What’s in a Word? Language Constructs Emotion Perception

Kristen A. Lindquist
Department of Psychology, University of North Carolina, USA

Maria Gendron
Department of Psychology, Boston College, USA
Department of Psychology, Northeastern University, USA

Abstract

In this review, we highlight evidence suggesting that concepts represented in language are used to create a perception of emotion from the constant ebb and flow of other people’s facial muscle movements. In this “construction hypothesis,” (cf. Gendron, Lindquist, Barsalou, & Barrett, 2012) (see also Barrett, 2006b; Barrett, Lindquist, & Gendron, 2007; Barrett, Mesquita, & Gendron, 2011), language plays a constitutive role in emotion perception because words ground the otherwise highly variable instances of an emotion category. We demonstrate that language plays a constitutive role in emotion perception by discussing findings from behavior, neuropsychology, development, and neuroimaging. We close by discussing implications of a constructionist view for the science of emotion.

Keywords
construction, context, emotion perception, language

Although the ability to correctly name a facial expression is generally low, the meaning of it is readily seen when its true name is given.
Floyd Allport (1924, p. 255)

Common sense tells us that emotions are expressed on the face and easily decoded by a perceiver without the use of language. Yet growing evidence suggests that emotion words help a perceiver understand the meaning of another person’s facial muscle movements. In this review, we demonstrate that language plays a constitutive role in emotion perception, even in psychological studies that use artificial, posed, and caricatured pictures of facial muscle movements as stimuli. We begin by situating this hypothesis in the relevant theoretical context by introducing two competing theories on emotion perception: the basic emotion and constructionist views. We next demonstrate that there is a paradox in the emotion literature: People report seeing emotion on others’ faces, yet counter to the predictions of a basic emotion view, it is far from clear that others reliably produce certain facial muscle movements for certain emotions. This leaves open the possibility that emotion perceptions are constructed in the mind of a perceiver when concept knowledge grounded by language is brought to bear to make meaning of someone’s facial muscle movements in context. We review evidence in favor of this constructionist view that language helps construct emotion perception by discussing findings from behavior, neuropsychology, development, and neuroimaging.

What’s in a Face?

The commonsense view that emotion perception is automatically given by information on the face is formalized in the psychological literature as the basic emotion approach (for recent reviews, see Ekman & Cordaro, 2011; Izard, 2011; Shariff & Tracy, 2011). This view hypothesizes that certain combinations of facial muscle movements code emotion in a specific and consistent manner. Researchers often use the term “facial expression” (cf. Darwin, 1872/1965) to refer to facial muscle movements because they assume that the movements observed are the emotion seeking expression on the face (see Russell, Bachorowski, & Fernández-Dols, 2003). In this view, emotion perception merely requires that a perceiver “decode” the information that is “encoded” in someone else’s facial muscle movements. This process is assumed to occur
automatically and without any effort on the part of the perceiver. Consistent with the basic emotion view, studies demonstrate that people can perceive emotion on others’ faces relatively quickly (e.g., Tracy & Robins, 2008) and to some degree across cultural contexts (Ekman et al., 1987). These data are often interpreted as evidence for the automaticity and universality of emotion production and perception, and it is concluded that the categories of emotion expressed and perceived on others’ faces are biologically innate.

