Relativity in the perception of emotion across cultures

Author: Maria Therese Gendron

Boston College Electronic Thesis or Dissertation, 2013

Boston College

The Graduate School of Arts and Sciences

Department of Psychology

RELATIVITY IN THE PERCEPTION OF EMOTION ACROSS CULTURES

a dissertation

by

MARIA GENDRON

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

May 2013

Relativity in the Perception of Emotion Across Cultures

Maria Gendron

Supervised by Lisa Feldman Barrett

A central question in the study of human behavior is whether or not certain categories of emotion, such as anger, fear and sadness (termed "discrete emotions"), are universally recognized in the nonverbal behaviors of others (termed the "universality of attribution hypothesis"). In this dissertation, the universality of attribution hypothesis was revisited in order to examine whether individuals from remote cultural contexts perceive the same mental states in nonverbal cues as individuals from a Western cultural context. The studies described in this dissertation removed certain features of prior universality studies that served to obscure the underlying nature of cross-cultural perceptions. In study 1, perception of posed emotional vocalizations by individuals from a US cultural context were compared to those of individuals from the Himba ethnic group, who reside in remote regions of Namibia and have limited contact with individuals outside their community. In contrast to recent data claiming to support the universality hypothesis, we did not find evidence that emotions were universally perceived when participants were asked to freely label the emotion they perceived in vocalizations. In contrast, our findings did support the hypothesis that affective dimensions of valence and arousal are perceived across cultural contexts. In the second study, emotion perceptions based on facial expressions were compared between participants from US and Himba cultural contexts. Consistent with the results of Study 1, Himba individuals did not perceive the Western discrete emotion categories that their US counterparts did. Our data did support the hypothesis that Himba participants were routinely engaging in action perception, rather than mental state inference. Across both cultural contexts, when conceptual knowledge

about emotions was made more accessible by presenting emotion words as part of the task, perception was impacted. In US participants, perceptions conformed even more strongly with the previously assumed "universal" model. Himba participants appeared to rely more on mental state categories when exposed to concepts, but a substantial amount of cultural variation was still observed. Finally, in Study 3, perceptions of emotion were examined in a US cultural context after the focus of participants was manipulated, either onto mental states (broadly), emotions or behaviors. Perceptions of emotion did not differ substantially across these three conditions, indicating that within a US cultural context the tendency to infer mental states from facial expressions is somewhat inflexible. Overall, the findings of this dissertation indicate that emotion perception is both culturally and linguistically relative and that attempts to apply the Western cultural model for emotions as a universal one obscures important cultural variation.

TABLE OF CONTENTS

Chapter I. Introduction

Relativity and universality in the perception of emotion

An introduction to the emotion universality debate, a review of prior research on universality in emotion perception and the critiques of this literature, and a focused review on research suggesting that emotion perception is a relative process that is shaped by culturally constructed concepts.

"It is the mind which creates the world around us, and even though we stand side by side in the same meadow, my eyes will never see what is beheld by yours, my heart will never stir to the emotions with which yours is touched."
-George Gissing

One of the most compelling story lines in Psychology is that individuals from remote corners of the world (e.g., Ekman, 1972; Ekman & Friesen, 1971; Ekman, Sorenson, & Friesen, 1969; Izard, 1971) can recognize emotion in Western facial muscle movements ("facial expressions"). This claim of universality is part of the "basic emotion" view, where emotions are assumed to be evolutionarily given and biologically basic capacities that all humans share. For example, in this view, the emotion fear triggers a set of facial muscle movements, including the widening of the eyes and (often) a gaping of the mouth. This external "signal" on the face can then be recognized as "fear" by a perceiver. This recognition process should occur regardless of the culture of either the individual experiencing fear or the individual perceiving fear. Further, this recognition process has been argued to be pre-linguistic (Izard, 1994), such that even infants can recognize emotions in others. Based, in part, on the assumptions of universality, it was further postulated that innate circuitry in the brain is responsible for producing emotions, with each emotion circuit capable of producing a distinct set of facial behaviors, in addition to distinct physiological responses, phenomenology and other behaviors (Ekman, 1972). This "basic" emotion perspective has thrived in psychology and neuroscience and has even gained popularity in the popular media. Indeed the implications of a basic emotion perception perspective are profound—we can get a direct read out on the face (and perhaps in the body and brain) of the internal states of all other humans. In this sense, the face is a common form of communication that should allow us to overcome basic linguistic and cultural boundaries that divide us. Further, this opens up new vistas not only for communication, but also for extending the psychology of emotion into applied domains such as cross-cultural relations, security, law and so on.

The Universality of Emotion Perception

The literature on universal perception of emotion gained momentum in the late 1960s and early 1970s when several researchers, notably Paul Ekman and Carol Izard, tested whether individuals from other cultures (outside of the United States) perceived emotion from a set of facial poses that were constructed in an a priori fashion (i.e., people were directed to pose certain facial actions and the "best" portrayals were selected) to "signal" different emotional experiences. While this literature has grown considerably since that time, with hundreds of experiments testing cross-cultural perception of emotion (for a meta-analysis see Elfenbein & Ambady, 2002), research that tested individuals from the most remote cultures is relatively spare.[i]

A handful of experiments did test the universality of emotion perception in the strictest sense of the "two-culture" approach (Norenzayan & Heine, 2005), where individuals from remote cultural contexts were tested. The two-culture approach is key to arguments of universality, since this approach can rule out shared cultural practice, language or direct cultural contact as alternative explanations for similarities in psychological phenomena across cultures. In two distinct cultures, people without prior exposure to Western media representations of emotion were able to select the correct facial portrayals of *anger, fear, sadness, disgust*, etc., to match to a corresponding story about emotion (Ekman, 1972; Ekman & Friesen, 1971; Ekman, Heider, Friesen, & Heider, 1972). Further, individuals from the United States were able to select the correct emotion term to describe the emotion portrayed in videotapes of individuals from the Fore tribe in Papau New Guinea (Ekman, 1972, Study 3). (Although accuracy rates were relatively low —at 36% overall.) This line of experiments is often described as the definitive evidence that the face reveals internal state—or that facial behaviors are linked to categories of emotional experience in a one-to-one manner.

Extending Universality to Vocalizations?

The neurocultural model put forth by Ekman (1972) included vocal changes as part of the signature for each emotional state, but it was only recently that this hypothesis was empirically tested using the two-culture approach (Sauter, Eisner, Ekman, & Scott, 2010). Specifically, it was hypothesized that non-verbal vocal utterances like cries and screams are automatically generated by basic emotional states, such that these vocal utterances will be recognized in a remote cultural context. This recent experiment employed the situation story paradigm (as in the original two-culture work; Ekman & Friesen, 1971) in order to examine whether there is cross-cultural recognition of emotion from brief vocalizations called "vocal bursts". Sauter and colleagues tested individuals from the Himba, an ethnic group that resides in the North West region (historically called the Kaokoland) of the African nation Namibia where they live in small, semi-nomadic, pastoral communities that have limited contact with other groups. The Himba are a strong sample for the two-culture approach due to the relative cultural isolation of the communities. Evidence of psychological similarity in perceptions of discrete emotion from the voice between individuals in a Western culture and individuals from the Himba cultural context, then, should not be accounted for by cultural similarity or contact. This line of reasoning opens up possibility of an "innate" or universal psychological capacity being discovered (although it doesn't ensure that alternatives are ruled out, a point I will return to next). The results of this work are fairly striking, when taken at face value. Himba participants were able to select the vocalization (from a foil) that matched a word/story pairing at levels that exceeded chance. These data were used to support the conclusion that perception of vocalizations of emotion is a universal capacity.

