Published in final edited form as:

Psychol Sci. 2014 April 1; 25(4): 911–920. doi:10.1177/0956797613517239.

Cultural relativity in perceiving emotion from vocalizations

Maria Gendron^{1,2}, Debi Roberson³, Jacoba Marietta van der Vyver⁴, and Lisa Feldman Barrett^{5,6}

¹Department of Psychology, 300 McGuinn Hall, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA, USA 02467

²Affective Science Institute and Department of Psychology, 125 Nightingale Hall, 360 Northeastern University, 360 Huntington Avenue, Boston, MA, USA 02115

³Department of Psychology, Room 4.709, University of Essex, Colchester CO4 3SQ, U.K.

⁴School of Nursing and Public Health, University of Namibia, Main Campus, Block M, Room M209, Mandume Ndemufayo Avenue, Windhoek, Namibia

⁵Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA 02129

⁶Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA 02144

Abstract

A central question in the study of human behavior is whether or not certain emotions, such as anger, fear and sadness, are recognized across cultures in non-verbal cues. We predicted and found that in a concept-free experimental task, participants from an isolated cultural context (the Himba ethnic group from Northwest Namibia) do not freely label Western vocalizations with expected emotion terms. Responses indicated Himba participants perceived more basic affective properties of valence (positivity or negativity) and to some extent arousal (high or low activation). In a second concept-embedded task, we manipulated whether a given trial could be solved using only affective content or discrete emotion content based on the foil choice. Above chance accuracy in Himba participants occurred only when foils differed from targets in valence, indicating that the voice can reliably convey affective meaning across cultures, but that perceptions of emotion from the voice are culturally variable.

Think about the last time you heard someone sigh, chuckle, or groan and concluded that the person was tired, amused, or frustrated (respectively). The *universality hypothesis* (UH) states that (barring illness) all humans innately express and recognize the same emotions in non-verbal behaviors, including vocalizations. Universalist views agree that each emotion has a “fixed set of neural and bodily expressed components” (Tracy & Randles, 2011, p. 398). In strong versions of the UH, vocal cues are hypothesized to contain sufficient perceptual regularities that broadcast discrete emotion information to perceivers (Scherer, 1994; Sauter, Eisner, Ekman, & Scott, 2010). As a consequence, it is hypothesized that

emotions can be “recognized” independent of language or conceptual knowledge (Hoehl & Striano, 2010; Izard, 1994). In fact, vocalizations are thought to allow for improved “*detectability*” over facial expressions because they “can travel omni-directionally and over long distances” (Hawk, van Kleef, Fischer, & van der Schalk, 2009, p. 294). Even in a strong UH, some cultural variation in perception is expected, but the mechanisms thought to produce variability (display and decoding rules) are independent of the hypothesized innate mechanisms of expression and perception (Buck, 1984; Ekman, 1972; Dailey Matsumoto, 1989; Schimmack, 1996). In a weaker UH versions, cultural dialects for universal expressions are hypothesized (e.g., Marsh, Elfenbein & Ambady, 2003; Elfenbein, 2013). In all UH versions, however, cross cultural “recognition” levels for discrete emotion categories are expected to be *greater than chance*, even if they are not uniformly high across all cultural groups.

Of hundreds of cross-cultural experiments on emotion perception (Elfenbein & Ambady, 2002), only five provide a stringent test of the UH (see Table S1) by using a *two-culture approach* where participants are asked to decipher emotion cues from a culture with which they have limited exposure (Norenzayan & Heine, 2005). To our knowledge, only one published study has examined the UH with vocal cues in participants from a remote culture (Sauter et al., 2010). Sauter et al. tested whether Himba individuals residing in remote villages in northwestern Namibia perceived Western non-verbal vocal utterances (laughs, screams, sighs, etc.) in line with their intended “universal” emotional meaning (i.e., the Western model of “amusement”, “fear”, “relief”, and so on). On each trial, participants’ task was to select which of two vocalizations (e.g., a sigh vs. a scream) corresponded to a story about an emotional situation described with an emotion word (e.g., “Someone is suddenly faced with a dangerous animal and feels very scared”). More frequently than chance, Himba participants chose the vocalization that best fit the Western model (e.g., the scream for “fear”), leading Sauter et al. to claim support for the UH.

