven by appraisal and generates appraisal inferences. In J-M. Fernandez-Dols & J. A. Russell (Eds.), The science of facial expression (p. 353-374). NY: Oxford U Press.

Smith, C. & Lazarus, R. (1990) “Emotion and Adaptation,” in *Handbook of Personality: Theory and Research*, L. A. Pervin, Ed., 1990, pp. 609–637.

Tickle-Degnen, L., & Rosenthal, R. (1990). The Nature of Rapport and Its Nonverbal Correlates. Psychological Inquiry, 1(4), 285-293.

Traum, D., Rickel, J., Gratch, J. & Marsella, S. (2003)."Negotiation over Tasks in Hybrid Human-Agent Teams for Simulation-Based Training.", in Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, Australia.

Wörtwein, T. & Scherer, S. (2017) What Really Matters - An Information Gain Analysis of Questions and Reactions in Automated PTSD Screenings. Proceedings of Affective Computing and Intelligent Interaction (ACII).

Yongsatianchot, N. & Marsella, S. (2016) "Integrating Model-Based Prediction and facial Expressions in the Perception of Emotion", in International Conference on Artificial General Intelligence, pp. 234-243, Springer.

Box 12: The In-Group Advantage in Emotion Perception

People's perceptions of emotion are in better agreement with scientists' expectations when judging faces from their own culture or heritage, referred to as an **in-group advantage** (Elfenbein, 2017; Elfenbein & Ambady, 2002a, 2002b, 2003), suggesting that people are better able to infer emotional states from configurations of muscle movements when they are more familiar with the structural features of the faces they are asked to label (Neth & Martinez, 2010). This suggests that perceptual learning may play an important role in perceiving emotion. Perceptual learning may be necessary for making sense of even the most basic visual details in photographs. For example, it is only when we are familiar with other people that we realize that different photos of the same person are, in fact, the same person (i.e., perceptual learning and familiarity are necessary to categorize faces by identity; 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).

References

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

Elfenbein, H. A. (2017). Emotional dialects in the language of emotion. In J-M. Fernandez-Dols & J. A. Russell (Eds.), The science of facial expression (p. 479-496). NY: Oxford U Press.

Elfenbein, H. A., & Ambady, N. (2002). Is there an in-group advantage in emotion recognition? Psychol Bull, 128(2), 243-249.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. Psychological Bulletin, 128(2), 203-235.

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. Current Directions in Psychological Science, 12(5), 159-164.

Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. Cognition, 121(3), 313-323. doi:10.1016/j.cognition.2011.08.001

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(3), 573-599

Neth, D., & Martinez, A. M. (2010). A computational shape-based model of anger and sadness justifies a configural representation of faces. Vision research, 50(17), 1693-1711.

Viviani, P., Binda, P., & Borsato, T. (2007). Categorical perception of newly learned faces. Visual Cognition, 15(4), 420-467.

Young, A. W., & Burton, A. M. (2018). Are We Face Experts? Trends Cogn Sci, 22(2), 100-110. doi:10.1016/j.tics.2017.11.007

Box 13: Some Details of the Emotion Perception Studies by Crivelli and Colleagues

Crivelli and colleagues have published three articles examining emotion perception in small-scale, remote cultural contexts that are relevant to this paper. If we only pay attention to the *reliability* with which participants infer an emotional cause for the facial configurations in Figure 4, then it appears that several of these experiments provide moderate support for the commonsense hypothesis that anger, disgust, fear, happiness, sadness and surprise each are expressed with the facial configurations in Figure 4. But a closer examination of the specificity, in addition to reliability, makes it clear that they do not.

1. 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(7)*, 830-843.

In Study 1 of Crivelli, Jarillo, et al. (2016), photographs of six posed facial configurations similar to those in Figure 4 of the main paper were spread before participants from the Trobriand Islands; participants were asked to point to the person who felt a specific emotion (so, this was a choice-from-array task: matching facial configurations to words).

	Reliability	**Specificity**
Smiling	Labeled as happiness (.58 [.39, .74])	Labeled as anger (.20 [.09, .38])
Pouting	Labeled as sadness (.46 [.29, .65])	Labeled as fear (.27 [.13, .46])
Gasping	Labeled as fear (.31 [.16, .50])	Labeled as anger (.30), disgust (.29)
Nose-scrunch	Labeled as disgust (.25)	Labeled as fear (.27), sad (.23), anger (.20)
Scowling	Labeled as anger (.07)	Labeled as disgust (.08)

Results from Table 1 of Crivelli et al., Study 1. Facial configurations are listed in the first column. 95% confidence intervals (CIs) presented in square brackets.