Limitations to the Universality Literature

The universality literature, although wildly popular in the media and its applications in other disciplines, also remains contentious within the psychology of emotion. Chief among the critiques of this literature is that the methods provided conceptual framing that bolstered participants' performance and may open up these experiments to alternative explanations. Specifically, the two main experiments conducted by Paul Ekman and colleagues in the Dani and Fore of Papau New Guinea involved providing participants with a situational prompt and asking the participant to select the face which best matched the prompt. For example, in response to the prompt "he/she is looking at something which smells bad" participants would be presented with several options, one of which was a "disgusted" portrayal of emotion. Since participants tended to select the face with a wrinkled-nose, furrowed brow and pulled up lip (i.e., the Western pose selected by the researchers as the facial expression for disgust) more often than would be expected by chance, these data were interpreted as evidence that the emotion of "disgust" was perceived in the facial expression of emotion.

The conceptual frame presented in universality experiments attempted to accommodate the complexities of testing participants from a preliterate culture and difficulties associated with obtaining exact translations of emotion terms (which is itself an interesting area of research in the cross-cultural study of emotions). Yet the inclusion of a situational story may have also added an interesting source of context that might explain the high agreement rates in universality studies (c.f. Russell, 1994). Specifically, Russell argues that narrowing the responses of participants down to a few key emotions (i.e., the forced-choice paradigm) coupled with the inclusion of conceptual content (which in many of the studies in the larger cross-cultural literature were the response options themselves) serves to produce uniform responses that participants wouldn't otherwise endorse. This forced choice method appears in so much of the cross-cultural work on emotion perception that a meta-analysis of this literature was unable to examine method as a factor, for lack of studies that use alternative tasks (Elfenbein & Ambady, 2002).

The uniformity of methods employed poses a problem for drawing firm conclusions based on the extant cross-cultural literature, particularly given that studies directly comparing the forced-choice method to emotion recognition levels based on free labeling reveal dramatically different results (Boucher & Carlson, 1980; Izard, 1971; Rosenberg & Ekman, 1995; for a discussion see Russell, 1994; Widen, Christy, Hewett, & Russell, 2011). Specifically, providing participants with response options helps to increase "accuracy" levels. One older experiment reported a 16% increase in accuracy when words were provided in an emotion perception task versus when they were not (Kline & Johannsen, 1935), and more recently, this increase in perception accuracy was observed at 26% (Izard, 1971). Indeed, there are significantly higher cross-cultural accuracy findings within certain research groups that employed the forced-choice method (see Elfenbein & Ambady, 2002), suggesting further that the specific methods employed in the cross-cultural literature have a substantial impact on the conclusions that can be drawn from that data.

One explanation is that the forced choice method simply taps recognition memory whereas the free labeling method taps recall (c.f. Ekman & Rosenberg, 1995). Indeed, it is well established that there are higher accuracy rates in the former type of task. Yet perceivers will also agree on the incorrect label when it is provided (Russell, 1993). And there are similar findings in the early literature on emotion perception, where participants actually accepted labels that did not “match” the emotion a face was intended to portray as correct (Buzby, 1924; Langfied, 1918). Together these data suggest it is unlikely that language merely impacts on accuracy because it makes the “correct” response more accessible. Instead, the presence of concepts in the task seem to shift what the perceived appropriate emotion category is.

If language does more than simply make the “correct” category more accessible, what could it be doing? A possible explanation of these effects is that language is actually structuring perception (Barrett, Lindquist, & Gendron, 2007)—that even posed, static, highly stereotyped portrayals of emotion (such as the ones used in the universality work) are somewhat ambiguous as to their psychological meaning, and emotion words (as well as scenarios in the case of the universality literature) appear to help shape those percepts into more specific perceptions of “anger” or “fear”.

Existing Evidence that Language Structures Perception in a Western Cultural Context

There are a number of findings within a Western cultural context demonstrating that language has the capacity to shape emotion perception (reviewed in Barrett et al., 2007; Barrett, Mesquita, & Gendron, 2011; Lindquist & Gendron, 2013; Roberson, Damjanovic, & Kikutani, 2010). First, emotion words can produce biases in perceptual memory for a face, such that participants' memory for a facial portrayal is shifted toward the emotion that was paired (as a word) with the face at encoding (Halberstadt, 2005). Recent data suggest that this may be due to the grounding of emotion words in sensorimotor representations, given that the tendency to engage facial muscles consistent with an emotion word is related to the magnitude of later perceptual bias (Halberstadt, Winkielman, Niedenthal, & Dalle, 2009). Further, a completely false perceptual memory (i.e., remembering a smile) can be created based on a context that primes a specific semantic category (e.g., wining a sporting event), even when there were no category relevant facial actions actually present (Fernandez-Dols, Carrera, Barchard, & Gacitua, 2008). Words also appear to support categorical perception of facial expressions (Fugate, Gouzoules, & Barrett, 2010). Perceivers learned chimpanzee expressions (e.g., a *hoot*) either with an arbitrary label or without. Only those perceivers who learned the expressions with labels showed the hallmark of categorical perception—an advantage at discriminating morphs that crossed the categorical boundary between two expressions. These data suggest that it is not just the structural features of expressions that drive categorical perception, but the labels that are linked to those expressions also drive this effect. Importantly, these data can explain why many studies conducted in cultures with words for discrete emotions may also demonstrate categorical perception for portrayals of those emotions (e.g., Young et al., 1997).

Words also appear to provide an advantage to young children during emotion perception. Specifically, young children are more accurate to sort matching emotional faces into a box, and leave out mismatching faces, when that box is marked by a word as compared to when it is marked with a perceptually similar face with an identical posed set of facial actions (Russell & Widen, 2002). Further, emotion perception accuracy increases in parallel with the development of children's vocabulary for emotion words (Widen & Russell, 2008a). This happens in a category specific manner that is fairly consistent across children. For example, children acquire the term 'disgust' relatively late, at a mean age of 56 months (4.6 years) and only around this time do children distinguish between negative high arousal fear and disgust portrayals (for a review see Widen & Russell, 2008b).

Likewise emotion perception can be impaired by reducing the accessibility of emotion words, even when such words are incidental to the task at hand. When emotion words are made less accessible by a standard laboratory task called semantic satiation (for a review see Black, 2003), accuracy on a perceptual matching task drops to 36% from 42% (even though emotion words are not necessary to say whether two faces match in their emotional content or not) (Lindquist, Barrett, Bliss-Moreau, & Russell, 2006). Further, when participants are placed under verbal load, categorical perception for posed, caricatured faces is eliminated, such that typical perceptual advantages for distinguishing between stimuli from two different emotion categories are wiped out (Roberson, Damjanovic, & Pilling, 2007). Most recently, data suggests that these emotion perception effects resulting from the inaccessibility of emotion concepts may be due to shifts in the perceptual representation of faces (Gendron, Lindquist, Barsalou, & Barrett, 2012). Specifically, perceptual priming of a given facial expression is disrupted following semantic satiation of relevant emotion words (Gendron et al., 2012). These data indicate that words do more than shift judgments or memory for faces, but have an impact on the perceptual encoding of faces.