Yet increasing data from a variety of domains call the basic emotion view into question by demonstrating that people do not consistently produce the specific configurations of facial muscle movements predicted by a basic emotion model. For instance, the facial electromyography (EMG) literature does not find evidence for prototypical patterns of facial muscle movements that distinguish between discrete emotions (e.g., EMG cannot reliably distinguish facial muscle patterns for anger vs. sadness; Cacioppo, Bernston, Larsen, Poehlmann, & Ito, 2000; for discussions, see Barrett, 2006a; Mauss & Robinson, 2009). Congenitally blind infants (Fraiberg & Fraiberg, 1977), children (Roch-Levecq, 2006), and adults (Galati, Scherer, & Ricci-Bitti, 1997) do not make the combinations of emotional facial muscle movements that are predicted by the basic emotion view (but then again, neither do sighted adults; Galati et al., 1997; for a discussion, see Barrett, 2011; Barrett, Lindquist, & Gendron, 2007). A wealth of findings additionally demonstrate that facial muscle movements do not correspond in a 1:1 way with reported emotional experiences (e.g., Fernández-Dols & Ruiz-Belda, 1995; Fernández-Dols, Sánchez, Carrera, & Ruiz-Belda, 1997; for a review, see Russell et al., 2003). Sometimes people report an emotion experience (e.g., happiness), but do not make the facial muscle movements hypothesized to be specific for that emotion (e.g., they do not smile) (Fernández-Dols & Ruiz-Belda, 1995). Other times, people make the facial muscle movements hypothesized to be specific for an emotion (e.g., disgust), but report the experience of multiple emotions (e.g., disgust, pain, fear) (e.g., Table 13.3 in Ekman, Frank, & Ancoli, 1980; see Matsumoto, Keltner, Shiota, O’Sullivan, & Frank, 2008, for additional data). When the combinations of facial muscle movements predicted by the basic emotion model do occur, they tend to do so only in social contexts (e.g., participants who were in a nonsocial context made combinations of facial muscle movements consistent with basic emotion hypotheses in only two out of nine studies summarized in Matsumoto et al., 2008, Table 13.2). Facial muscle movements might therefore be considered cultural symbols used in communication rather than internal cultural models. Because emotion words are explicitly available in most emotion perception experiments and implicitly available in the mind of healthy adults at all times, they can thus serve as a form of context that transforms one person’s facial muscle movements into perceptions of anger, disgust, sadness, etcetera? In this article, we argue that emotion perceptions are constructed in the mind of a perceiver when concepts represented in language help create a perception of emotion from the constant ebb and flow of other people’s facial muscle movements.

What’s in a Word?

According to our construction hypothesis (cf. Gendron, Lindquist, Barsalou, & Barrett, 2012; also see Barrett, 2006b; Barrett et al., 2007; Barrett, Mesquita, & Gendron, 2011), language plays a constitutive role in emotion perception because words ground the otherwise highly variable instances of an emotion category and are brought to bear to make meaning of facial muscle movements in a given context. Cognitive science demonstrates that in the absence of clear statistical regularities, humans use a word as the “glue” that holds perceptual instances together as members of a category (see Barsalou & Wiemer-Hastings, 2005). For instance, infants routinely use the phonological form of words to make conceptual inferences about novel objects that share little structural similarity (Dewar & Xu, 2009; Ferry, Hespós, & Waxman, 2010; Xu, 2002). We hypothesize that adults do the same thing with abstract categories like emotion. Our hypothesis is that people see instances where someone frowns at a coworker, pouts after receiving a parking ticket, seethes silently at an insult, and smiles at a misbehaving child as instances of anger because the facial muscle movements and the contexts in which they occur are all linked by the same word. Without the word “anger” to bind them, the behaviors and contexts share too few statistical regularities (Barrett, 2006a; Mauss & Robinson, 2009) to form a coherent category. Because emotion words are explicitly available in most emotion perception experiments and implicitly available in the mind of healthy adults at all times, they can thus serve as a form of context that transforms one person’s facial muscle movements into perceptions of anger, disgust, sadness, etcetera (for additional discussions of language and context in emotion perception, see Fugate, 2013; Hassin, Aviezer, & Benton, 2013; Widen, 2013).
Language Constructs Emotion Perception

Emotion Words in Standard Emotion Perception Tasks Construct Perception

Some of the clearest evidence that words are constitutive in emotion perception comes from typical emotion perception studies, which use emotion words as response options. Although response options are typically considered an innocuous feature of the task, it has been shown that including emotion words in the experiment inflates participants’ “accuracy” at identifying the emotion on the face. As it is typically used, the word “accuracy” implies that there is an unambiguous signal on a face that the perceiver correctly detects. Since we do not believe that people reliably produce unambiguous facial muscle movements that code for specific emotions, we use the term “accuracy” here to mean the agreement between what the participant reports seeing (e.g., “anger”) and what the experimenter intends the participant to see (e.g., a scowl as anger). Said another way, “accuracy” is interrater reliability between the perceiver and researcher.