There is only above chance reliability and specificity for smiling as an expression of happiness. This result has an untested alternative explanation: participants were able to pick the correct label based on valence alone (smiling faces depict a pleasant state and the only pleasant word offered was happiness). For a discussion, see main text.

In Crivelli, Jarillo et al. (2016) Study 2, the method was the same (matching faces to words), except that some participants saw posed static photographs of facial configurations and some saw posed dynamic videos. Participants were the Mwani of Mozambique and the experiment was conducted in their native language (Kimwani). There was no difference between the static and dynamic conditions. Only the static condition is discussed here for brevity.

	Reliability	**Specificity**
Smiling	Labeled as happiness (.58)	Labeled as sadness (.21)
Pouting	Labeled as sadness (.16)	Labeled as anger (.16), fear (.16), disgust (.32)
Gasping	Labeled as fear (.58)	Labeled as anger (.21)
Nose-scrunch	Labeled as disgust (.37)	Labeled as sadness (.16), anger (.32), fear (.11)
Scowling	Labeled as anger (.26)	Labeled as happiness (.11), sadness (.16), fear (.11), disgust (.11)

Results from Table 5, of Crivelli et al., Study 2. Facial configurations are listed in the first column. The Specificity column lists results that are not statistically different from those in the Reliability column.

No facial configuration was labeled specifically as hypothesized. The results for the dynamic facial configurations was similar, except that the smiling configuration was specifically labeled as happiness, which is tempered by an alternative explanation (perceiving valence).

2. 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*, 113(44), 12403-12407.

In this study, Trobriand Islanders were asked to point to the posed facial configuration of a person who felt a specific emotion (in the emotion condition) or who was communicating a specific social motive (in the social motive condition). Participants could select a facial configuration from an array of six, pick a card with a black cross meaning “face not present in the array,” or answer that they did not know the response.

	Reliability	Specificity
Smiling	Labeled as happiness (1.0)	
Pouting	Labeled as sadness (.53 [.15, .57])	Labeled as hunger (.31), fear (.25), submission (.25)
Gasping	Labeled as fear (.39)	Labeled as anger (.56), threat (.69)
Nose-scrunch	Labeled as disgust (.22)	Labeled as fear (.28), rejection (.56)
Scowling	Labeled as anger (.06)	Labeled as disgust (.36), hunger (.14), help (.19), submission (.14), rejection (.22), about to eat (.19)

Results from Table S1, of Crivelli et al. Facial configurations are listed in the first column. The Specificity column lists results that are not statistically different from those in the Reliability column.

There was strong reliability and above chance specificity for smiling as an expression of happiness, which is tempered by an alternative explanation (perception of valence).

3. 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(2), 337-347.

In Study 1, Trobriand Islanders were asked to freely label spontaneous expressions of happiness, sadness, anger, surprise and disgust produced by the Fore people of Papua New Guinea; the photographs were labeled by and published by Ekman (1980).

	Reliability	Specificity
Smiling	Labeled as happiness (.13 [.04, .29])	Labeled as laughing or smiling (.44)
Pouting	Labeled as sadness (.16 [.06, .32]	Labeled as avoidance (.19)
Startled	Labeled as surprise (.00)	Labeled as avoidance (.19), sadness (.16)
Nose-scrunch	Labeled as disgust (.06 [.01, .21]),	Labeled as avoidance (.22)
Scowling	Labeled as anger (.03 [.00, .17])	Labeled as avoidance (.56)

Results from Table 1, of Crivelli et al., Study 1. Facial configurations are listed in the first column. The Specificity column lists results that are not statistically different from those in the Reliability column.

No spontaneous expression was labeled as predicted by Ekman (1980) or by the commonsense view.

In Study 2 of Crivelli at al. (2017), Trobriand Islanders matched posed facial configurations to emotion labels that were provided by the experimenter (i.e., they performed the matching faces to words method that was used in Crivelli, Jarillo et al., 2016). According to the paper, "The response format consisted of nine written terms. Five of the labels were predicted by Ekman (1980): *mwasawa* (happiness), *ninamwau* (sadness), *leya* (anger), *eyowa lopola* (surprise, startle), and *minena* (disgust). Two of the labels were Study 1's modal categories: *gigila* (laughing, smiling) and *gibulwa* (feels like avoiding social interaction). And, two of the remaining labels were *itwali* (other emotion) and *gala anukwali* (I do not know). On the actual questionnaire, only the Kilivila terms were listed. The items were always presented in the same order: *gilbuwa*, *ninamwau/mwau*, *minena*, *eyowa lopola*, *gigila*, *leya*, *mwasawa*, *itwali*, and *gala anukwali*."