There are several brain-based studies that provide convergent evidence with the perceptual effects at the behavioral level, demonstrating that the distributed neural representation of an emotional face is shaped by conceptual processing involving emotion language. Providing perceivers with emotion (but not gender) words significantly reduced amygdala response to emotional faces (Lieberman et al., 2007). Amygdala response reductions presumably occurred because words helped to resolve competing "perceptual hypotheses" that arose from a structural analysis of the face alone (faces were presented in the absence of other contextual information). Furthermore, when perceivers judged a structurally neutral face as emotional, this engaged a network of regions (e.g., right superior temporal sulcus, bilateral orbitofrontal cortex, right anterior insula) (Thieschler & Pessoa, 2007) that are typically thought of as the distributed network for emotional face perception (Haxby, Hoffman, & Gobbini, 2000) and would not be predicted a priori for a structurally neutral face. Neural adaptation to emotional faces in this same network also appears to be driven by perceiver conceptualization (i.e., judgments of the emotion category), rather than by stimulus-defined properties (Fox, Moon, Iaria, & Barton, 2009). Specifically, even when repeated stimuli in an adaptation paradigm are changed so drastically that they are "perceptually" drawn from a different category, and theoretically should release the brain from adaptation (i.e., neural responses should go back up because the stimuli has changed), they fail to do so when perceivers judge that the category membership has not changed. Taken together data suggest that the neural representation of an emotional face is shaped by conceptual processing and is not determined by the stimulus features. Indeed, language may be routinely involved in emotion perception. A recent meta-analysis comparing brain activity during the perception and experience of emotion observed consistently greater activation in language-related regions, including inferior frontal gyrus (IFG), extending from the pars opercularis (Broca's area, BA 44) through pars triangularis (BA 45) and pars orbitalis on the inferior frontal convexity (BA 47/12 l) (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). These data suggest that language participates in emotion perception routinely because these activations hold across many distinct studies of emotion perception, with distinct task constraints and varying dependency on verbal labeling.

Future Directions of the Universality Literature

Limitations to the methods used in the universality literature were pointed out nearly 20 years ago (Russell, 1994), and mounting evidence (reviewed above) indicates that conceptual context importantly impacts emotion perception and, at a minimum, is artificially increasing inter-rater agreement in emotion perception studies. An important next step in this literature is to employ the two-culture approach, but to advance beyond the methods first introduced in the 1970s. Indeed, implicit in citations of the prior cross-cultural literature on emotion perception is a naïve assumption that if the features of the two-culture approach are met, this provides positive support for whatever "universal" capacity is under exploration. Of course, data from the two-culture literature (e.g., Sauter, et al. 2010) support the conclusion that individuals from remote cultural contexts are able to extract some meaning from the Western style vocalizations or faces they listened to. Otherwise, their selections on the task should have been random and not exceeded chance levels. But this does not suggest that it is necessarily a basic emotional state being perceived. This was the point that Russell (1995) made when he suggested that the literature supports minimal universality—clearly some meaning making can occur across cultures, just as it can for any type of action another person is engaged in. In order to determine what those sources of minimal universality may be, it is instructive to dissect the tasks used in the two-culture literature. In the tasks used, there appear to be at least three main sources of "information" that participants' performance may rest on. I will outline these three sources of information below and briefly indicate how the present dissertation attempts to examine these sources in more depth.

Concepts for emotions. The first form of information, concepts, are embedded in the majority of tasks in the two-culture literature on emotion perception. In the task used recently by Sauter et al. (2010), words appeared as part of the situation-story cue that participants were asked to match to vocalizations. On a given trial, participants were presented with a single emotion word (as well as a story) that matched one of the vocalizations. As described in the section on language and emotion perception above, narrowing the participant in on a specific concept may have a profound impact on how emotion perception proceeds. At a minimum, introducing concepts should lead to higher agreement rates than are typically observed when the word is not present. Perhaps more interestingly, concepts may be shaping how the brain uses basic sensory information to arrive at a "percept" for emotion—leading perceivers to perceive distinctions between vocal or facial cues as more extreme and categorical (although the language and categorical perception work has not directly been extended to the voice). Thus future experiments in the two-culture literature would benefit from stripping away this form of context in order to compare what accuracy rates and perceptions are like without words to bolster performance (for an example in a Western cultural context, see Widen et al., 2011). This was a primary aim of Study 1 in this dissertation. We examined whether in a free-labeling task (without the context of words in the task), individuals from the Himba ethnic group perceived the predicted discrete emotions in vocal bursts. In Study 2, we then manipulated the presence or absence of emotion concepts in the task and examined the subsequent impact on performance.

Affective dimensions of valence and arousal. The second potential source of information in prior two-culture studies is affective. This source of information is not typically directly highlighted in tasks, but instead derives from how the task is structured. For example, when response options are limited and the options differ in their basic affective properties (as was the case on half of the experimental trials in Sauter et al., 2010), participants can use affective, rather than discrete emotion information to solve the trial sucessfully. For example, if the task is to pick out which of two vocalizations match a disgust story, and the two choices are the correct "ewww" vocalization and a foil emotion conveying positive affect such as a giggle for amusement, then the perceiver could select the "correct" emotion by relying on perception of valence (whether the person feels good or bad). While the task feature of limited response options undoubtedly served to narrow the complexity of the task, and thus was deemed appropriate for testing individuals from remote cultures with limited exposure to testing contexts, it likely shaped the nature of the decision that participants were engaged in. This, in turn, complicates attempts to draw conclusions regarding discrete emotion perception from those data. While some of the two-culture literature did attempt to include within valence (e.g., two negative or two positive) choices, this does not rule out the possibility that affect perception was carrying performance on many of the trials—particularly because the degree of arousal was rarely controlled for across the two choices. Since the affective dimensions of valence (pleasure and displeasure) and arousal (activated to deactivated), to a lesser extent, are recovered in nearly all cross-cultural studies aimed at examining the structure of affect (Russell, 1991), a second aim of this dissertation was to examine whether affective dimensions of valence and arousal are perceived by individuals from a remote cultural context. In Study 1, we examined affect perceptions in vocal portrayals of emotion. There is evidence within a Western cultural context that dimensions of valence and arousal can be predicted based on the acoustical properties of the voice (although arousal to a greater extent) (Bachorowski & Owren, 2003). These findings suggest that that there may be an affective "signal" value in vocalizations that maps on to these more simple dimensions and may thus be cross-culturally similar.

Situation-tuned behavior. A final feature of all of the two culture experiments providing positive support for the universality hypothesis was the presence of situational stories as a cue in each of the tasks. This type of contextual information is important because it allows for "accurate" performance to ride on a very different type of inference than that of a discrete emotion. Specifically, presenting a situational context allows participants to identify situation-tuned actions rather than mental states and still perform "accurately". Take this example from Ekman and Friesen (1971): A participant is presented with a story "He/She is looking at something which smells bad", and is asked to select which of two faces, an anger portrayal or a disgust portrayal, match the story. An important feature of a disgust portrayal is that the nose appears to be crinkling up—which is argued to function as an oral-nasal rejection of "aversive chemosensory stimuli" (Chapman, Kim, Susskind, & Anderson, 2009). Thus it may be that the simple behavior of nose-crinkling, might be sufficient for participants to match the stimulus to the cue, without any mental inference about, or knowledge of, disgust. The question of whether perceivers from other cultures routinely engage in mental state inference (e.g., inferring a disgusted internal state) as opposed to action perception (e.g., someone smelling followed by a nose-crinkle) has not been directly examined in the prior literature and was the final major aim of the present dissertation. If it is the case that individuals from remote cultural contexts are routinely engaging in action perception (rather than mental state inference) this would suggest a very different interpretation of the prior universality studies is in order. In Study 2 we examined whether there is cultural variation in the extent to which individuals engage in mental state inference versus action perception. In Study 3, we then attempted to manipulate the tendency to engage in mental state inference within a Western cultural context. This final study was aimed at assessing the relative stability versus malleability of mental state inference depending on the manner in which participants are instructed.