Despite the ubiquity of universality claims in popular and scientific circles, empirical evidence questioning the reliability of universal emotion perception steadily accumulates (for a review, see Barrett, 2011; for a review see Barrett, Mesquita, & Gendron, 2011). First, there is growing evidence for deeper cross-cultural variation in mental representations of emotion (e.g., Jack, Garrod, Yu, Caldara, & Schyns, 2012). Second, studies providing the strongest support for the UH include emotion concept cues within the task; tasks that do not prime emotion concept knowledge (by asking participants to freely label expressions rather than choose the label from provided response options) or reduce accessibility of emotion concepts (e.g., using semantic satiation) impair emotion perception, even in US participants (Boucher & Carlson, 1980; Gendron, Lindquist, Barsalou, & Barrett, 2012; Ekman & Rosenberg, 1997; Lindquist, Barrett, Bliss-Moreau, & Russell, 2006; Widen, Christy, Hewett, & Russell, 2011; see also Table S1).

The present set of experiments seek to explicitly examine the role that conceptual context plays in shaping perceptions of vocalizations across cultures. We travelled to remote part of northwestern Namibia to examine whether individuals from the Himba ethnic group (who live in villages that are relatively isolated from Western cultural practices and norms) perceive the intended emotions in Western vocal portrayals of emotion.¹ The Himba ethnic

group speaks the Otji-Herero language, which contains words that can be translated to English words for emotion (see Sauter et al., 2010).

Study 1

In Study 1, Himba and US participants completed a free-labeling task of Western (US) vocal portrayals of emotion (see Table S2). Participant-provided labels were coded as in “agreement with the presumed universal pattern” if they fit the expected emotion according to the discrete model (e.g., “angry” or “anger” for a growl) or a close synonym (e.g., “mad”). We also computed indices of valence-based agreement (e.g., “sad” is a valence-consistent label for a growl because both sadness and anger are negative states) and arousal-based agreement (e.g., “angry” is an arousal-consistent label for “woohoo” because both triumph and anger are prototypically states of high activation). We tested for agreement on affective dimensions given ample evidence that facial and vocal cues are perceived across cultures in terms of the valence and the level of arousal that they communicate (Russell, 1991; Russell, Bachorowski, & Fernandez-Dols, 2003; Russell & Barrett, 1999).

Materials and Methods

Participants

Himba: Participants were 24 native Herero speakers from the Himba ethnic group located in the remote and mountainous Northwest region of Namibia (12 male, 12 female; mean age = 35.96, SD = 14.5)² (for details see SOM).

American: Participants were 24 individuals tested at the Boston Museum of Science in Boston, MA, United States (13 male, 11 female; mean age = 38.41, SD = 18.71) (for details see SOM).

Stimuli—Stimuli were 36 non-word vocalizations (two male and two female native English speakers, each producing a vocalization to depict amusement, anger, disgust, fear, sensual pleasure, relief, sadness, surprise and *triumph*; Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009); similar to those used in Sauter et al (2010), except with *triumph* rather than *achievement* vocalizations. Participants heard a subset of 18 stimuli each (a male and female exemplar for each emotion), counterbalanced across participants. Stimuli were cleaned for ambient noise and adjusted for mean peak amplitude using Audacity (<http://audacity.sourceforge.net/>).

Procedure—All participants were tested individually. Himba participants were instructed and responded through a translator (the same translator used in Sauter et al., 2010). Participants were outfitted with headphones and verbally instructed to label the emotion they heard in the voice with a word or phrase (researchers were naive to the particular stimulus presented on a trial). The translator provided an immediate translation of the participant's

¹Note that our goal in this study was to examine the veracity of the UH using Western cues, not fully explore cultural variation in emotion perception.