	Reliability	**Specificity**
Smiling	Labeled as happiness (.17 [.06, .37])	Labeled as laughing, smiling (.69), sadness (.13)
Pouting	Labeled as sadness (.29 [.15, .49])	Labeled as anger (.17), surprise (.17)
Startled	Labeled as surprise (.21 [.09, .41])	Labeled as sadness (.21), happiness (.13), avoidance (.17)
Nose-scrunch	Labeled as disgust (.38 [.21, .57])	Labeled as avoidance (.33)
Scowling	Labeled as anger (.13 [.04, .32])	Labeled as sadness (.29), avoidance (.50)

Results from Table 3 of Crivelli et al., Study 2. Facial configurations are listed in the first column. The Specificity column lists results that are not statistically different from those in the Reliability column.

No spontaneous expression was labeled as predicted by the commonsense view. The smiling configuration was consistently chosen as the expression for "happy" just barely above chance, it was also labeled sadness as frequently. And the smiling configuration was also labeled as laughter more reliably than it was labeled "happy," replicating Study 1. Nor did a pouting configuration meet the specificity criterion; proportionally, participants were just as likely to label a pouting face as anger or surprise. The startled configuration (proposed as the surprise expression) was as reliably labeled as sadness, happiness or "feels like avoiding a situation" as it was labeled surprise. A nose-wrinkled configuration (the proposed disgust expression) was consistently labeled as "disgust" better than what would be expected by chance, but it was not specifically (uniquely) labeled as “disgust,” as it was just as often labeled as “feels like avoiding a social interaction.”

Box 14: The Power of Words in Emotion Perception Experiments

The emotion words provided during choice-from-array tasks may have a psychological impact that extends beyond just constraining participants' word choice – they may actually help to create reliable emotion inferences.

Words Support Perception

In people who already possess conceptual knowledge for U.S. concepts of anger, fear, and so on, emotion words appear to encourage participants to see certain emotions in the facial configurations of Figure 4 more so than they would otherwise. Many experiments now show that words have a more basic function in supporting perception, even for unfamiliar objects (Lupyan, Rakison & McClelland, 2007), contradicting the widespread assumption that people simply learn names for categories they already know, and supporting the hypothesis that words shape how categories are learned in the first place (Gelman & Roberts, 2017; Waxman & Gelman, 2010). Consistent with these broader findings, emotion words have been shown to shape how participants perceive and even literally *see* faces (Gendron et al., 2012) because they influence how people encode and remember facial features (Fugate et al., 2010, 2017; Doyle & Lindquist, 2018). Additional evidence also shows that the conceptual knowledge linked to words dynamically shapes the perception of facial configurations: when participants believe that two emotion categories are conceptually more similar to one another, facial configurations depicting those categories were also perceived as more similar, even when controlling for the actual perceptual similarity of the facial configurations (using novel research methods like reverse correlation and computer mouse tracking; Brooks & Freeman, 2018).

Words Invite Concept Learning

Research shows that the words provided in choice-from-array tasks may actually quickly *teach* participants the expected answers in an experiment. Participants in an experiment might see anywhere from a dozen to a hundred proposed expressive configurations, and for each one they see, the same handful of emotion words are presented over and over again. Under these conditions, participants quickly learn which words are supposed to correspond to each facial configuration. For example, children learn to label an artificially constructed facial expression (e.g., a blowfish expression) with the word “pax” in a choice from array task at levels that are comparable to the proposed expressive configurations in Figure 4 (Nelson & Russell, 2016). Participants also use a process of elimination strategy: words that are not chosen on prior trials are selected more frequently, inflating agreement levels (DiGirolamo & Russell, 2017).