References

Bachorowski, J. A., & Owren, M. J. (2003). Sounds of emotion: production and perception of affect-related vocal acoustics. *Ann N Y Acad Sci, 1000*, 244-265.

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends Cogn Sci, 11*(8), 327-332. doi: 10.1016/j.tics.2007.06.003

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in Emotion Perception. *Current Directions in Psychological Science, 20*(5), 286-290. doi: Doi 10.1177/0963721411422522

Black, S. R. (2003). A review of semantic satiation. In S. P. Shohov (Ed.), *Advances in Psychology Research, Volume 26* (pp. 95-106). Huntington, NY: Nova Science Publishers, Inc.

Boucher, J. D., & Carlson, G. E. (1980). Recognition of Facial Expression in 3 Cultures. *Journal of Cross-Cultural Psychology, 11*(3), 263-280. doi: Doi 10.1177/0022022180113003

Buzby, D. E. (1924). The interpretation of facial expression. *The American Journal of Psychology, 35,* 602-604.

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In Bad Taste: Evidence for the Oral Origins of Moral Disgust. *Science, 323*(5918), 1222-1226. doi: Doi 10.1126/Science.1165565

Ekman, P. (1972). Universals And Cultural Differences In Facial Expressions Of Emotions. In J. Cole (Ed.), *Nebraska symposium on motivation, 1971* (pp. 207-283). Lincoln, Nebraska: University of Nebraska Press.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J Pers Soc Psychol, 17*(2), 124-129.

Ekman, P., Heider, E., Friesen, W. V., & Heider, K. (1972). *Facial expression in a preliterate culture*.

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

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

Fernandez-Dols, J. M., Carrera, P., Barchard, K. A., & Gacitua, M. (2008). False recognition of facial expressions of emotion: Causes and implications. *Emotion, 8*(4), 530-539. doi: Doi 10.1037/A0012724

Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. S. (2009). The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *Neuroimage, 44*(2), 569-580. doi: Doi 10.1016/J.Neuroimage.2008.09.011

Fugate, J. M. B., Gouzoules, H., & Barrett, L. F. (2010). Reading Chimpanzee Faces: Evidence for the Role of Verbal Labels in Categorical Perception of Emotion. *Emotion, 10*(4), 544-554. doi: Doi 10.1037/A0019017

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion, 12*(2), 314-325. doi: 10.1037/a0026007

Halberstadt, J. (2005). Featural shift in explanation-biased memory for emotional faces. *J Pers Soc Psychol, 88*(1), 38-49. doi: Doi 10.1037/0022-3514.88.1.38

Halberstadt, J., Winkielman, P., Niedenthal, P. M., & Dalle, N. (2009). Emotional Conception: How Embodied Emotion Concepts Guide Perception and Facial Action. *Psychol Sci, 20*(10), 1254-1261.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn Sci, 4*(6), 223-233. doi: Doi 10.1016/S1364-6613(00)01482-0

Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Izard, C. E. (1994). Innate and Universal Facial Expressions - Evidence from Developmental and Cross-Cultural Research. *Psychol Bull, 115*(2), 288-299. doi: Doi 10.1037/0033-2909.115.2.288

Kline, L. W., & Johannsen, D. E. (1935). Comparative role of the face and of the face-body-hands as aids in identifying emotions. *The Journal of Abnormal and Social Psychology, 29*(4), 415-426.

Langfeld, H. S. (1918). Judgments of facial expression and suggestion. *Psychological Review, 25,* 488-494.

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words - Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychol Sci, 18*(5), 421-428. doi: Doi 10.1111/J.1467-9280.2007.01916.X

Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion, 6*(1), 125-138. doi: 10.1037/1528-3542.6.1.125

Lindquist, K. A., & Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emotion Review, 5*(1), 66-71.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behav Brain Sci, 35*(3), 121-143. doi: 10.1017/S0140525X11000446

Norenzayan, A., & Heine, S. J. (2005). Psychological universals: what are they and how can we know? *Psychol Bull, 131*(5), 763-784. doi: 10.1037/0033-2909.131.5.763

Roberson, D., Damjanovic, L., & Kikutani, M. (2010). Show and Tell: The Role of Language in Categorizing Facial Expression of Emotion. *Emotion Review, 2*(3), 255-260. doi: Doi 10.1177/1754073910361979

Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition, 35*(7), 1814-1829. doi: Doi 10.3758/Bf03193512

Rosenberg, E. L., & Ekman, P. (1995). Conceptual and Methodological Issues in the Judgment of Facial Expressions of Emotion. *Motivation and Emotion, 19*(2), 111-138. doi: Doi 10.1007/Bf02250566

Russell, J. A. (1991). Culture and the Categorization of Emotions. *Psychol Bull, 110*(3), 426-450. doi: Doi 10.1037/0033-2909.110.3.426

Russell, J. A. (1993). Forced-Choice Response Format in the Study of Facial Expression. *Motivation and Emotion, 17*(1), 41-51. doi: Doi 10.1007/Bf00995206

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

Russell, J. A. (1995). Facial Expressions of Emotion - What Lies Beyond Minimal Universality. *Psychol Bull, 118*(3), 379-391. doi: Doi 10.1037//0033-2909.118.3.379

Russell, J. A., & Widen, S. C. (2002). A label superiority effect in children's categorization of facial expressions. *Social Development, 11*(1), 30-52. doi: Doi 10.1111/1467-9507.00185

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 of the United States of America, 107*(6), 2408-2412. doi: 10.1073/pnas.0908239106

Thielscher, A. & Pessoa, L. (2007). Neural correlates of perceptual choice and decision making in Fear-Disgust discrimination. *Journal of Neuroscience, 27,* 2908-2917.

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? *Cogn Emot, 25*(5), 898-906. doi: 10.1080/02699931.2010.508270

Widen, S. C., & Russell, J. A. (2008a). Children acquire emotion categories gradually. *Cognitive Development, 23*(2), 291-312. doi: Doi 10.1016/J.Cogdev.2008.01.002

Widen, S. C. & Russell, J. A. (2008b). Young children's understanding of other's emotions. In M. Lewis, J. M. Haviland-Jones (Eds.), *Handbook of emotions, third edition* (pp. 348-363). New York, NY: Guilford.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63*(3), 271-313. doi: Doi 10.1016/S0010-0277(97)00003-6

Footnotes

[i] Unfortunately, misconceptions are present in the published record regarding which experiments tested the most remote cultural samples. For example, Norenzayan and Heine (2005) cite Ekman and colleagues' 1969 paper as an example of "comparing expressions across cultures with minimal cultural history and contact with each other" (p. 767). (Technically, a remote cultural sample that fits this description was not reported on until two years later in Ekman and Friesen (1971).)

Chapter II

Cultural relativity in perceiving emotion from vocalizations

A test of the universality of attribution hypothesis for vocal cues. This study replicated and extended recent work in order to examine whether vocalizations are cross-culturally perceived from the voice in a task that does not provide conceptual context. Further, we examined whether perception of affective dimensions are perceived from the voice across cultural contexts.