²The age distribution we report for our Himba participants is not exact. While there is a numerical system in Otji-Herero, most Himba individuals did not keep track of their age. Participants either provided estimates themselves or we estimated age based on reproductive history, number and age of children (if applicable) and physical appearance.

verbal response and this translated content (or original response for American participants) was entered into the laptop computer by the experimenter. If the participant initially provided a description of a situation, behavior or bodily state, they were prompted: “Can you think of a single word to describe the feeling, the emotion?” If the participant provided a vague affective response (e.g., “good” or “bad” feeling), they were prompted: “Can you think of a more specific feeling word to describe the emotion?” Any contextual content (i.e., situational, behavioral, or physical state) provided was always recorded in addition to any mental state terms generated.

Data Coding—The data were independently coded by two trained individuals. Trials were randomized such that the coders were blind to culture and individual responder. Using criteria from Russell (1990), coders rated whether a participant's response on a given trial agreed or disagreed with the stimulus according to discrete emotion, valence and arousal. In addition to coding for agreement, both coders indicated responses where “no mental content” was available. Reliability between the two coders (Cohen's Kappa) was high for each of the sub-codes: discrete emotion = .957, valence = .943, arousal = .958. Discrepancies in coding were resolved by review and discussion among the coders and the first author. Data were analyzed by comparing the mean percent agreement (for the discrete emotion, valence and arousal responses) against zero. This was a liberal test of the UH, because any agreement statistically above zero would be considered intact perception within a cultural context. Comparisons against what would be expected by chance (a more stringent test of universality) are presented in the SOM.

Results

Emotion Perception from Western Vocalizations is Culturally Variable—Our results indicate that individuals from a remote culture do not “recognize” the intended emotions in Western vocal utterances, contrary to UH predictions (Figure 1, and Table S3). An ANOVA on the mean percent agreement, with cultural group as the between subjects factor (Himba vs. US) and the emotion category as the within subjects factor (amusement, anger, disgust, fear, relief, sadness, sensory pleasure, surprise, triumph), revealed that, in contrast to US participants, the Himba participants rarely produced the expected emotion label for the vocal utterances, $F(1, 46) = 146.351, p < .001, \eta_p^2 = .761$. This main effect for cultural group was qualified by a significant emotion $\times$ cultural group interaction $F(8, 368) = 12.113, p < .001, \eta_p^2 = .208$ (See Figure 1). Both US and Himba participants had the highest agreement for labeling laughter as amusement, 79% and 69% respectively. Both groups were also likely to label screams as fear, although for the Himba group, their labeling of screams as fear was not different from what would be expected by chance. Furthermore, the Himba participants used “fear” to label many different vocalizations, indicating that higher than zero agreement was due to a high base rate of term use more generally (SOM Tables S4 and S5 for confusion matrices). For all other categories of emotion, the Himba participants labeled the vocalizations in agreement with the presumed universal pattern less than 5% of the time (which did not differ significantly from zero), indicating that most of the vocalizations were not perceived similarly across cultures.

Himba participants appeared to have a cultural tendency to describe vocalizations in behavioral terms initially, performing action identification rather than mental state inference on most trials (Kozak, Marsh, & Wegner, 2006; Vallacher & Wegner, 1987). For example, instead of describing a vocalization as fearful, participants would use a term that translates to “scream”. On average, Himba participants provided non-mental state content on 69% of the trials compared to US participants who provided such content only 12% of the time (SOM for additional analyses).³

While US participants tended to produce labels that agreed with the intended emotions portrayed, they did so at lower levels of agreement than reported in experiments where emotion perception was assessed by choosing a vocalization to match an emotion word provided by the experimenter (Sauter et al., 2010) or when asked to match a vocalization to an emotion word from a small set of words provided by the experimenter (Hawk et al., 2009; Simon-Thomas et al., 2009).