There is some evidence that choice-from-array tasks inadvertently allow participants to learn emotion categories during the course of an experiment. In a recent study, emotion categories that are untranslatable with a single word in English, in Mandarin Chinese and in Hadza culture (and that do not exist in those cultures) were presented to participants from those cultures in a choice-from-array task along with contrived cues (in this case, made up vocalizations). Participants free-labeling of the vocalizations indicated that they were unfamiliar with them; they did not label the vocalizations with words for the novel emotion concepts or with words for anger, disgust, fear, and so on. Nonetheless, participants labeled the vocalizations with reliability and specificity when they were offered the novel category words in a choice-from-array task, making those emotion categories and their (completely made up) vocalizations appear universal (Hoemann, Crittenden, Ruark, Gendron, & Barrett, in press; also see Gendron et al., 2015).

Emotion Words, Emotion Concepts and Emotion Perception

Developmental evidence is also consistent with the hypothesis that emotion words and associated conceptual knowledge play a powerful, and perhaps even necessary, role in emotion perception. A careful line of research provides robust, replicable evidence that children implicitly learn the affective meaning of facial movements in infancy, but only learn to explicitly infer an emotional meaning for facial configurations when they acquire the relevant emotion concept (see Box 10). These studies also suggest that emotion words play an important role in the development of emotion perception during early and middle childhood. For example, children between the ages of three and ten find it easier to match an expressive stereotype to an emotion word than to another example of the same stereotype (i.e., children find it easier to match the word "angry" to a scowling face than to perceptually match two scowling faces; see Widen, 2016). In a story-telling task, three and four-year-old children find it harder to state the cause of an expressive stereotype (e.g., a scowling face) than for an emotion word (e.g., angry) or a corresponding behavior (e.g., a scream) (Widen & Russell, 2004). This **label superiority effect** is robust and is observed in a variety of experiments (Balconi & Carrera, 2007; Camras & Allison, 1985; Reichenbach & Masters, 1983; Russell & Widen, 2002a, 2002b; Widen & Russell, 2002, 2004, 2010a, 2010b). Children between the ages of four and ten years of age find it more difficult to freely label an expression stereotype like those in Figure 4 than brief stories describing anger, fear, surprise, disgust, compassion, embarrassment, shame and contempt (where the stories do not contain any emotion label; Widen & Russell, 2010a); for example, children are more likely to freely offer the word "disgusted" (or a synonym) to label a story describing disgust than the stereotyped disgust expression (e.g., see also Camras & Allison, 1985 N. Eisenberg, murphy, shepherd, 1997; for a review see Widen & Russell, 2013a). This phenomenon has been called a **face inferiority effect**. It suggests that the ability to infer emotions emerges later for facial movements than for stories, an effect that could result from a number of factors.

Figure S8-1 may reflect more about children's' ability to incorporate facial movements into their emotion concepts, rather than their acquisition of those concepts. This interpretation is suggested by the developmental trajectory of the human visual system during these same years (Mondloch et al., 2003). Adults process the configurations in Figure 4 configurally (Martinez, 2017a; Neth & Martinez, 2009), meaning that they perceive second-order dependencies in image features (first-order dependencies involve the ordering of features (e.g., eyes on top of the nose on top of the mouth); second-order dependencies involve the relative distances between these features). There is some debate over when children become proficient in the visual interpretation of configural features (some evidence suggests at birth (Turati et al., 2010) whereas other evidence suggests around eight years of age (Mondloch et al., 2003; Le Grand et al., 2004; Maurer et al., 2002; Sinha et al., 2006)).[16] This may help to explain, at least in part, why children increasingly improve in their ability to match words to faces as they age. It may also help explain why toddlers (and even older children) have difficulty perceptually matching photos of different people who are posing the proposed expressive configurations for the same emotion category (e.g., different people scowling, different people frowning, etc.; Widen, 2016).[17]

Despite these ambiguities, it is possible that infants begin to learn emotion concepts and infer emotional meaning in facial movements earlier than they can explicitly label those movements with emotion words. One (as yet untested) hypothesis proposes that early in their development, infants hear emotion words being used by their parents and caregivers, and these emotion words serve to scaffold the ability of infants to begin to form emotion categories and learn emotion concepts (this idea is thought to operate more generally for other abstract categories and concepts; see Barrett, 2017a; Barrett et al., 2007; Lindquist & Gendron, 2013).

Words can initiate and scaffold the formation of concepts and categories (Balaban & Waxman, 1997; Waxman & Markow, 1995), particularly when the instances of a concept vary in how they look, sound and feel (as is the case with emotion concepts). These are called **abstract** or **artifact categories**. A growing body of research shows that infants and toddlers use words as

---

[16] Experience with faces allows children to more quickly learn how to interpret the meaning of facial movements (Oakes & Ellis, 2013); scientists speculate that this experience allows children to differentiate facial information into categories that are functional in their social environments.