In the typical emotion perception study, participants see pictures of posed facial muscle movements (such as pouts, scowls, wrinkled noses, wide eyes, and smiles) and are asked to match those pictures to the words “sad,” “anger,” “disgust,” “fear,” or “happy” (see Russell, 1994). We refer to these posed faces as “caricatures” because they are artificial and contain strong statistical regularities (e.g., all angry faces are scowling) that are not representative of the within-category variability that exists in daily life (e.g., people do not always scowl when angry). Consistent with our construction hypothesis, participants are generally better than chance at “accurately” identifying the emotion on the face when words are available in the experiment as response options (>63%; e.g., Boucher & Carlson, 1980; Izard, 1971; Kline & Johannsen, 1935; Rosenberg & Ekman, 1995; Widen, Christy, Hewett, & Russell, 2011). Studies that do not include emotion words in the task find substantially lower “accuracy” rates, however. For instance, the “accuracy” of responses is quite low when participants are asked to freely label an emotional caricature without being given a set of words to choose from (e.g., between 7.5% and 54%; Kanner, 1931; for a discussion, see Russell, 1994). One interpretation of this finding is that emotion words merely boost “accuracy” because they facilitate recognition memory for an otherwise clear emotional signal (cf. Rosenberg & Ekman, 1995). Yet if this interpretation were correct, then including additional, or “incorrect,” response options should have no effect on participants’ performance. On the contrary, “accuracy” is generally very low (e.g., 2–63% accuracy; Buzby, 1924) when participants are presented with up to 18 “incorrect” labels plus a “correct” label (i.e., the label intended by the researcher) for an emotional caricature. More strikingly, providing labels can even cause participants to perceive a face as an instance of an “incorrect” emotion (e.g., participants perceive a scowling face as “disgust” rather than “anger” when the word “disgust” is available but “anger” is not; Russell, 1993).

Manipulating Language Impairs Perception of Emotion

Consistent with the hypothesis that language is constitutive in emotion perception, a growing body of literature shows that manipulating language impairs the perception of emotion (for evidence that manipulating language produces categorical perception of discrete emotion categories, see Fugate, 2013). In a number of studies, we have impaired participants’ ability to see emotion on faces using a technique called semantic satiation, in which a word is repeated out loud 30 times until its meaning becomes temporarily inaccessible (see Black, 2004). After satiating a relevant emotion word (e.g., “anger”), participants were slower and less “accurate” to see two caricatures from the same category (e.g., two scowling faces) as perceptually similar (Lindquist, Barrett, Bliss-Moreau, & Russell, 2006). In a more recent article, we demonstrated that manipulating language impaired emotion perception in a perceptual priming task, even when the task itself did not require language or explicit categorization of the emotion. Perceptual priming is typically observed when perceivers see the same stimulus more than once; this is measured as a faster response to the stimulus on its second presentation (for a review of visual priming, see Grill-Spector, 2008). We found that reducing accessibility to the meaning of a relevant emotion word (e.g., “anger”) prior to the first perception of a caricature (e.g., scowling face) prevented that caricature from perceptually priming itself on a subsequent presentation (Gendron et al., 2012). Because this study demonstrated an effect of language in a task that does not require the categorization of emotional faces, it suggests that language has a role in perception beyond the mere labeling of faces.

Pathology in Brain Areas Associated with Language Impairs Emotion Perception

Like studies that experimentally manipulate language, studies of patients with impaired access to the meaning of emotion words show that language helps construct emotion perception. An early finding came from patient LEW, who suffered a stroke that resulted in loss of object knowledge and naming. When LEW was asked to sort photographs of emotional caricatures into piles, he produced disorganized piles that did not correspond to discrete emotion categories (Roberson, Davidoff, & Brasby, 1999). More recently, we examined the impact of semantic dementia on emotion perception (Lindquist, Gendron, Barrett, & Dickerson, 2012). Semantic dementia results in a loss of semantic knowledge due to progressive neurodegeneration in the left anterior temporal lobe (e.g., patients are no longer able to say what the word “anger” means, or to identify situations in which “anger” might occur; Lindquist, Gendron, et al., 2012). Two patients were able to match emotional caricatures to other emotional caricatures based on the perceptual features of the face, indicating intact visual perception. Yet neither patient was able to freely sort the caricatures into categories reflecting discrete emotional meaning (i.e., piles for anger, disgust, fear, happiness, sadness, and neutral caricatures). Instead, patients...
produced piles consistent with valence: piles of pleasant, unpleasant, and neutral faces. Without access to emotion concept knowledge patients could not perceive discrete emotions on faces, although they could perceive the basic valenced meaning of facial muscle movements.

