

Psychological Science

<http://pss.sagepub.com/>

Cultural Relativity in Perceiving Emotion From Vocalizations

Maria Gendron, Debi Roberson, Jacoba Marieta van der Vyver and Lisa Feldman Barrett

Psychological Science published online 5 February 2014

DOI: 10.1177/0956797613517239

The online version of this article can be found at:

<http://pss.sagepub.com/content/early/2014/02/05/0956797613517239>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Association for Psychological Science](http://www.sagepublications.com)

Additional services and information for *Psychological Science* can be found at:

Email Alerts: <http://pss.sagepub.com/cgi/alerts>

Subscriptions: <http://pss.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [OnlineFirst Version of Record](#) - Feb 5, 2014

[What is This?](#)

Cultural Relativity in Perceiving Emotion From Vocalizations

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

¹Affective Science Institute, Northeastern University; ²Department of Psychology, Northeastern University;

³Department of Psychology, University of Essex; ⁴School of Nursing and Public Health, University of

Namibia; ⁵Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown,

Massachusetts; and ⁶Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School

Psychological Science

1–10

© The Author(s) 2014

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0956797613517239

pss.sagepub.com



Abstract

A central question in the study of human behavior is whether certain emotions, such as anger, fear, and sadness, are recognized in nonverbal cues across cultures. We predicted and found that in a concept-free experimental task, participants from an isolated cultural context (the Himba ethnic group from northwestern Namibia) did not freely label Western vocalizations with expected emotion terms. Responses indicate that 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 the target and foil on a given trial matched in both valence and arousal, neither valence nor arousal, valence only, or arousal only. Himba participants achieved above-chance accuracy only when foils differed from targets in valence only. Our results indicate that the voice can reliably convey affective meaning across cultures, but that perceptions of emotion from the voice are culturally variable.

Keywords

emotions, affect, cross-cultural differences, vocalizations

Received 1/31/13; Revision accepted 11/19/13

Think about the last time you heard someone sigh, chuckle, or groan and concluded that the person was tired, amused, or frustrated. The *universality hypothesis* states that (barring illness) all humans innately express and recognize the same emotions in nonverbal behaviors, including vocalizations. Universalist views agree that each emotion has a “fixed set of neural and bodily expressed components” (Tracy & Randles, 2011, p. 398). According to strong versions of this hypothesis, vocal cues contain perceptual regularities sufficient to broadcast discrete emotion information to perceivers (Sauter, Eisner, Ekman, & Scott, 2010; Scherer, 1994). As a consequence, it is hypothesized that emotions can be “recognized” independently of language or conceptual knowledge (Hoehl & Striano, 2010; Izard, 1994). In fact, compared with facial expressions, vocalizations are thought to allow for better “detectability” because they “can travel omnidirectionally and over long distances” (Hawk, van Kleef, Fischer, & van der Schalk, 2009, p. 294). Even according to a strong universality hypothesis, 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; Matsumoto, 1989; Schimmack, 1996). Weaker versions of the universality hypothesis posit cultural dialects for universal expressions (e.g., Elfenbein, 2013; Marsh, Elfenbein, & Ambady, 2003). According to all versions of the hypothesis, 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 have provided a stringent test of the universality hypothesis

Corresponding Author:

Lisa Feldman Barrett, Department of Psychology, Northeastern University, 253 Nightingale Hall, Boston, MA 02115
E-mail: l.barrett@neu.edu

(see Table S1 in the Supplemental Material available online) by using a *two-culture approach*, in which 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 universality hypothesis 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 nonverbal vocal utterances (laughs, screams, sighs, etc.) in line with their intended “universal” emotional meaning (i.e., the Western model of amusement, fear, relief, etc.). 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), which led Sauter et al. to claim support for the universality hypothesis.