[17] It has been suggested that children have a difficult time perceptually matching faces because they are limited in their ability to process faces configurally (Pascalis et al., 2002). Configural processing is necessary to gain expertise in face recognition (Maurer et al., 2002) and plays an important role in the visual perception of emotion (Neth & Martinez, 2009).

a powerful tool for learning artifact or nominal kind categories (i.e., objects are treated as similar for performing some function when they are perceptually dissimilar; e.g., Fulkerson & Waxman, 2007; Landau & Shipley, 2001; Plunkett et al. 2008; Addyman & Mareschal, 2010; Althaus & Westerman, 2016; Baldwin & Markman, 1989; Dewar & Xu, 2007, 2009; Welder & Graham, 2006). It has been proposed that emotion words help children form emotion categories and learn emotion concepts precisely because emotion concepts are abstract concepts (Barrett, 2017; Barrett et al., 2007; Lindquist & Gendron, 2013).

If variation is the norm when it comes to emotion categories, then emotion categories are abstract categories: people can tremble, jump, freeze, scream, gasp, hide, attack, and even laugh in the face of fear; the same appears to be true for anger. Physiological changes such as heart rate, breathing rate or blood pressure increase, decrease, or stay the same across instances of all emotion categories that have been studied (see Box 9). The variation is not random – it is situated -- and it is beyond what can be accounted for by common beliefs about emotion (see Box 2). The fact that the instances of an emotion category, like fear, can vary considerably in their facial movements, their physical changes, and their behaviors implies that when it comes to learning emotion concepts, children are faced with the task of making inferences about deeper commonalities across perceptually variable instances (i.e., they must learn a nominal kind category). That is, they must learn that instances of fear belong to the same category because they serve the same purpose, even if those instances look, sound, and feel different.

Evidence from Congenitally Deaf Children

Another opportunity to study the role of emotion words and their associated concepts in emotion perception comes from observing children who are born deaf. Congenitally deaf children are often born to hearing parents who do not know sign language and who may subsequently struggle to learn it, reducing crucial opportunities to communicate with their infants early in life. Communication with parents and caregivers is the basis of emotion concept learning (Harris, de Rosnay, & Pons, 2016) and language learning (Kuhl, 2014) more generally, suggesting that congenitally deaf children who are born to non-signing, hearing parents might be slower to learn mental words and concepts. This may offer a window to observe the effects of early exposure to language, or its delay, on emotion perception competency. Numerous studies now show that congenitally deaf children who are born to non-signing, hearing parents have no opportunities to engage in or hear conversations about emotion or benefit from emotion labeling. These children are, in fact, slower to learn mental words and concepts because of this lack of access (e.g., Levrez et al., 2012; Remmel & Peters, 2008; Russell et al., 1998; Schick et al., 2007; Steeds, Rowe, & Dowker, 1997). A variety of studies do, in fact, suggest that deaf children have a difficult time inferring emotions from scowls, frowns, smiles and so on, when compared to hearing children or children who are raised by parents who are fluent in sign language (for a review of evidence, see Sidera et al., 2017). This difficulty extends to inferring mental causes for physical movements, more generally (e.g., Ludlow et al., 2012).

Some scientists working in this area hypothesize that deaf children's difficulty inferring emotion from facial movements is an example of their larger difficulties inferring mental events in general, primarily due to a delay in their ability to learn language (e.g., Dyck & Denver, 2003; Ludlow et al., 2010; Schick et al., 2007; Spencer & Marschark, 2010; Walker-Andrews & Lennon, 1991). Hearing children who are delayed in learning language also have difficulties inferring emotion from facial movements (Nelson et al., 2011). Taken together, then, this research is consistent with the hypothesis that emotion perception competency emerges in the context of word and concept learning.

Summary

Taken together, these findings suggest that emotion words are not psychologically inert – they may shape how emotion is inferred in facial movements and encourage participants to assign emotional meaning to facial configurations differently than they would if the words were not present. Since most of the studies that support the common view of emotion perception are choice-from-array tasks that include emotion words, the potency of those words provides an alternative explanation for the hundreds of studies that seem to strongly support the hypothesis that people perceive specific emotions in specific facial configurations with reliability.