Acquisition of Language Shapes Emotion Perception in Children

As opposed to patients with semantic dementia, who lose the ability to perceive emotion on faces as they lose the meaning of emotion words, children become able to perceive emotion on faces as they learn the meaning of emotion words (for reviews, see Roberson, Damjanovic, & Kikutani, 2010; Widen, 2013; Widen & Russell, 2008b). Prior to the development of language, infants are unable to perceive discrete emotions on faces, although they are able to perceive general pleasant, unpleasant, and neutral affect. For example, 5-month-old infants look longer at startled (or scowling, or pouting) faces after habituating to smiling faces (e.g., Bornstein & Arterberry, 2003), which is evidence that infants can distinguish between faces of different valence. As toddlers begin to learn emotion words, they start to construct perceptions of discrete emotions on faces. For instance, 2-year-olds only reliably use the words “sad” and “happy,” and like prelinguistic infants, can only perceive differences between unpleasant and pleasant faces (e.g., they categorize all unpleasant caricatures as “sad”). Yet around the ages of 3 and 4, children begin to reliably use the words “anger” and “fear” and become able to perceive differences between unpleasant and pleasant faces (e.g., they differentiate between sad, angry, and fearful caricatures; Widen & Russell, 2003, 2008a). Children do not learn the term “disgust” until relatively late in early childhood (mean age of 4.6 years) and accordingly, cannot reliably distinguish disgust caricatures from other unpleasant faces until later in childhood (Widen & Russell, 2003, 2008a).

Once children have learned the meaning of emotion words, including them in an experimental task improves children’s emotion perception “accuracy,” as it does for adults. Children demonstrate a “label superiority effect” in which they are more accurate at putting pictures of scowling faces in a box labeled with the word “anger” than in a box identified by a picture of another scowling face (Russell & Widen, 2002). These findings suggest that emotion words help children ignore the variability present across even caricatured facial muscle movements by cohering them into a single meaningful category.

Language Shapes the Neural Representation of Faces

Finally, findings from neuroimaging studies are consistent with the behavioral evidence that language is constitutive in the perception of emotion. For example, language alters neural representations of faces in visual cortex (Thielscher & Pessoa, 2007). These findings are interesting because if language played only a superficial role in emotion perception, changes in neural activity should occur exclusively in brain areas related to language retrieval (e.g., inferior frontal gyrus; Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997). Yet when participants perceive a neutral face as “fearful,” there is increased activity in primary visual (e.g., the calcarine fissure) and visual association cortex (e.g., superior temporal sulcus, fusiform face area) that is nearly identical to the pattern of brain activity observed when participants actually view a fearful caricature (Thielscher & Pessoa, 2007). These findings suggest that perceiving the face as an instance of “fear” literally changes how visual cortex represents that face.

Another study more directly demonstrates that the emotional content perceived on a face—rather than the features of the face itself—are reflected in activity in visual association cortex (Fox, Moon, Iaria, & Barton, 2009). Participants were asked to judge whether two sequentially presented pictures of facial muscle movements were indicative of the same emotion category (e.g., both fearful) or not (e.g., fearful and disgusted). The researchers assessed whether participants’ perceptions of emotion or the actual features of the face caused neural adaptation in visual association cortex. Neural adaptation refers to decreased regional brain activity when the same stimulus is perceived repeatedly (Grill-Spector, Henson, & Martin, 2006). Consistent with a construction hypothesis, neural adaptation occurred in posterior superior temporal sulcus (pSTS) and the fusiform face area (FFA) when participants perceived that the second picture was from the same emotion category as the first, even if the facial muscle movements were of two different emotion caricatures (e.g., a face with wide eyes and a face with a wrinkled nose). On the contrary, neural adaptation did not occur if participants perceived the faces as different emotions, even if they were in fact caricatures of the same emotion category (e.g., both faces with wide eyes) (Fox et al., 2009). These findings suggest that even activity in brain areas once thought to code for the perceptual features of the face alone reflects the linguistic emotion category perceived on the face, rather than the structural features of the face itself.