Affect Perception from Western Vocalizations is Consistent Across Cultures

Valence: Our results support the hypothesis that valence perception (distinguishing pleasant, neutral, and unpleasant states) in vocal utterances is relatively cross-culturally stable (See Figure 2; Table S6). An ANOVA on mean percent valence-based agreement, with cultural group as the between subjects factor and emotion category as the within subjects factor, revealed that both US (Mean=75%, SD=28.35) and Himba (Mean=50.46% SD=26.26%) participants labeled vocal utterances with a valence-appropriate term at levels greater than zero (with the exception of Himba labels for portrayals of surprise). In contrast to US participants, however, the Himba participants offered fewer valence-consistent labels for the vocal utterances, $F(1, 46)=40.20, p<.001, \eta_p^2=.466$. The effect of cultural group was qualified by an interaction between emotion category $\times$ cultural group, $F(8, 368)=13.273, p<.001, \eta_p^2=.224$. When compared to US participants, Himba participants were less likely to freely label the putative vocalizations for disgust, fear, and, sadness with negative emotion or affect words, and they were less likely to freely label vocalizations of sensory pleasure with positive emotion or affect words. Again, this may reflect the tendency of Himba participants to engage in action identification rather than mental state inference (see SOM). US and Himba participants were equivalently likely to perceive positivity in vocalizations for triumph and negativity in vocalizations for anger. Himba participants were more likely than US participants to label vocalizations of amusement as positive ($p<.001$, 2-tailed).

Arousal: Our results provide some limited support for the cross-cultural stability of arousal perception (distinguishing activated, neutral, and deactivated states) in vocal utterances (see Figure 3; Table S7). Perception of arousal was less robust cross-culturally than valence, particularly because Himba participants appeared to have difficulty correctly labeling lower arousal states in the putative vocalizations for relief, sensory pleasure, and surprise. An ANOVA on mean percentage agreement, with cultural group as the between subjects factor

³This finding alone does not explain the cultural differences in discrete emotion perception, however, since Himba participants produced a comparable amount of mental state content (on 71% of trials) to US participants (83% of trials), but often following probes for that type of content.

and the portrayed emotion category as the within subjects factor, revealed that, in contrast to US participants, the Himba participants produced less arousal-consistent labels for vocal utterances overall, $F(1, 46)=60.259, p<.001, \eta_p^2=.567$. Whereas US participants perceived arousal with comparable agreement levels to valence (Mean =72.69%, $SD=32.46$), the Himba level of agreement was considerably lower (Mean=37.03%, $SD=32.95$). The effect of cultural group was qualified by an interaction between emotion category $\times$ cultural group, $F(8, 368)=6.15, p<.001, \eta_p^2=.118$, indicating that Himba participants did not as frequently freely label the vocalizations with words that were in agreement with the level of arousal portrayed in the utterance (all p 's<.005, 2-tailed), with the exception of utterances for amusement and surprise. Again, this limited evidence for universality may be because Himba perceivers often labeled the vocalizations using something other than mental state terms.

Study 2

In Study 2, we again tested whether Himba individuals could perceive affective properties of valence and arousal, and discrete emotions. Following Sauter et al. (2010), a second sample of Himba individuals listened to a series of situations (e.g., “Someone is suddenly faced with a dangerous animal and feels very scared”) and for each story selected which of two vocalizations corresponded to the emotional context. We also examined whether the foils of the task provided a context for improving performance. Specifically, participants received a foil vocalization that matched the target in valence on some trials (e.g., an anger story with a growl target and a scream foil); on others, foil and target did not match (e.g., an anger story with a growl and a laugh). A similar procedure was followed for arousal. The trials where foil and target matched on valence and arousal (e.g., anger and disgust) provided the clearest test of whether discrete emotions are perceived universally. Because the design was optimized to test for discrete emotion versus affect perception, we were unable to include enough trial types to allow for individual emotion analyses as in Study 1.