References

Addyman, C., & Mareschal, D. (2010). The perceptual origins of the abstract same/different concept in human infants. Animal cognition, 13(6), 817-833.

Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. Journal of experimental child psychology, 151, 5-17.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? Journal of experimental child psychology, 64(1), 3-26.

Balconi, M., & Carrera, A. (2007). Emotional representation in facial expression and script: A comparison between normal and autistic children. Research in developmental disabilities, 28(4), 409-422.

Baldwin, D. A., & Markman, E. M. (1989). Establishing word-object relations: A first step. Child development, 381-398.

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. Trends in cognitive sciences, 11(8), 327-332.

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

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behavior*, https://doi.org/10.1038/s41562-018-0376-6

Camras, L. A., & Allison, K. (1985). Children's understanding of emotional facial expressions and verbal labels. Journal of nonverbal Behavior, 9(2), 84-94.

Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. Psychological Science, 20(2), 252-257.

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(3), 538-546. doi:10.1037/emo0000247

Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. J Exp Psychol Gen, 147(1), 62-73. doi:10.1037/xge0000361

Dyck, M. J., & Denver, E. (2003). Can the Emotion Recognition Ability of Deaf Children be Enhanced? A Pilot Study. J Deaf Stud Deaf Educ, 8(3), 348-356. doi:10.1093/deafed/eng019

Eisenberg, N., Murphy, B. C., & Shepard, S. (1997). The development of empathic accuracy. Empathic accuracy, 73-116.

Fugate, J. M. B., Gendron, M., Nakashima, S., & Barrett, L. F. (2017). Emotion words: Adding face value. Emotion

Fugate, J., Gouzoules, H., & Barrett, L. F. (2010). Reading chimpanzee faces: evidence for the role of verbal labels in categorical perception of emotion. Emotion, 10(4), 544-554.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6- and 12-month-olds. Cognition, 105(1), 218-228.

Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences*, *114*(30), 7900-7907.

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

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.

Harris, P. L., de Rosnay, M., & Pons, F. (2016). Understanding emotion. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), Handbook of Emotion, 4th Edition. New York: Guilford.

Hoemann, K., Crittenden, A. N., Ruark, G., Gendron, M., & Barrett, L. F. (in press). Context facilitates the cross-cultural perception of emotion. *Emotion*.

Hoemann, K., Xu, Fei, & Barrett, L. F. (in press). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology*.

Kuhl, P. K. (2014). Early language learning and the social brain. Cold Spring Harbor symposia on quantitative biology, 79, 211-220.

Landau, B., & Shipley, E. (2001). Labelling patterns and object naming. Developmental science, 4(1), 109-118.

Le Grand, R. L., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. Psychological Science, 15(11), 762-768.

Levrez, C., Bourdin, B., Le Driant, B., & Baudouin, F. D. (2012). The impact of verbal capacity on theory of mind in deaf and hard of hearing children. American Annals of the Deaf, 157, 66-77.

Lindquist, K.A. & Gendron, M. (2013). What's in a word? Language Constructs emotion perception. Emotion Review, 5, 66-71.

Ludlow, A. K., Heaton, P., & Deruelle, C. (2013). Decoding Actions and Emotions in Deaf Children: Evidence From a Biological Motion Task. Journal of Cognition and Development, 14(4), 561-572. doi:10.1080/15248372.2012.698434

Ludlow, A., Heaton, P., Rosset, D., Hills, P., & Deruelle, C. (2010). Emotion recognition in children with profound and severe deafness: do they have a deficit in perceptual processing? J Clin Exp Neuropsychol, 32(9), 923-928. doi:10.1080/13803391003596447

Lupyan, G., Rakison D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. Psychological Science, 18, 1077-1083.

Martinez, A. M. (2017a). Computational models of face perception. Current Directions in Psychological Science, 26(3), 263-269.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. Trends in cognitive sciences, 6(6), 255-260.

Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. Journal of experimental child psychology, 86(1), 67-84.

Nelson, N. L., & Russell, J. A. (2016). Building emotion categories: Children use a process of elimination when they encounter novel expressions. J Exp Child Psychol, 151, 120-130. doi:10.1016/j.jecp.2016.02.012

Nelson, K. E., Welsh, J. A., Trup, E. M. V., & Greenberg, M. T. (2011). Language delays of impoverished preschool children in relation to early academic and emotion recognition skills. First Language, 31, 164–194.