Despite the ubiquity of universality claims in popular and scientific circles, empirical evidence questioning the reliability of universal emotion perception steadily accumulates (for reviews, see Barrett, 2011, and 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 universality hypothesis include emotion-concept cues within the task; tasks that do not prime emotion-concept knowledge (by requiring participants to freely label expressions rather than choose a label from provided response options) or reduce accessibility of emotion concepts (e.g., by using semantic satiation) impair emotion perception, even in U.S. participants (Boucher & Carlson, 1980; Ekman & Rosenberg, 1997; Gendron, Lindquist, Barsalou, & Barrett, 2012; Lindquist, Barrett, Bliss-Moreau, & Russell, 2006; Widen, Christy, Hewett, & Russell, 2011; see also Table S1).

In the experiments reported here, we sought to explicitly examine the role that conceptual context plays in shaping perceptions of vocalizations across cultures. We traveled to a 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 Herero language, which contains words that can be translated to English words for emotion (see Sauter et al., 2010).

Study 1: Free-Labeling Experiment

In Study 1, Himba and U.S. participants completed a free-labeling task of Western (U.S.) vocal portrayals of

emotion (see Table 1). Participant-provided labels were coded as in “agreement with the presumed universal pattern” if they fit the expected emotion term that was used to elicit the vocalization presented (e.g., “angry” or “anger” for a growl) or if they were a close synonym (e.g., “frustrated”). All vocalizations were produced by speakers of English. We predicted that Himba perceivers would have much lower agreement with the expected Western emotion (and presumed universal) pattern compared with U.S. perceivers. 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 prototypically 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 these affective dimensions given the ample evidence that across cultures, facial and vocal cues are perceived in terms of the valence and the level of arousal that they communicate (Russell, 1991; Russell, Bachorowski, & Fernandez-Dols, 2003; Russell & Barrett, 1999).

Method

Participants. The Himba participants were 24 native Herero speakers from the remote and mountainous northwest region of Namibia (12 male, 12 female; mean age = 35.96, $SD = 14.5$).² The U.S. participants were 24 individuals tested at the Boston Museum of Science in Boston, Massachusetts (13 male, 11 female; mean age = 38.41, $SD = 18.71$; for details on the two groups, see the Supplemental Material).

Stimuli. Stimuli were 36 nonword vocalizations. Two male and two female native English speakers each produced a vocalization to depict each of the following emotions: amusement, anger, disgust, fear, relief, sadness, sensory pleasure, surprise, and triumph (Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009). These vocalizations were similar to those used in Sauter et al. (2010), except that we substituted triumph vocalizations for achievement vocalizations. Each participant heard a

Table 1. Descriptions of the Vocalizations in Study 1

Discrete emotion portrayed	Vocalization
Amusement	Giggle, laughter
Anger	Guttural yell, growl
Disgust	“Ewww”
Fear	Scream
Relief	Sigh
Sadness	Cry
Sensory pleasure	“Mmm mmm”
Surprise	“Ahhh-ahhh”
Triumph	“Woohoo”

subset of 18 stimuli (a male and a female exemplar for each emotion), with the particular subset of posed vocalizations used 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 each vocalization with a word or phrase (experimenters were naive to the particular stimulus presented on a given trial). After each trial, the translator's immediate translation of the participant's verbal response (for Himba participants) or the participant's original response (for U.S. participants) was entered into a laptop computer by the experimenter. A participant who initially provided a description of a situation, behavior, or bodily state was prompted: "Can you think of a single word to describe the feeling, the emotion?" A participant who provided a vague affective response (e.g., "good" or "bad" feeling) was 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. (For additional procedural details, see the Supplemental Material.)

Data coding. The data were independently coded by two trained individuals. Trials were coded in a randomized order such that the coders were blind to the identity and culture of the responders. Using Russell's (1990) approach, coders rated whether the response on a given trial agreed or disagreed with the discrete emotion, valence, and arousal of the stimulus. In addition, both coders indicated responses for which "no mental content" was available (this coding was done on the full response for a given trial, i.e., for the content both before and after prompting). Reliability between the two coders (Cohen's kappa) was high for each of the subcodes—discrete emotion: $\kappa = .957$; valence: $\kappa = .943$, arousal: $\kappa = .958$. Discrepancies in coding were resolved by review and discussion among the coders and the first author. Data were analyzed by comparing the mean percentage of agreement between response and stimulus (for discrete emotion, valence, and arousal) against zero. This was a liberal test of the universality hypothesis, 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 Supplemental Material.