Study 2 was specifically designed to examine whether providing conceptual content within the emotion perception task itself would improve performance in the Himba cultural group, making them their responses more closely resemble US participants. Specifically, we drew on prior research indicating that emotion perception performance is improved in forced-choice (when compared to free-labeling) tasks (see Russell, 1994). Since US participants produced emotion labels that were largely consistent with the expected category in our free labeling experiment (Study 1), it was not necessary to use a forced choice task to test whether the vocalizations were culturally meaningful cues to emotion. Furthermore, providing Himba participants with emotion concept information as part of the perception task allowed us to rule out the possibility that cultural variation in Study 1 was due to *decoding rules* (i.e., reporting on percepts in a socially desirable way within a given culture, such as under reporting negative emotions to enhance social harmony). The influence of decoding rules is minimized when the emotion category is embedded in the task as the stimulus.

Materials and Methods

Participants—Participants were 37 native Herero speakers from the Himba ethnic group (13 male, 24 female; mean age= 27.14, SD=13.04) (see SOM for details).

Materials

Vocalizations: Vocalizations were the same audio files used in Study 1.

Emotion Words and Scenarios: The scenarios and words originally by Sauter et al (2010) were recorded by a translator in Otji-Herero. We used a different translator for Study 2, because our original translator passed away.

Procedure—On a given trial, participants listened to an audio recording of an emotion scenario (with an emotion word embedded) followed by two vocalizations. As the first vocalization played, an icon appeared on the left side of the screen; as the second played, the icon appeared on the right side of the screen. Both icons then appeared simultaneously and participants were instructed to press the icon (left or right) corresponding to the sound which best matched the scenario. Scenarios and/or vocalizations were repeated for participants who wished to hear them again. On each trial, there was always a “correct” vocalization that matched the story in discrete emotion content. The foil vocalization was then varied in relationship to the correct vocalization in one of four conditions, listed in Table 2, allowing us to distinguish whether Himba participants perceived valence, arousal, or discrete emotions. Participants completed a total of 18 trials with 4-5 trials of each type.

Results

Our results indicate that Himba participants perceived only valence in the vocalizations better than what would be expected by chance. An ANOVA on mean percentage accuracy, with foil type (valence-matched, arousal-matched, affect-matched, affect-mismatched) and target valence (positive or negative) of the target word/story as within subject factors. This analysis revealed a main effect of foil type, $F(3, 96) = 3.355, p < .05, \eta_p^2 = .095$. One-sample t-tests revealed that participants' performance was significantly above chance only in the arousal-matched condition ($M = 60.88, SD = 28.90, t(32) = 2.163, p < .05$, 2-tailed, where valence-based information can be used to distinguish between the target and foil.⁴ The ANOVA also revealed a main effect of target valence, $F(1, 32) = 8.85, p < .01, \eta_p^2 = .217$, such that participants were more accurate when the target was a negative ($M = 56.25, SD = 17.92$) compared to a positive ($M = 46.61, SD = 15.73$) vocalization, $t(32) = 2.975, p < .005$. This finding is consistent with higher accuracy for positive (compared to negative) vocalizations in Sauter (2010).

⁴Affect-mismatched trials did not yield above-chance accuracy, despite the fact that valence-based information could be used to discriminate between the target and foil. This unexpected finding may indicate that the mis-match of arousal-based information may have made the use of valence-based information more difficult. Reliance on different stimuli as foils to construct the four trial types may also explain the lower accuracy levels.

General Discussion

Taken together, these two experiments demonstrate important boundary conditions to claims that emotions can be universally recognized in vocal cues. In both the free labeling experiment (where emotion concept information was not provided for participants) and the forced choice experiment (where concept information was provided), Himba participants' did not perceive the intended emotional states in vocal utterances. These findings indicate that links between specific vocalizations (e.g., crying) and specific perceived mental states (e.g., sadness) are not always cross-culturally preserved. Study 2 results run contrary to even weak universalist accounts (e.g., dialect theory), where cultural variation is always expected, but better than chance agreement in emotion perception across cultures is predicted. Our findings are consistent with a growing number of studies showing that *emotion perception* is culturally relative (for a review see Barrett et al., 2011), and that performance is highly dependent on the conceptual context provided to participants (Russell, 1994; Nelson & Russell, 2013). Our results are also consistent with recent evidence from our lab demonstrating that Himba individuals do not perceive the intended emotion categories in Western facial portrayals (Gendron, Roberson, van der Vyver & Barrett, in press). Both studies reported here point to the conclusion that *valence perception*, rather than discrete emotion perception per se, is robust across cultures, with valence coming closer to a core human capacity (Russell, 1991).