Neth, D., & Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. Journal of vision, 9(1), 5.

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(1), 134-148.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life?. Science, 296(5571), 1321-1323.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. Cognition, 106(2), 665-681.

Reichenbach, L., & Masters, J. C. (1983). Children's use of expressive and contextual cues in judgments of emotion. Child Development, 993-1004.

Remmel, E., & Peters, K. (2008). Theory of mind and language in children with cochlear implants. Journal of Deaf Studies and Deaf Education, 14, 218–236.

Russell, J. A., & Widen, S. C. (2002a). A label superiority effect in children's categorization of facial expressions. Social Development, 11(1), 30-52.

Russell, J. A., & Widen, S. C. (2002b). Words versus faces in evoking preschool children's knowledge of the causes of emotions. International Journal of Behavioral Development, 26(2), 97-103.

Russell, J. A., Hosie, C. D., Gray, C., Scott, C., Hunter, N., Banks, J. S., & Macaualy, M. C. (1998). The development of theory of mind in deaf children. Journal of Child Psychology and Psychiatry, 39, 903–910.

Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. Child Development, 78, 376–396.

Sidera, F., Amado, A., & Martinez, L. (2017). Influences on facial Emotion Recognition in Deaf Children. J Deaf Stud Deaf Educ, 22(2), 164-177. doi:10.1093/deafed/enw072

Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. Proceedings of the IEEE, 94(11), 1948-1962.

Spencer, P. E., & Marschark, M. (2010) Evidence-based practice in educating deaf and hard-of-hearing students. New York, NY: Oxford University Press.

Steeds, L., Rowe, K., & Dowker, A. (1997). Deaf children's understanding of beliefs and desires. Journal of Deaf Studies and Deaf Education, 185-195.

Turati, C., Di Giorgio, E., Bardi, L., & Simion, F. (2010). Holistic face processing in newborns, 3-month-old infants and adults: Evidence from the composite face effect. Child Development, 81, 1894-1905.

Walker-Andrews, A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. Infant Behavior and Development, 14(2), 131-142.

Waxman, S. R., & Gelman, S. A. (2010). Different kinds of concepts and different kinds of words: What words do for human cognition. In D. Mareschal, P. C. Quinn & S. E. G. Lea (Eds), The making of human concepts (p. 101-130). New York Oxford University Press.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. Cognitive psychology, 29(3), 257-302.

Welder, A. N., & Graham, S. A. (2006). Infants’ categorization of novel objects with more or less obvious features. Cognitive psychology, 52(1), 57-91.

Widen, S. C., & Russell, J. A. (2002). Gender and preschoolers' perception of emotion. *Merrill-Palmer Quarterly, 48*(3), 248-262. doi: 10.1353/mpq.2002.0013

Widen, S. C., & Russell, J. A. (2004). The relative power of an emotion's facial expression, label, and behavioral consequence to evoke preschoolers' knowledge of its cause. *Cognitive Development, 19*(1), 111-125. doi: 10.1016/j.cogdev.2003.11.004

Widen, S. C., & Russell, J. A. (2010a). Children's scripts for social emotions: Causes and consequences are more central than are facial expressions. *British Journal of Developmental Psychology, 28*(3), 565-581. doi: 10.1348/026151009X457550d

Widen, S. C., & Russell, J. A. (2010b). Differentiation in preschooler's categories of emotion. *Emotion, 10*(5), 651-661. doi: 10.1037/a0019005

Widen, S. C. (2016). The development of children's concepts of emotion. In L. F. Barrett, M. Lewis, J. M., & Haviland-Jones (Eds.), *Handbook of emotions* (4th ed., pp. 307-318). New York: Guilford Press.

Box 15: The Habituation Task Used in Studies of Emotion Perception in Infants

The most popular experimental design that is used to study emotion perception in preverbal infants is called the **habituation task**. In a standard habituation task, an infant is presented with a stimulus that experimenters believe represents an emotion category (say, a scowling face to represent anger) and the infant will look at it. As subsequent stimuli are presented (say, a series of scowling faces), the infant is assumed to identify or categorize each face according to its emotional meaning. If the infant categorizes the stimuli as belonging to the same emotion category, it will look for a shorter amount of time (because the infant is presumed to become bored). Once the infant's looking time drops below a certain threshold, experimenters assume that the child is habituated to the emotion category (i.e., the infant has become uninterested in looking at "scowling faces"). Then, a new stimulus is presented, and looking time is again recorded (e.g., after a viewing a series of scowling faces, the infant is then shown a smiling face). If the infant looks for a *longer* time, scientists infer that she has categorized the stimulus as belonging to an emotion category that is different from before (i.e., the infant is assumed to be interested in novelty; this novelty is supposed to reflect a category difference in the context of the experiment).