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 the prediction of the universality hypothesis (Fig. 1; see also Table S2 in the Supplemental Material). An analysis of variance (ANOVA) on the mean percentage of agreement, with cultural group as a between-subjects factor (Himba, U.S.) and portrayed emotion category as a within-subjects factor (amusement, anger, disgust, fear, relief, sadness, sensory pleasure, surprise, triumph), revealed a main effect of cultural group, $F(1, 46) = 146.351$, $p < .001$, $\eta_p^2 = .761$; in contrast to the U.S. participants, the Himba participants rarely produced the expected emotion label for the vocal utterances.

This main effect was qualified by a significant Emotion Category \times Cultural Group interaction, $F(8, 368) = 12.113$, $p < .001$, $\eta_p^2 = .208$ (see Fig. 1). Both U.S. and Himba participants showed the highest agreement with the intended emotion for the laughter stimuli, which they most frequently labeled as indicating amusement or a close synonym (e.g., happiness; 69% and 79%, respectively). Both groups also labeled screams as fear at a level significantly greater than zero, 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, which indicates that the higher-than-zero agreement was due to a high base rate of using this term more generally (see Tables S3 and S4 in the Supplemental Material for confusion matrices for the two groups). For all other categories of emotion, the Himba participants' labels for the vocalizations agreed with the presumed universal pattern less than 5% of the time (and these percentages did not differ significantly from zero). U.S. participants, in contrast, labeled all categories of vocalization in line with the presumed universal emotions at levels significantly above chance (see the Supplemental Material for these analyses and additional analyses examining whether responses referring to discrete emotions were more accurate than would be expected if responses were based more generally on perceived valence and arousal). Thus, most of the vocalizations were not perceived similarly across the two cultures.

Himba participants appeared to have a cultural tendency to describe vocalizations in behavioral terms initially; that is, on most trials, they first identified the action instead of making a mental-state inference (Kozak, Marsh, & Wegner, 2006; Vallacher & Wegner, 1987). For example, instead of describing a vocalization as fearful, they often used a term that translates to "scream." On average, Himba participants provided non-mental-state content on

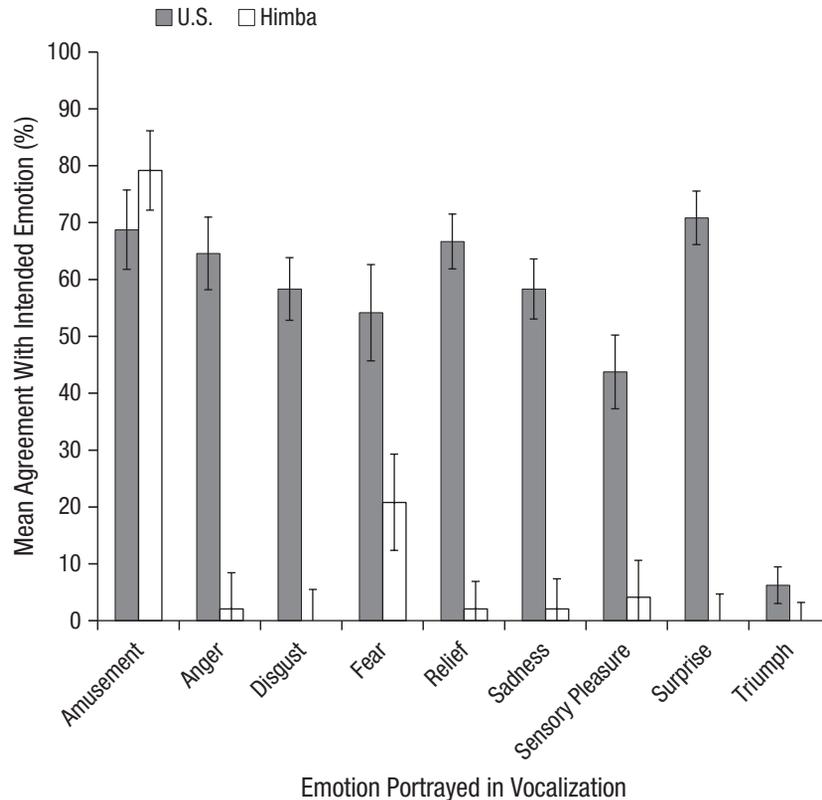


Fig. 1. Results from Study 1: mean percentage of responses that agreed with the intended emotion of the vocalization as a function of intended emotion. Results are presented separately for U.S. and Himba participants. Error bars indicate ± 1 SEM.