Limitations and Future Directions—The present experiments are not without limitations. We tested only two samples from a single remote culture. Additional research is needed to explore relativity versus universality of emotion and affect perception in other cultural contexts and using other non-verbal cues. Additionally, our experiments used posed, highly caricatured vocal utterances (according to the “portrayal paradigm”; Scherer, Johnstone & Klasmeyer, 2003), which might fail to capture the range of vocal acoustics in spontaneous vocalizations. For example, in contrast to spontaneous vocalizations which are thought to be driven by a “production-first” system (Owren, Amoss, & Rendall, 2011) associated with stronger physiological changes, posed vocalizations might be produced by a learned and volitional “receptive-first” system, generating different acoustical patterns from affectively driven ones. This might explain why acoustical properties of prosody (i.e., fundamental frequency and amplitude) in spontaneous utterances typically correlate well with arousal (Bachorowski, 1999), but arousal-based perception was not robust across both of our experiments. Furthermore, posed and spontaneous utterances might be more similar for some emotion categories (where learning and experience are not necessary) than for others (where learning and prior experience are more important). This might explain unexpected differences in affect perception across vocalizations in Study 1 and overall lower accuracy for positive vocalizations in Study 2. Future research is needed to explore these possibilities. Nonetheless, the fact that we did not find evidence to support the UH cannot be explained by stimuli lacking in sufficient statistical regularity or “source clarity”, a critique leveled against much of the older (pre-1970s) studies reporting support for relativity in emotion perception (Naab & Russell, 2007; c.f. Scherer, 2003). Furthermore, use of posed stimuli rules out the alternative explanations that low recognition levels result from *display rules* (Ekman, 1972). Specifically, display rules to mask felt expressions, such that non-verbal cues contain fewer perceptual regularities, could account for correspondingly lower

recognition levels. Posed stimuli circumvent this problem since these cues are artificially constructed by target individuals and thus do not contain “masked” displays.

Another limitation is that the current experiments were not designed to *fully* characterize the extent of cultural relativity in emotion perception (we did not have vocalizations from Himba individuals). However, asking non-Western participants to evaluate Western vocalizations as we did is sufficient to examine whether the Western cultural model holds in other cultural contexts. Future research must examine whether other cultural models for emotion do not necessarily extend to Western cultures.

Finally, Study 1 revealed that Himba participants frequently understood vocalizations in situational (e.g., someone is sick) and action (e.g., growling) terms. Action Identification Theory (Kozak, Marsh, & Wegner, 2006) demonstrates that physical movements can be understood as an action or as evidence of a mental state. Emotion perception, at least in a Western cultural context, involves both action identification and mental state inference (Spunt & Lieberman, 2012), but our results indicate that Himba participants disproportionately understood the vocalizations in action terms. While speculative, this finding suggests that Himba conceptions of emotion may be more action-based and situated rather than rooted in mental feelings. Cross-cultural variability in the concept of emotion has been documented (e.g., Wierzbicka, 1999), and future research is required to explore this possibility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Kemuu Jakurama and Tjakazapi Mbunguha for translation services, Julia Reading, Hanna Negami and Sharon Feldman for coding assistance, Emiliana Simon-Thomas and colleagues for use of their vocal stimulus set, and Jules Davidoff and Serge Caparos for the use of field equipment. This research was supported by a National Institutes of Health Director's Pioneer Award (DP1OD003312) to Lisa Feldman Barrett.