Implications

The findings we have reviewed suggest that emotion perception might not proceed passively, where emotions encoded in facial muscle movements are automatically decoded by a perceiver. Instead, emotions appear to be constructed in the minds of perceivers, and language plays an important role in constituting what is seen on another person’s face. These findings have important implications for the study of emotion. For instance, future research must address the extent to which evidence for so-called universality in emotion perception across cultures is driven by the typical laboratory procedure, in which a limited set of words are provided as response options (Ekman et al., 1987). Even studies of “abnormal” emotion perception (e.g., autism; Baron-Cohen & Wheelwright, 2004; e.g., schizophrenia; Kohler, Walker, Martin, Healey, & Moberg, 2010; e.g., Alzheimer’s disease; Phillips, Scott, Henry, Mowat, & Bell, 2010) likely inflate the degree to which patients can perceive emotions on faces by providing linguistic context in the laboratory that might not be chronically accessible to the patient in daily life.
Another question that remains outstanding is whether organisms without language perceive emotion in the same manner as healthy adult humans. Although there is evidence that infants (e.g., Bornstein & Arterberry, 2003) and nonhuman primates (e.g., Parr, Hopkins, & de Waal, 1998) perceive affect on faces, the evidence that they perceive facial muscle movements as instances of discrete emotion awaits further experimentation (for a discussion, see Barrett et al., 2007; Lindquist, Wager, Bliss-Moreau, Kober, & Barrett, 2012). Future studies in infants and nonhuman primes would need to explicitly rule out that emotion perception is driven by the ability to perceive facial muscle movements in terms of more basic affective dimensions (e.g., Russell & Bullock, 1986; Widen & Russell, 2003, 2008a), or simply by sensitivity to structural changes that are not experienced as psychologically meaningful (e.g., whether teeth are visible or not; Caron, Caron, & Myers, 1985).

Another question that remains outstanding concerns the relative contribution of language versus structural information from the face during the perception of emotion (for additional discussion, see Fugate, 2013; Hassin et al., 2013; Widen, 2013). One possibility is that contextual effects on emotion perception are constrained by the statistical regularities that are present in the facial muscle movements that occur when a person experiences a certain emotion (i.e., “emotion seeds”; cf. Aviezer, Hassin, Bentin, & Trope, 2008). Indeed, existing studies assessing the role of context in emotion perception (including our own) may be biased towards this interpretation because they use highly caricatured stimuli that contain a lot of statistical regularity. Yet another possibility is that there are not strong statistical regularities in facial muscle movements when a person experiences a certain emotion—a person is no more likely to scowl in anger than in sadness or fear. We would thus expect language to play an even more constitutive role in emotion perception outside of the lab, where caricatured facial muscle movements are rare. Of course, this hypothesis must be borne out by further research using more naturalistic stimuli.

**Conclusions**

In this article, we reviewed the existing evidence suggesting that language helps construct instances of emotion perception from the continuous ebb and flow of other people’s facial muscle movements. These findings underscore the ever-increasing recognition that emotions are psychologically constructed experiences that are made meaningful in context (Aviezer et al., 2008; Barrett, 2006a, 2006b, 2009; Barrett et al., 2011; Fernández-Dols & Carroll, 1997; Hassin et al., 2013; Lindquist, Wager, Kober, Bliss, & Barrett, 2012; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011). They also contribute to growing evidence that emotions are not natural kind categories that are given by biology—discrete categories of emotions are not evidenced as consistent and specific patterns in facial muscle movements (see Barrett, 2006a; Cacioppo et al., 2000; Mauss & Robinson, 2009; Russell et al., 2003), vocal acoustics (Russell et al., 2003), peripheral nervous system activity (Barrett, 2006a; Cacioppo et al., 2000; Mauss & Robinson, 2009), or central nervous system activity (Lindquist, Wager, Kober, et al., 2012). This leaves open the possibility, as the data reviewed here suggest, that emotions seen on other people’s faces are constructed in the mind of the perceiver.

**References**