69% of the trials, whereas U.S. participants provided such content on only 12% of the trials (see the Supplemental Material for additional analyses relevant to this point).³

Although the U.S. participants tended to produce labels that agreed with the intended emotions, they did so at lower levels than reported for previous experiments in which emotion perception was assessed by having participants match a vocalization to an emotion scenario (Sauter et al., 2010) or match an emotion word, from a small provided set, to a vocalization (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 stable cross-culturally (Fig. 2; see also Table S5 in the Supplemental Material). Both U.S. ($M = 75.00\%$, $SD = 28.35$) and Himba ($M = 50.46\%$, $SD = 26.26\%$) participants labeled vocal utterances with a valence-appropriate term at levels greater than zero (with the exception of Himba labels for portrayals of surprise). An ANOVA on mean percent-

age of valence-based agreement, with cultural group as a between-subjects factor and portrayed emotion category as a within-subjects factor, revealed a main effect of cultural group, $F(1, 46) = 40.20$, $p < .001$, $\eta_p^2 = .466$; Himba participants offered fewer valence-consistent labels for the vocal utterances, compared with U.S. participants.

The effect of cultural group was qualified by an interaction between emotion category and cultural group, $F(8, 368) = 13.273$, $p < .001$, $\eta_p^2 = .224$. Compared with U.S. participants, Himba participants were less likely to freely label the vocalizations of disgust, fear, and sadness with negative emotion or affect words; to label the vocalizations of relief and sensory pleasure with positive emotion or affect words; and to label the vocalizations of surprise with neutral affect words (all $ps < .01$, two-tailed). These results may reflect the tendency of Himba participants to engage in action identification rather than mental-state inference (see the Supplemental Material). The U.S. and Himba participants were equivalently likely to perceive positivity in vocalizations for triumph and negativity in vocalizations for anger. The Himba participants were more likely than the U.S. participants to label vocalizations of amusement as positive ($p < .001$, two-tailed).

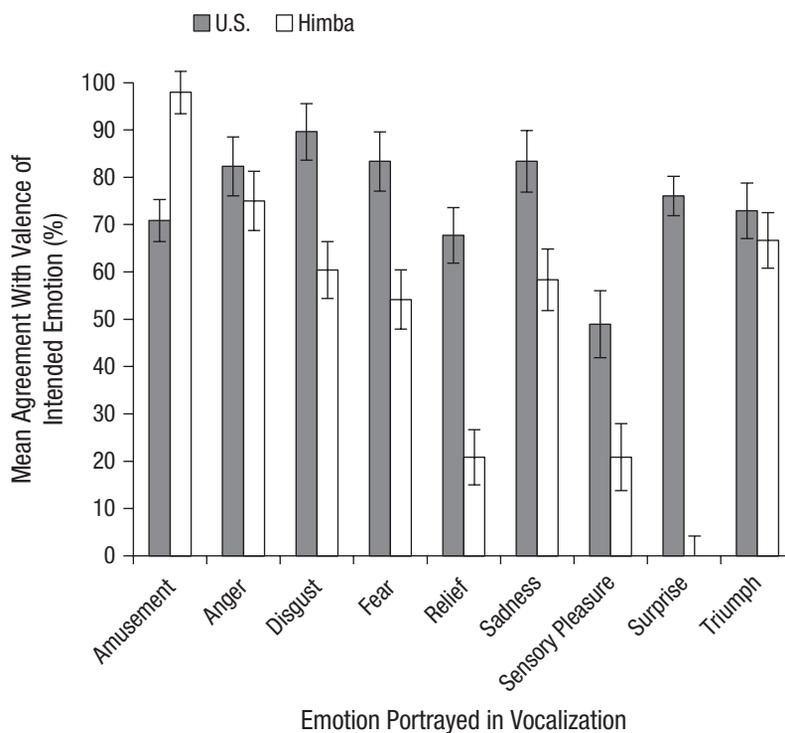


Fig. 2. Results from Study 1: mean percentage of responses that agreed with the valence of the vocalization (positive, negative, or neutral) as a function of the intended emotion. Results are presented separately for U.S. and Himba participants. Error bars indicate ± 1 SEM.