References

- Bachorowski JA. Vocal expression and perception of emotion. *Current Directions in Psychological Science*. 1999; 8(2):53–57.
- Barrett LF. Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science*. 2011; 20:400–406.
- Barrett LF, Bliss-Moreau E. Affect as a Psychological Primitive. *Adv Exp Soc Psychol*. 2009; 41:167–218. [PubMed: 20552040]
- Barrett LF, Mesquita B, Gendron M. Context in emotion perception. *Current Directions in Psychological Science*. 2011; 20(5):286–290.
- Boucher JD, Carlson GE. Recognition of facial expressions in three cultures. *Journal of Cross-Cultural Psychology*. 1980; 11:263–280.
- Buck, R. *The communication of emotion*. New York: Guilford Press; 1984.
- Ekman, P. Universals And Cultural Differences In Facial Expressions Of Emotions.. In: Cole, J., editor. *Nebraska symposium on motivation*. University of Nebraska Press; Lincoln, Nebraska: 1972. 1971. p. 207-283.

- Ekman P, Friesen W. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*. 1971; 17(2):124–129. [PubMed: 5542557]
- Ekman P, Heider E, Friesen WV, Heider K. Facial expression in a preliterate culture. 1972
- Ekman P.; Rosenberg, EL. Oxford University Press; New York: 1997. What the face reveals : basic and applied studies of spontaneous expression using the facial action coding system (FACS)..
- Marsh AA, Elfenbein HA, Ambady N. Nonverbal “Accents” Cultural Differences in Facial Expressions of Emotion. *Psychological Science*. 2003; 14(4):373–376. [PubMed: 12807413]
- Matsumoto D. Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*. 1989; 20(1):92–105.
- Elfenbein HA. Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review*. 2013; 5(1):90–96.
- Elfenbein HA, Ambady N. On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol Bull*. 2002; 128(2):203–235. [PubMed: 11931516]
- Gendron M, Lindquist KA, Barsalou L, Barrett LF. Emotion words shape emotion percepts. *Emotion*. 2012; 12(2):314–325. [PubMed: 22309717]
- Gendron M, Roberson D, van der Vyver MJ, Barrett LF. Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*. in press.
- Hawk ST, van Kleef GA, Fischer AH, van der Schalk J. “Worth a Thousand Words”: Absolute and Relative Decoding of Nonlinguistic Affect Vocalizations. *Emotion*. 2009; 9(3):293–305. [PubMed: 19485607]
- Hoehl S, Striano T. Discrete Emotions in Infancy: Perception without Production? *Emotion Review*. 2010; 2(2):132–133.
- Hontz CR, Hartwig M, Kleinman SM, Meissner CA. Credibility Assessment at Portals, Portals Committee Report. 2009
- Izard CE. Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychol Bull*. 1994; 115(2):288–299. [PubMed: 8165273]
- Jack RE, Garrod OGB, Yu H, Caldara R, Schyns PG. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(19):7241–4. [PubMed: 22509011]
- Kozak MN, Marsh AA, Wegner DM. What do I think you're doing? Action identification and mind attribution. *J Pers Soc Psychol*. 2006; 90(4):543–555. [PubMed: 16649854]
- Lindquist KA, Barrett LF, Bliss-Moreau E, Russell JA. Language and the perception of emotion. *Emotion*. 2006; 6(1):125–138. [PubMed: 16637756]
- Nelson NL, Russell JA. Universality revisited. *Emotion Review*. 2013; 5(1):8–15.
- Naab PJ, Russell JA. Judgments of emotion from spontaneous facial expressions of new guineans. *Emotion*. 2007; 7(4):736–744. [PubMed: 18039042]
- Norenzayan A, Heine SJ. Psychological universals: what are they and how can we know? *Psychol Bull*. 2005; 131(5):763–784. [PubMed: 16187859]
- Owren MJ, Amoss RT, Rendall D. Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*. 2011; 73(6):530–544. [PubMed: 21509789]
- Russell JA. The Preschoolers Understanding of the Causes and Consequences of Emotion. *Child Dev*. 1990; 61(6):1872–1881.
- Russell JA. Culture and the categorization of emotions. *Psychol Bull*. 1991; 110(3):426–450. [PubMed: 1758918]
- Russell JA. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol Bull*. 1994; 115(1):102–141. [PubMed: 8202574]
- Russell JA, Bachorowski JA, Fernandez-Dols JM. Facial and vocal expressions of emotion. *Annual Review of Psychol*. 2003; 54:329–349.
- Russell JA, Barrett LF. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality and Social Psychology*. 1999; 76(5):805–819. [PubMed: 10353204]