Cultural relativity in perceiving emotion from vocalizations

Maria Gendron,[1,2] Debi Roberson,[3] Jacoba Marietta van der Vyver[4] & Lisa Feldman Barrett[5,6]

[1]Department of Psychology, 300 McGuinn Hall, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA, USA 02467
[2]Affective Science Institute and Department of Psychology, 125 Nightingale Hall, 360 Northeastern University, 360 Huntington Avenue, Boston, MA, USA 02115
[3]Department of Psychology, Room 4.709, University of Essex, Colchester CO4 3SQ, U.K.
[4]School of Nursing and Public Health, University of Namibia, Main Campus, Block M, Room M209, Mandume Ndemufayo Avenue, Windhoek, Namibia
[5]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA 02129
[6]Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA 02144

Classification: SOCIAL SCIENCES
Correspondence to:
Maria Gendron
E-mail: gendroma@bc.edu

Abstract

A central question in the study of human behavior is whether or not certain categories of emotion, such as anger, fear and sadness (termed "discrete emotions"), are universally recognized in human vocalizations (termed the "universality hypothesis"). A recent paper published in *PNAS* [Sauter DA, Eisner F, Ekman P, & Scott SK, (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America* 107(6):2408-2412.] reported that members of the Himba ethnic group in Namibia recognized discrete emotions in non-verbal human utterances such as sighs, laughs, and screams, providing apparent support for the universality hypothesis. Yet, prior research has shown that emotion concept words embedded in the experimental task inflate the evidence for universality, and that removing the influence of those words can disrupt the perception of discrete emotions in US perceivers. In the present experiment, we predicted, and found, that when emotion concept words are not introduced as part of the experimental task, Himba participants failed to perceive discrete emotions in Western vocalizations intended to portray emotion. Importantly, Himba participants were able to perceive more basic properties of valence (positivity or negativity) and arousal (high or low activation). US participants did perceive discrete emotions in the vocalizations, but had lower agreement with the expected emotion than reported in prior studies that introduce concept words. Our findings indicate that the voice can reliably convey affective meaning across cultures, but that perceptions of emotion in a voice are culturally specific.

Think about the last time you heard someone sigh, chuckle, or groan. Perhaps you concluded that the person was tired, amused, or frustrated (respectively). Perceiving mental states from vocal cues is automatic and effortless, leading to the assumption that as perceivers, we "recognize" emotions in non-verbal cues and that emotion recognition is a universal human capacity (1). This is called the *universality hypothesis*. Originally developed to understand how people read facial expressions, the universality hypothesis assumes that (barring illness), all humans around the globe innately express the same internal mental states (i.e., emotions) on the face (as the universal, non-verbal behaviors); for example, scowling in anger is supposed to be innate. Furthermore, people from very different cultures should be able to recognize emotions in the same non-verbal expressions; for example, in all cultures, perceivers should be able to gaze upon a scowling person from any culture in the world and instantly recognize that he or she is angry. In strong versions of the universality hypothesis, emotion "recognition" is presumed to proceed directly from perceiving the perceptual regularities in non-verbal behaviors (e.g., people scowl in anger and only anger) and is not dependent on language and associated conceptual knowledge (2). Furthermore, emotion expression and perception are presumed to have co-evolved (3, 4) so that they are inborn and neither depends on learning across development (5). The universality hypothesis is so popular that it has become part of the story of what it means to be human in Western culture: It is standard curriculum in psychology textbooks, has been the subject of magazine articles in the New Yorker and National Geographic, has been a topic for science shows like "Radiolab", and is even the subject of a syndicated television show ("Lie to Me"). Government spending on security training (6, 7) is based, in part, on the assumption that it is possible to read a person's intent from non-verbal behaviors, regardless of cultural background or life experiences.

Of the hundreds of published cross-cultural experiments on emotion perception (8), only six provide a strong test of the universality hypothesis (see Table S1) because they used a method called the *two-culture approach* (9): they examined whether participants from maximally distinct cultural backgrounds, with limited exposure to Western culture, could decipher emotions in Western-style expressions. A recent paper in *PNAS* (10) is a good example of the two culture approach, testing whether Himba individuals residing in remote villages in the Kunene region of northwest Namibia (and largely isolated from Western cultural influences) perceived Western non-verbal vocal utterances (laughs, screams, sighs, etc.) in line with their intended "universal" emotional meaning (which looks identical to the Western model); for example, were perceivers able to recognize laughs as "amusement", screams as "fear", sighs as "relief" and so on? On each trial, participants were asked to select which of two vocalizations (e.g., a sigh and 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), leading Sauter et al. to claim support for the universality hypothesis. To our knowledge, this was the first and only study to examine the universality hypothesis with vocal cues using participants from a remote culture.

Although the universality hypothesis has largely been accepted as fact in both popular and many scientific circles, evidence against the idea of universal emotion perception steadily accumulates. First, there is growing evidence for substantial cross-cultural variation in emotion perception (for a review see 11). In a recent paper published in *PNAS*, for example, researchers used random generation of facial expressions to show that people from different cultures have dramatically different mental representations of emotions on the face (12), presumably because they have been exposed to different expressions in daily life. Second, and perhaps most important for our purposes here, is the observation that two-culture studies supporting the universality hypothesis (listed in Table S1) have a common element that appears to inflate the degree of agreement in emotion perception across cultures: they explicitly include emotion words in the experimental task. Participants in these studies were asked to select an expression to match a single emotion word, typically embedded in a longer description (e.g., 10). Of the six studies that provide the best empirical test of the universality hypothesis, the four experiments including emotion words within the method provided support for the universality hypothesis (i.e., people from culturally isolated groups and Western participants both perceived emotion in non-verbal stimuli according to the Western pattern). The two experiments that did not constrain responses with a small set of emotion words but asked participants to freely label the non-verbal expressions (i.e., to think of and nominate their own words as labels) did not find support for universality (i.e., people from culturally isolated groups and Western participants did not perceive the same emotions in the non-verbal stimuli). More generally, free-labeling experiments do not find strong agreement on the putative "universal" response even in Western participants viewing posed Western-style expressions (13-15). By presenting an non-verbal expression to a participant and asking him or her to choose an appropriate label from a small set of alternatives (or presenting several non-verbal expressions and asking the participant to pick the one that best matches an emotion adjective), experiments dramatically increase what has been called “recognition accuracy” (i.e., their ability to choose an answer that matches the Western pattern that the experimenters expect) over what would be observed when the expressions are presented alone and participants are asked to report on their own spontaneous perception.

The power of words in producing “emotion recognition accuracy” (11, 16) is clearly demonstrated in a number of lines of recent research. For example, an experimental procedure called “semantic satiation” (e.g., 17) can be used to reduce the accessibility of emotion word knowledge, which in turn dramatically reduces a person’s ability to perceive emotion in facial expressions (18), because words help to shape the underlying perceptual representation of those faces in the first place (19-21). Individuals who have a permanent loss of word and object knowledge resulting from neurodegeneration of the left anterior temporal lobe (i.e., people with semantic dementia) do not perceive scowls as anger, pouts as sadness, smiles as happiness, and so on, although these patients can distinguish which facial expressions are positive, negative, and neutral (i.e., they can make simple “affective” distinctions) (22). These findings are consistent with recent meta-analytic findings that the anterior temporal lobes and left ventrolateral prefrontal cortex (regions involved in language processes such as semantic representation and retrieval) show a consistent increase of activity during neuroimaging studies of emotion perception (23). Furthermore a patient suffering from language impairments (i.e., semantic aphasia) demonstrated difficulty grouping together posed facial expressions into the discrete emotion categories that those faces are designed to portray (24). Even the developmental psychology literature points to the importance of emotion words in emotion perception young in children. While young infants can perceive the positive/negative or arousing properties in non-verbal behavior (25) (i.e., affect perception is present), carefully controlled experiments reveal that what looks like an infant's ability to recognize emotion in non-verbal displays (i.e., discriminating a toothy smile depicting happiness from a closed mouth scowl depicting anger) actually turns out to be more simplisitic featural processing (infants were distinguishing "toothy" vs "nontoothy" faces) (26). Furthermore, young children who do not yet have a fully differentiated emotion lexicon group facial expressions together based on valence (whether the face is depicting a pleasant or an unpleasant facial expression) but the ability to distinguish scowls, frowns, nose-wrinkles and so on only emerges as children acquire linguistic categories for discrete emotions (27).