Arousal. Our results provide some limited support for the cross-cultural stability of arousal perception (distinguishing activated, neutral, and deactivated states) in vocal utterances (Fig. 3; see also Table S6 in the Supplemental Material). Perception of arousal was less robust cross-culturally than perception of valence, particularly because the Himba participants appeared to have difficulty correctly labeling low-arousal states in the vocalizations of relief, sensory pleasure, and sadness. An ANOVA on mean percentage of agreement, with cultural group as a between-subjects factor and portrayed emotion category as a within-subjects factor, revealed a main effect of cultural group, $F(1, 46) = 60.259, p < .001, \eta_p^2 = .567$; compared with the U.S. participants, the Himba participants produced fewer arousal-consistent labels for vocal utterances overall. Whereas the U.S. participants perceived arousal with agreement levels ($M = 72.69\%$, $SD = 32.46$) comparable to those for valence ($M = 75.00\%$, $SD = 28.35$), there was asymmetry in levels of agreement for valence ($M = 50.46\%$, $SD = 26.26\%$) and arousal ($M = 37.03\%$, $SD = 32.95$) among the Himba participants.

The effect of cultural group was qualified by an interaction between emotion category and cultural group, $F(8, 368) = 6.15, p < .001, \eta_p^2 = .118$; although the Himba

participants' arousal-based agreement was lower than that for the U.S. participants in the case of most of the intended emotions ($ps < .005$, two-tailed), this was not true for the vocalizations of amusement and surprise. Again, this limited evidence for universality may be due to Himba perceivers often labeling the vocalizations using something other than mental-state terms (despite explicit prompting for mental-state content).

Study 2: Forced-Choice Experiment

In Study 2, we again tested whether Himba individuals could perceive affective properties of valence and arousal, as well as discrete emotions, in vocalizations. Following Sauter et al. (2010), we recruited a second sample of Himba individuals to listen to a series of situations (e.g., "Someone is suddenly faced with a dangerous animal and feels very scared") and to select which of two vocalizations corresponded to the emotional context of each story. We also examined whether the particular foils used provided a context for improving performance. Specifically, on some trials, participants heard a foil vocalization that matched the target in valence (e.g., an anger story with a growl target and a scream foil); on other