- Sauter DA, Eisner F, Ekman P, Scott SK. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(6):2408–2412. [PubMed: 20133790]
- Scherer, KR. Affect bursts.. In: van Goozen, SHM.; Van de Poll, NE.; Sergeant, JA., editors. *Emotions: Essays on emotion theory*. Erlbaum; Hillsdale, NJ: 1994.
- Scherer KR. Expression of emotion in voice and music. *Journal of Voice*. 1995; 9(3):235–248. [PubMed: 8541967]
- Scherer KR. Vocal communication of emotion: A review of research paradigms. *Speech Communication*. 2003; 40:227–256.
- Scherer, KR.; Johnstone, T.; Klasmeyer, G. Vocal expression of emotion.. In: Davidson, RJ.; Scherer, KR.; Hill Goldsmith, H., editors. *Handbook of Affective Sciences*. Oxford; New York: 2003. p. 433-456.
- Schimmack U. Cultural influences on the recognition of emotion by facial expressions: Individualistic or Caucasian cultures? *Journal of Cross-Cultural Psychology*. 1996; 27(1):37–50.
- Simon-Thomas ER, Keltner DJ, Sauter D, Sinicropi-Yao L, Abramson A. The Voice Conveys Specific Emotions: Evidence From Vocal Burst Displays. *Emotion*. 2009; 9(6):838–846. [PubMed: 20001126]
- Sorensen, ER. Culture and the expression of emotion.. In: Williams, TR., editor. *Psychological Anthropology*. Mouton Publishers; Paris: 1975. p. 361-372.
- Spunt RP, Lieberman MD. An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *Neuroimage*. 2012; 59(3):3050–3059. [PubMed: 22019857]
- Weinberger S. Airport Security: Intent to deceive? *Nature*. 2010; 465:412–415. [PubMed: 20505706]
- Widen SC, Christy AM, Hewett K, Russell JA. Do proposed facial expressions of contempt, shame, embarrassment, and compassion communicate the predicted emotion? *Cogn Emot*. 2011; 25(5): 898–906. [PubMed: 21824027]
- Wierzbicka, A. *Emotions across languages and cultures: diversity and universals*. Cambridge University Press; Cambridge: 1999.

Figure 1. Cross-cultural comparisons for emotion

Performance for perception of discrete content from the voice. Mean percent ($\pm$ SEM) discrete emotion agreement is presented on the y-axis. Data are presented by culture and discrete emotion portrayal type (x-axis).

Figure 2. Cross-cultural comparisons for valence

Performance for perception of valence (positive, negative and neutral) content from the voice. Mean percent ($\pm$ SEM) valence agreement is presented on the y-axis. Data are presented by culture and discrete emotion portrayal type (x-axis).

Figure 3. Cross-cultural comparisons for arousal

Performance for perception of arousal/activation (high, mid, and low) content from the voice. Mean percent ($\pm$ SEM) of arousal agreement is presented on the y-axis. Data are presented by culture and discrete emotion portrayal type (x-axis).

Figure 4. Himba Forced-Choice Performance

Himba participants' performance in Study 2 based on the foil condition (x-axis). Mean percentage ($\pm$ SEM) accuracy to select the correct vocalization is presented on the y-axis.

Table 1

Examples of Task Conditions in Study 2.

Condition	Example Target Vocalization	Example Foil Vocalization
Affect-matched	Anger	Fear
Affect-mismatched	Anger	Relief
Valence-matched	Anger	Sadness
Arousal-matched	Anger	Triumph