Taken together, perceivers in emotion recognition experiments, whether conducted only in the US or cross-culturally, do not appear to be merely "recognizing" emotion in the perceptual features of a non-verbal expression alone. Instead, they match faces or vocalizations to a smaller, constrained set of word alternatives. These alternatives represent experimenters' expectations that all participants will be recognizing emotion according to a Western model. As a result, in the two-culture studies, the presence of emotion words in the experimental method appears to be a critical element producing higher agreement on the "correct" responses, thereby increasing the appearance of universality (c.f., 28).

A simple but strong test of the universality hypothesis with vocal cues, then, would require participants from a cultural context that is relatively isolated from Western cultural influences (such as Himba individuals who live in remote villages in northwestern Namibia) to *freely label* Western vocalizations, rather than choose a response from a small, highly selected group of emotion words. If both Himba and US participants spontaneously label the vocalizations as discrete emotions in agreement with the experimenter's expectations, then this would provide support for the universality hypothesis. Alternatively, cultural specificity would be supported if Himba participants fail to label the vocalizations with the predicted emotion words when US participants labeled the vocalizations with the expected discrete emotion words. Given the research showing that presumed universal pattern often fails to materialize even in US participants in free-labeling studies (13, 15, 29, 30), a third possibility is that neither the Himba nor the US participants would produce the expected discrete emotion words, or that the US participants would do so at a level of agreement that was below what was reported in Sauter et al (10). Finally, there is ample evidence that facial and vocal cues are perceived in terms of the pleasantness or unpleasantness (valence) and the level of activation or quiescence (arousal) that they communicate (31, 32). There is evidence that these "affective" properties are universally perceived (33). Support for the universality of affect perception would occur if Himba participants spontaneously labeled the vocalizations in agreement with the valence or arousal of the portrayed emotion, including responses that "confuse" discrete emotion categories (e.g., a scowl would be correctly labeled for valence as "anger", "fear", or "disgust").

Participants in our experiment were sampled from the Himba ethnic group, as in Sauter et al. (10). The Himba people live in remote villages spread throughout the Kunene region of Northwest Namibia (historically known as the Kaokoland). The Otji-Herero language spoken by the Himba contains a set of emotion words that have been translated into English words for emotion in prior research (see 10), suggesting that it should be possible to find evidence of universal emotion perception in vocalizations if such universality exists. Our comparison group was sampled from the Boston Museum of Science (MoS, located in Massachusetts, USA). All participants from the Boston MoS were English-language speakers and were tested within the museum environment which well approximates the social and distracting nature of testing in the field in Namibia.

Participants completed a free-labeling task where on each trial, a participant listened to one of 18 non-word vocalizations and nominated a word to label the emotion being portrayed that vocalization. The stimuli (taken from 34) portrayed the same emotions as in Sauter et al, but were posed by American (as opposed to British), individuals making them a within culture stimuli for our sample of participants. The vocalizations posed Western portrayals of nine different discrete emotions (two utterances for amusement, anger, disgust, fear, relief, sadness, sensory pleasure, surprise, and triumph were presented to each participant; see Table 1). Each participant wore headphones so that the researcher (and the translator, who was present for the Himba participants) were blind to the particular stimulus that the participant was exposed to on any given trial. The participant's response was recorded verbatim on each trial (in the case of the Himba participants, their exact response was translated into English and recorded). Each participant was tested individually and no feedback was offered or provided. A response was coded as having "discrete agreement" if the participant offered a specific emotion word that was expected (e.g., "angry" or "anger" for a growl) or a close synonym (e.g., "mad").

We also computed an index of valence-based agreement (e.g., the word “sad” was used to label a growl because both sadness and anger are unpleasant or negative states) and an index of arousal-based agreement (e.g., the word “angry” was used for a “woohoo” where the correct response would be “triumph” because both anger and pride are states of high activation).

Data for the discrete emotion, valence and arousal codes just described were analyzed by comparing the mean percent agreement against zero (following omnibus ANOVA tests). This was a liberal test of the universality hypothesis, because any agreement statistically above zero would be considered intact perception within a cultural context, despite the possibility that “correct” agreement could be driven by only a few participants’ correct response.

Results

Emotions are Not Universally Perceived in Human Vocalizations

Our results do not support the hypothesis that emotions are universally “recognized” in vocal utterances, and instead support the hypothesis that emotion is perceived in vocalizations in a culturally relative manner (Table S2). An ANOVA on mean percentage of 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 difference was qualified by a significant emotion x 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 more likely to labeled screams as fear with a frequency that was significantly greater than zero, at 54% and 20%, respectively. For all other categories of emotion, the Himba participants provided word labels that agreed with the universal solution less than 5% of the time (which did not differ significantly from zero), suggesting that most of the vocalizations were not perceived similarly across cultures. Further, Himba participants did not label “eww” vocalizations, growls or crying as disgust, anger and sadness, respectively, countering claims that these emotions are universally perceived in vocalizations. One reason why Himba participants had lower agreement rates may reflect a cultural tendency to describe vocalizations in behavioral/situational rather than mental terms (see SOM). It is important to note that these results reflect a liberal test of “cross-cultural” perception, since very low means could reflect the intact performance of only a few individuals. When we submit these data to more stringent “chance” criteria accounting for the number of stimuli in the experiment and the valence/arousal conveyed by a given stimulus, the results are even less consistent with the universality hypothesis (see SOM).

While US participants tended to produce labels that agreed with the putative universal solution, they did at much lower levels of agreement that reported in experiments that ask participants to match a vocalization to an emotion word from a small set of words provided by the experimenter (10, 34, 35). For example, in Sauter et al, British participants “recognized” between 75% and 100% of the vocalizations they were presented with, dwarfing performance in our experiment. Importantly, US perceivers ”recognition” in our experiment is relatively modest when considered from the standpoint of ecological validity – on average, almost 40% of the time, perceivers are mislabeling exaggerated vocalizations that are supposed to be diagnostic of disgust or sadness, almost 35% of the time mislabeling growls that are supposed to be diagnostic of anger, and so on. Furthermore, without providing a concept for participants, US perceivers did not correctly label sounds depicting triumph (achievement)--with agreement levels that were completely at floor, suggesting that this vocalization does not have a culturally conveyed, let alone universal, meaning.

Affect is Perceived Universally in Human Vocalizations

Valence. Our results support at least minimal universality of valence perception (distinguishing pleasant, neutral, and unpleasant states) in vocal utterances. An ANOVA on mean percentage of valence-based agreement, with cultural group as the between subjects factor and emotion category as the within subjects factor, revealed that both US and Himba participants were able to freely label the utterances correctly for valence at greater levels than would be expected by chance (with the exception of portrayals of surprise by Himba participants), at 75% vs. 50.46%, respectively. 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 x cultural group, $F(8, 368)=13.273$, $p<.001$, $\eta_p^2=.224$ (See Figure 2; Table S3). When compared to US participants, Himba participants were less likely to freely label vocalizations of disgust, fear, and, sadness with negative emotion or affect terms, and they were less likely to freely label vocalizations of sensory pleasure with positive emotion or affect terms. Again, this may reflect the tendency of Himba participants to provide content that was not related to mental states (see SOM). US and Himba participants were equivalently likely to perceive positivity in vocalizations for triumph and negativity in vocalizations for anger. Interestingly, Himba participants were more likely than US participants to level vocalizations of amusement as positive (p<.001, 2-tailed).