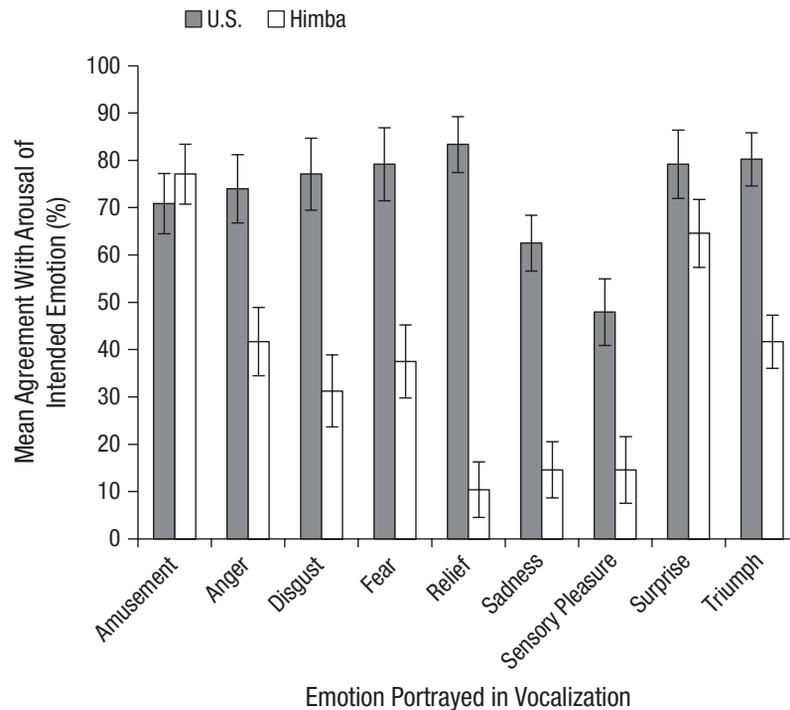


Fig. 3. Results from Study 1: mean percentage of responses that agreed with the arousal (activation) of the vocalization (high, mid, or low) as a function of the intended emotion. Results are presented separately for U.S. and Himba participants. Error bars indicate ± 1 SEM.

trials, the foil and target did not match in valence (e.g., an anger story with a growl and a laugh). A similar procedure was followed for arousal. The trials on which the foil and target matched in valence and arousal (e.g., “growl” and “eww” vocalizations portraying anger and disgust, respectively) provided the clearest test of whether discrete emotions are perceived universally. Because the design was optimized to separately examine perceptions of discrete emotion and affect perception by manipulating valence and arousal, we were unable to include enough trial types to allow analyses of individual emotions as in Study 1.

Study 2 was specifically designed to examine whether providing conceptual content within the emotion perception task itself would improve performance of the Himba participants, making their responses more closely resemble those of U.S. participants. Specifically, we drew on prior research indicating that emotion perception performance is improved in forced-choice (compared with free-labeling) tasks (see Russell, 1994). Because U.S. 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 for U.S. participants. Furthermore, providing Himba participants with

emotional concept information as part of the task allowed us to rule out the possibility that cultural variation in Study 1 was due to *decoding rules* (i.e., a culture’s rules for reporting on percepts in socially desirable ways, such as a rule to underreport negative emotions in order to enhance social harmony). The influence of decoding rules is minimized when the emotion categories are embedded within the stimuli for the task.

Method

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

Stimuli. The vocalizations were the same audio files used in Study 1. The scenarios and emotion words originally used by Sauter et al. (2010) were recorded in Herero by our translator. 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 computer screen; as the second played, the same icon appeared on the right side of the screen. Both icons then appeared simultaneously, and participants were instructed to press the touch-screen icon (left or right) corresponding to the sound that best matched the scenario. Scenarios and 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. Across trials, the relation between the foil vocalization and the correct vocalization was varied to create four conditions: affect-matched (foil matched the target in both valence and arousal), affect-mismatched (foil matched the target in neither valence nor arousal), valence-matched (foil matched the target in valence but not arousal), and arousal-matched (foil matched the target in arousal but not valence; see Table 2). This manipulation allowed us to distinguish whether Himba participants perceived valence, arousal, or discrete emotions in the vocalizations. Participants completed 4 or 5 trials of each type, for a total of 18 trials.

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 condition (valence-matched, arousal-matched, affect-matched, affect-mismatched) and target valence (positive, negative) as within subject factors, revealed a main effect of foil condition, $F(3, 96) = 3.355$, $p < .05$, $\eta_p^2 = .095$ (see Fig. 4). One-sample t tests revealed that participants’ performance was significantly above chance only in the arousal-matched condition ($M = 60.88\%$, $SD = 28.90$), $t(32) = 2.163$, $p < .05$, two-tailed, in which valence-based information could 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 ($M = 56.25\%$, $SD = 17.92$) rather than a positive ($M = 46.61\%$, $SD = 15.73$) vocalization, $t(32) = 2.975$, $p < .005$. Sauter et al. (2010)

also found higher accuracy for positive (compared with negative) vocalizations.

General Discussion

Taken together, these two experiments demonstrate important boundary conditions to claims that emotions can be universally recognized in vocal cues. In both Study 1 (in which emotion-concept information was not provided to participants) and Study 2 (in which emotion-concept information was provided), Himba participants did not perceive the intended Western 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 preserved cross-culturally. The results of Study 2 run contrary to even weak universalist accounts (e.g., dialect theory), according to which cultural variation is always expected, but cross-cultural agreement in emotion perception should be better than chance. 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 (Nelson & Russell, 2013; Russell, 1994). 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, such that valence comes closer to being a core human capacity (Russell, 1991).