The habituation task obviously requires that the experimenter infer what looking times mean. Such inferences call for having strong alternative hypotheses that are, in practice, rarely considered. For example, the proposed expressive forms in Figure 4 differ in their familiarity (e.g., most infants are more familiar with smiling faces than with scowls or frowns). This makes it difficult to know which features of a face are holding an infant's attention (familiarity or novelty), and can lead to potentially incorrect inferences. For example, several studies claim that infants are somehow born prepared to detect fearful faces. But the proposed expression for fear is less familiar than "happy" faces; infants may look longer at them because they are attempting to learn novel stimuli (e.g., Bayet et al., 2017; see also Peltola et al., 2008, 2009).

References:

Bayet, L., Quinn, P. C., Laboissière, R., Caldara, R., Lee, K., & Pascalis, O. (2017). Fearful but not happy expressions boost face detection in human infants. *Proc. R. Soc. B*, *284*(1862), 20171054.

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*(1), 60-68.

Peltola, M. J., Leppänen, J. M., Vogel-Farley, V. K., Hietanen, J. K., & Nelson, C. A. (2009). Fearful faces but not fearful eyes alone delay attention disengagement in 7-month-old infants. *Emotion*, *9*(4), 560.

Box 16: Information Theory as Applied to Emotion Communication

An information theory approach to understanding emotional communication asks: what is the information that one person's facial movements (the sender) convey to another person (the perceiver) in a particular context and how does that information help achieve synchrony between the sender and perceiver. The questions are cast in terms of how the sender and perceiver share information. The sender is communicating some information with his facial movements (sometimes emotional, sometimes not), and the perceiver uses that information to reduce uncertainty about what the sender is going to do next (in Western cultures, this typically means making a mental inference that allows the perceiver to predict the sender's actions). For recent examples of experiments that use information theory as a scientific tool, see Jack & Schyns, (2017).

This approach begins with the assumption that one person, the sender, is in some state of mind, makes some facial movements, and performs some action (speaks, moves his body, etc.) in a particular situation. The perceiver detects these movements and makes a prediction about what the sender will do next, often by making an inference about the sender's state of mind. Communication of information between the sender and the perceiver is understood as mutual information, which is a symmetric measure of the information shared between two variables, X and Y. Specifically, mutual information is a measure of how much knowing value of one variable, X, tell us, reduces the uncertainty, about the value of the other variable, Y, and vice versa.

Formally, mutual Information is a symmetric measure of the dependency between two random variables that provides a nonparametric way to assess how much knowing some random variable, Y, reduces the uncertainty about another random variable, X. Mutual information is based on Shannon's notion of information which is measured by entropy, $H(X)$, a measure of the uncertainty of a random variable. Mutual Information, $I(X;Y)$, is the reduction in uncertainty that results from knowing Y, I(X;Y) = H(X)-H(X|Y) (see Cover & Thomas 2006). If X and Y are independent events, knowing Y provides no information about X and the entropy H(X) is equal to the entropy H(X,Y) and I(X;Y) is zero. On the other hand, if knowing Y fully determines X and Y fully determines X, then the mutual information reduces to the entropy of X (or Y).

When asking whether facial movements express an instance of a certain emotion category emotions, where X is the sender's state of mind and Y is his facial movements, the forward inference is as follows: How much does X, his state of mind, reduce uncertainty about Y, his facial movements, perhaps measured with facial EMG? In the reverse inference, X is the sender's facial movements and Y is his state of mind: How much does X, his facial movements, again measured with facial EMG, reduce uncertainty about his state of mind, Y?)

References

Chen, C., Crivelli, C., Garrod, O., Schyns, P., Fernández-Dols, J. & Jack, R. (2018) Distinct facial expressions represent pain and pleasure across cultures. Proceedings of the National Academy of Sciences Oct 2018, 115 (43)

Jack, R. and Schyns, P. (2017) Toward a Social Psychophysics of Face Communication Annual Review of Psychology 2017 68:1, 269-297

Thomas M. Cover and Joy A. Thomas. (2006). Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, New York, NY, USA.