Arousal. Our results also provide some limited support for minimal universality of arousal perception (distinguishing activated, neutral, and deactivated states) in vocal utterances. The perception of arousal from vocalizations is less robust cross-culturally than valence, particularly because Himba participants appeared to have difficultly correctly labeling lower arousal states of relief, sadness, and sensory pleasure (see Figure 3; Table S4). An ANOVA on mean percentage of agreement, with cultural group as the between subjects factor 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 than did US participants, $F(1, 46)=60.259$, $p<.001$, $\eta_p^2=.567$. Whereas US participants perceived arousal with comparable accuracy to valence at 72.69% (SD=32.46), Himba participants' agreement level was considerably lower, at 37.03% (SD=32.95). The effect of cultural group was qualified by an interaction between emotion x 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 employed descriptions that did not contain mental state terms.

Discussion

Which mental states are universally perceived from vocalizations?

When emotion words are embedded in emotion perception tasks, researchers find evidence for universal emotion "recognition" (36-38). When the experimenter does not provide emotion words to constrain perceiver options, evidence for universality is considerably weaker, if it is observed at all (39), even in US participants (13-15). In our experiment, where participants listened to human vocalizations that were explicitly designed by experimenters to universally depict emotions (34), and freely labeled those vocalizations, evidence for universal emotion perception was not found. Himba participants did not label the vocalizations with discrete emotion words that would be expected according to the hypothesized universal (but largely Western) pattern, even though such words are available in the Herero language spoken by Himba participants. Even US participants, who produced the expected labels did so at a considerably lower level of agreement than was observed for exactly the same stimuli in another US sample (34). Our findings are consistent with prior research showing that spontaneously produced labels produce much lower agreement than responses selected from a pre-determined list (c.f., 28). Taken together, our findings are consistent with a growing number of studies showing that *emotion perception* is culturally relative (for a review see 11), and that evidence for universal emotion "recognition" is highly dependent on the experimental context (28).

Our findings suggest that *affect perception* shows some evidence of universality. Himba individuals (who are relatively isolated from Western influence) were able to perceive vocalizations from US individuals as valenced (pleasant, neutral or unpleasant), and to some extent as high or low arousal (activated, neutral, or deactivated). The affect perception findings are somewhat inconsistent with evidence that acoustical markers (i.e., fundamental frequency and amplitude) correlate more with the arousal than with the valence of an utterance (40). One reason for this discrepancy may be that the posed, categorical stimuli used in our experiment fail to capture the full range of vocal qualities in naturalistic vocalizations. Furthermore, the stimulus set is optimized to distinguish perceptions of emotion and therefore might lead participants to emphasize valence over arousal perception when selecting a verbal label response. A direct arousal perception task might reveal that arousal is indeed universally perceived from the voice more readily.

The idea that affect perception is closer to a core human capacity (41) might provide an alternative explanation for the Sauter et al findings that have previously been interpreted as evidence for the universality hypothesis. Recall that in their experiment, Himba and US participants were presented with an emotion word (e.g., disgust) embedded in a story and then heard two vocalizations, and their task was to choose the vocalization that best matched the story. On half of the experimental trials, participants were asked to choose between two utterances that differenced in valence. For example, in response to the story “Someone has just eaten rotten food and feels very disgusted”, participants would have heard a positive utterance such as a portrayal of amusement (laughter) and a negative utterance such as a portrayal of disgust (“ewww”); participants could have picked the correct utterance by perceiving affective valence alone, having nothing to do with emotion perception per se. On other trials, the distractor vocalization shared the same valence as the correct choice but may have differed in arousal level. In response to the same story, for example, participants might have heard a low arousal utterance such as someone portraying sadness (crying) vs. a high arousal utterance such as someone portraying disgust ("ewww"); participants could have distinguished the two vocalizations on their arousal level alone. Thus, the "above chance" performance of Himba participants in the Sauter et al. study might have reflected valence perception on half the trials and arousal perception on some additional proportion of trials (left unspecified in their method description), rather than the universality of emotion perception per se.

Why Only Minimal Universality of Affect? Perceiving Actions Instead of Mental States in Vocalizations

Himba participants' failure to perceive the intended discrete mental states suggests that linkages between specific vocalizations (e.g., crying) and specific mental states (e.g., sadness) are not cross-culturally preserved. Yet Himba participants produced a fair amount of content to describe the vocalizations that did not reference mental states. That is, Himba participants often described the vocalizations they heard in terms of situations (e.g., someone is sick) and actions (e.g., screaming), rather than as evidence of mental states per se (disgust and anger, respectively) (see SI for additional analyses of this content). Himba participants were able to provide behavioral and situational responses that were tailored to the stimuli on 12% of the trials. This is perhaps not surprising given that the stimuli used in this experiment (e.g., a disgust vocal pose) were originally elicited by asking the posers to imagine a particular situation (e.g., "You come in contact with something physically noxious or contaminating"), and to produce the vocalization that would occur in that context. It may be that there is some minimal universality to the tuning of these discrete vocalizations to different situations, regardless of culture. This raises the question of whether the Himba participants were, in some capacity, having a "percept" of emotion. On the one hand, social psychological research on action identification has demonstrated that describing an action at a physical level is distinct from the mental level, such that the former does not require a representation of the other person's internal state (42). Furthermore, emotion perception, as defined in a Western cultural context, is also meant to involve a representation of another individual's mental state. Thus the action identification literature suggests that the Himba participants did not perceive emotion on trials where they produced behavioral descriptors. Alternatively, it is possible that the Himba conception of emotion does not involve mental states to the same degree; although this is speculative, there is documented cross-cultural variability in the concept of emotion itself (43), with some mental taxonomies lacking a word for emotion entirely, such as in early Buddhist writings (44) or in the German language (43), where no generic term for emotion exists. The prevalence of behavioral labels of vocalizations in this experiment may reveal a cultural difference in the tendency to infer mental states as causes for actions, more generally, although future research would be required to explore this possibility.

Limitations

One potential limitation of the present experiment was the use of translated, verbal responses as our primary dependent variable of interest. This concern is mitigated by the fact that our translator also worked on the Sauter et al study (45), which reduces the possibility that differences in results between the two studies are attributable to simple differences in translation. Another potential limitation is that we did not use Himba vocalizations as stimuli with both Himba and US perceivers. Although the one-directional test of Western vocalizations (that had been previously normed and published for the purposes of studying the universality hypothesis) (34) should be sufficient to examine whether the Western cultural model is in fact the universal one. To properly explore the observed cultural relativity in emotion perception, however, future research should involve collection of vocalization stimulus sets from other cultures as well as in Sauter (10). Finally, like the majority of experiments that are published on emotion perception, we used posed, portrayed vocalizations of emotion (i.e., the "portrayal paradigm"; 46), rather than spontaneously elicited vocalizations. Prior research has shown that evidence of universality is typically enhanced with the use of posed stimulus sets which tend to be less variable and more caricatured than spontaneous stimuli (c.f. 47, see also 48). The fact that we did not find evidence of universality cannot be explained by stimuli lacking in sufficient statistical regularity or "source clarity", a critique leveled against much of the literature on relativity in perception. Nonetheless, future research on cultural relativity in emotion perception might be best served by using spontaneous, naturally occurring vocalizations, in order to gain more ecological validity .