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 nonverbal 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, Owren, Amoss, and Rendall (2011) have proposed that spontaneous vocalizations are driven by a “production-first” system associated with physiological changes, whereas posed vocalizations are produced by a learned and volitional “receptive-first” system that generates acoustical patterns different from those of spontaneous vocalizations. This framework might explain why acoustical properties of prosody (e.g., fundamental frequency and amplitude) in spontaneous utterances typically correlate well with arousal (Bachorowski, 1999), but arousal-based perception was not robust in either of our experiments.

Table 2. Examples of the Emotions Portrayed in the Vocal Stimuli in the Four Conditions of 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

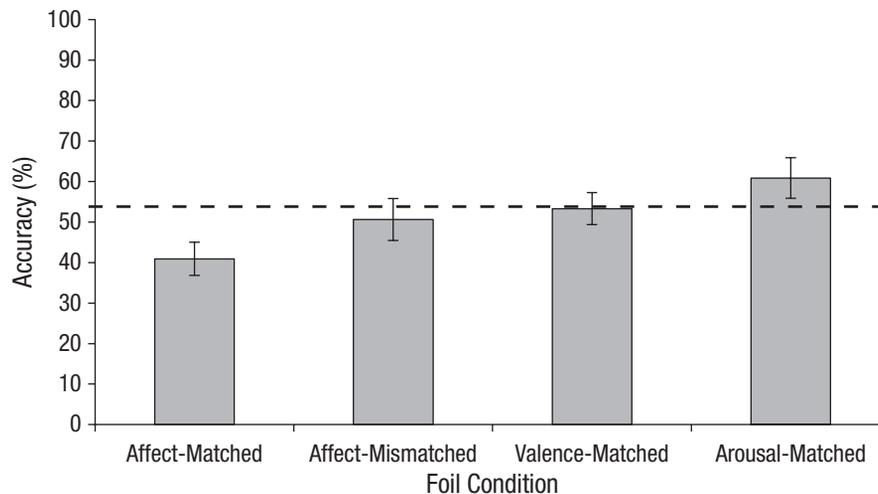


Fig. 4. Results from Study 2: Himba participants' mean percentage accuracy in selecting the correct vocalization as a function of foil condition. The dashed line shows the level of chance performance. Error bars indicate ± 1 SEM.

Furthermore, posed and spontaneous utterances might be more similar for some emotion categories (for which learning and experience are not necessary) than for others (for which learning and prior experience are more important). This might explain the unexpected differences in affect perception across vocalizations in Study 1 and the overall lower accuracy for positive vocalizations than for negative vocalizations in Study 2. Future research is needed to explore these possibilities.

Nonetheless, the fact that we did not find evidence to support the universality hypothesis cannot be attributed to use of stimuli lacking sufficient statistical regularity or "source clarity," a critique leveled against much of the older (pre-1970s) studies finding support for relativity in emotion perception (Naab & Russell, 2007; cf. Scherer, 2003). Furthermore, our use of posed stimuli rules out the alternative explanation that low recognition levels result from display rules (Ekman, 1972). Specifically, display rules to mask felt expressions, which reduce perceptual regularities in nonverbal cues, could lower recognition levels. Posed stimuli circumvent this problem because they are artificially constructed by target individuals and thus are not masked displays.

Another limitation of the current experiments is that they 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 action terms (e.g., growling). Research on action-identification theory (Kozak et al., 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. This finding suggests that Himba conceptions of emotion may be based more on action than on 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.

Author Contributions

M. Gendron and L. F. Barrett designed the research. M. Gendron, D. Roberson, and J. M. van der Vyver conducted the experiments. M. Gendron analyzed the data. M. Gendron, L. F. Barrett, and D. Roberson wrote the manuscript.

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 her colleagues for use of their vocal stimulus set; and Jules Davidoff and Serge Caparos for the use of field equipment.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by a National Institutes of Health Director's Pioneer Award (DP1OD003312) to L. F. Barrett.

Supplemental Material

Additional supporting information may be found at <http://pss.sagepub.com/content/by/supplemental-data>

Notes

1. Note that our goal in this study was to examine the veracity of the universality hypothesis using Western cues, not to fully explore cultural variation in emotion perception.
2. The age distribution we report for our Himba participants is not exact. Although there is a numerical system in Herero, most Himba individuals do not keep track of their age. Some Himba participants provided estimates of their own age; for the others, we estimated age on the basis of reproductive history, including number and age of children (if applicable), and physical appearance.
3. This finding alone does not fully explain the cultural differences in perception of discrete emotion, however. The frequency of mental-state content was comparable for Himba participants (on 71% of trials) and U.S. participants (83% of trials), although Himba participants often produced mental-state content only following prompts for that type of content.
4. Accuracy was not above chance on affect-mismatched trials despite the fact that valence-based information could be used to discriminate between the target and foil. This unexpected finding may indicate that the mismatch of arousal-based information made the use of valence-based information more difficult. The fact that the four trial types used different stimuli as foils may also explain the lower accuracy levels for affect-mismatched trials.

References

- Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, 8, 53–57.
- Barrett, L. F. (2011). Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science*, 20, 400–406.
- Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, 20, 286–290.
- Boucher, J. D., & Carlson, G. E. (1980). Recognition of facial expressions in three cultures. *Journal of Cross-Cultural Psychology*, 11, 263–280.
- Buck, R. (1984). *The communication of emotion*. New York, NY: Guilford 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: University of Nebraska Press.
- Ekman, P., & Rosenberg, E. L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. New York, NY: Oxford University Press.
- Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review*, 5, 90–96.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 203–235.
- Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, 12, 314–325.
- Gendron, M., Roberson, D., van der Vyver, M. J., & Barrett, L. F. (in press). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*.
- Hawk, S. T., van Kleef, G. A., Fischer, A. H., & van der Schalk, J. (2009). “Worth a thousand words”: Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, 9, 293–305.
- Hoehl, S., & Striano, T. (2010). Discrete emotions in infancy: Perception without production? *Emotion Review*, 2, 132–133.
- Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115, 288–299.
- Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences, USA*, 109, 7241–7244.
- Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology*, 90, 543–555.
- Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion*, 6, 125–138.
- Marsh, A. A., Elfenbein, H. A., & Ambady, N. (2003). Nonverbal “accents”: Cultural differences in facial expressions of emotion. *Psychological Science*, 14, 373–376.
- Matsumoto, D. (1989). Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*, 20, 92–105.
- Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion*, 7, 736–744.
- Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, 5, 8–15.
- Norenzayan, A., & Heine, S. J. (2005). Psychological universals: What are they and how can we know? *Psychological Bulletin*, 131, 763–784.
- Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, 73, 530–544.
- Russell, J. A. (1990). The preschooler's understanding of the causes and consequences of emotion. *Child Development*, 61, 1872–1881.
- Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin*, 110, 426–450.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141.

- Russell, J. A., Bachorowski, J. A., & Fernandez-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, *54*, 329–349.
- 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.
- Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, USA*, *107*, 2408–2412.
- Scherer, K. R. (1994). Affect bursts. In S. H. M. van Goozen, N. E. Van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 161–196). Hillsdale, NJ: Erlbaum.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*, 227–256.
- Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433–456). New York, NY: Oxford University Press.
- Schimmack, U. (1996). Cultural influences on the recognition of emotion by facial expressions: Individualistic or Caucasian cultures? *Journal of Cross-Cultural Psychology*, *27*, 37–50.
- Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion*, *9*, 838–846.
- Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage*, *59*, 3050–3059.
- Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*, 397–405.
- Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*, 3–15.
- Widen, S. C., Christy, A. M., Hewett, K., & Russell, J. A. (2011). Do proposed facial expressions of contempt, shame, embarrassment, and compassion communicate the predicted emotion? *Cognition & Emotion*, *25*, 898–906.
- Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge, England: Cambridge University Press.