One unexpected finding is that participants were better able to perceive affect in certain vocalizations (e.g., a triumphant "woohoo") when compared to others (e.g., a sigh of relief). This may be the product of variability in the underlying mechanisms that are involved in generating these sounds to begin with. At least two mechanisms for producing vocalizations have been identified. In what Owren calls the "production-first" system (49) and Scherer calls "push" effects (50), vocalizations are produced reflexively by the physiological changes accompanying affective arousal. In what Owren calls the "receptive first" system, and Scherer's "pull" effects, vocalizations can also be "ritualized or conventionalized" acoustical patterns that require experience and can be volitionally produced even in the absence of affective changes in experience (e.g., the exclamation "yuck!"), even if it often accompanies affective states. It may be that vocalizations where affect is not perceived cross-culturally are the sole products of this second (non-affective) mechanism, and entirely dependent on experience and culture. If this is the case, what looks like evidence for "minimal universality" of affect perception in our findings may actually reflect universality of affect perception, but relativity of some culturally specific, non-affective vocalizations. Future research should examine these questions more directly.

The Universality of Emotion

Universal emotion recognition has been taken as evidence for the universality of emotion generation – if everyone around the world can recognize emotion expressions, then these expressions must be universally produced (i.e., there must be perceptual regularity in production across cultures; c.f. 28). Many psychologists and researchers in related disciplines have hypothesized that emotions like sadness, anger, disgust and fear have universal, biological cores (51) that are present from infancy (5), can be detected in the brain (52), peripheral nervous system (53), face (45), and voice (54), and can be readily decoded by perceivers (4). Yet recent reviews of the empirical evidence call into question claims about the biological basicness of each emotion category based on the brain (55), the peripheral nervous system (56), and the face (57). Such findings are consistent with emerging research, including the data presented here, that emotion perceptions appear not to be universal. Instead, the present findings are more consistent with an alternative view of emotions as conceptualized affective states (58, 59) that are highly culturally variable (60) and that are the product of more basic "core" human capacities.

Conclusion

In this study we show that discrete emotions are not universally perceived in vocalizations. Without the context of a word/story and a limited range of responses options, Himba participants did not perceive the same emotions from vocal burst portrayals as English speaking participants. In fact, removing the words reduced agreement in US participants compared to what has been demonstrated in other studies, suggesting that the prior literature also overestimates the robustness of within culture emotion perception from the voice. We also show that both Himba and US participants were able to perceive whether the target individual is experiencing positive or negative affect, and high or low arousal levels (although to a lesser extent), suggesting that these affective dimensions may be better candidates than emotion for a "core" human capacity.

Materials and Methods

Participants

Himba. Participants were 24 native Herero speakers from the Himba ethnic group (12 male, 12 female; mean age= 35.96, SD=14.5)[§]. Data were collected in three remote villages, all located in the Mountainous Northwest region of Namibia. None of these villages had an established school or other signs of outside presence. Most members of the Himba communities remained within their villages and had limited contact with outsiders (although one of the men in one village owned a truck that he used for trips to a nearby town for supplies). All testing locations were far from the regional towns, and there was no evidence of tourism to these villages as is typical for those closer to towns. None of the participants spoke fluent English (although one participant was able to exchange simple greetings in English).

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); participants included 22 Caucasian individuals, 1 Asian individual and 1 African American individual. All participants reported being native English speakers. Data were collected in the busy museum environment that was similar in activity level to the field testing environment in Namibia. 17 participants had at least a bachelor's degree and the sample had a median yearly household income range of $87,500-99,000.

Stimuli

The 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*). These vocalizations were previously validated as prototypical tokens of emotion within a Western cultural context (34) and are similar to those used in Sauter et al (2010), except that we used vocalizations for *triumph* rather than Sauter et al.'s use of *achievement* vocalizations (on the assumption that pride is a better candidate for universality;). We created two stimulus sets of 18 stimuli each. Each participant in our study was presented with one set (including one male and one female exemplar for each emotion category). The stimuli were cleaned for ambient noise on the audio track and the mean peak amplitude was adjusted to be equivalent in all stimuli using Audacity audio editing software (http://audacity.sourceforge.net/).

Procedure

This research was approved by the Northeastern University Institutional Review Board's Office of Research Protections. All participants were consented prior to the start of the experiment. Himba participants were verbally consented through a translator.[¶] US participants were consented with a standard form matched for informational content with the verbal consent. Participants were outfitted with headphones and verbally instructed (via the use of a translator for Himba participants) in the following manner: *"You are going to hear some people make sounds with their voices. The people you will hear are feeling different emotions. We want to you try and figure out what emotion they feel based on how their voice sounds. Please do your best to come up with a single label to describe the feeling the person is having. If you cannot come up with a single word, you can give us a few words to describe the feeling. Before you hear each sound, you will hear a [beep]. That sound tells you that the voice is coming and to pay close attention. Once you hear the voice, go ahead and tell us what emotion you think the person is feeling."*

On a given trial, a "ready?" prompt came on screen. If the participant appeared ready (i.e., not visibly distracted by the surroundings), the trial proceeded. On some occasions, we paused for several seconds to accommodate participants so that they would be able to fully attend to the stimulus presentation. In those cases, the participant was verbally asked if he/she was ready before proceeding. Once the experimenter pressed the spacebar, the trial started. After 250 ms a tone was played to cue the participant to the onset of the stimulus, followed 250 ms later by a vocalization stimulus. Following the stimulus, participants typically verbally responded to the stimulus without any prompting. If the participant did not start to provide a verbal response after a second or two, we prompted the participant *"What emotion did you hear?"* The translator provided an immediate translation[‖] of the verbal response of the participant and this translated content (or original response for American participants) was entered into the laptop computer by the experimenter. If the participant 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?"* Responses were entered in the order they were provided such that any contextual content (i.e., situational, behavioral, or physical state) provided was always recorded in addition to any mental state terms generated.

Data Coding

Prior to any coding, all task and individual level trial information was removed and the trials were fully randomized such that the data were not nested by culture or individual. That is, coders were blind to culture as well as the individual a given response was from. Removing participant level information ensured that no clues to a given participant's cultural context that could be extracted from a given response would affect the coding of the remainder of that participant's data. The data were coded in two different sets modeled after Russell (1990) and overviewed below.

The data were independently coded by two trained individuals. Researchers subjectively coded responses on a given trial relative to the emotion portrayed by the stimulus vocalization (no other trial level information was available to the coders). Three sub-codes were created: discrete emotion, valence and arousal. Participants rated whether content agreed or disagreed with the stimulus. In addition to coding for agreement, both coders indicated instances when “no content” was available. This allowed for us to establish whether there was reliability regarding when mental content on each of the three dimensions was specified. Final percentage agreement was computed across all trials of a given emotion (i.e., regardless of when mental content was specified or not). Reliability between the two researchers (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. Support for universality of discrete emotion perception would be evidenced by high (and significantly above 0) accuracy for the discrete emotion code, regardless of cultural context. Support for relativity of discrete emotion perception, on the other hand, would be evidenced by high agreement for American participants, but low (and not significantly different than 0) accuracy for Himba participants.

Acknowledgements

We thank Kemuu Jakurama for his translation services and assistance in collecting these data, Julia Reading and Hanna Negami for their assistance in coding these data, Emiliana Simon-Thomas and colleagues for the use of their vocal stimulus set, and Jules Davidoff and Serge Caparos for the use of their equipment and resources in the field. We also thank dissertation committee members Hiram Brownell and James Russell for their guidance and feedback on this research. This research was supported by a National Institutes of Health Director's Pioneer Award (DP1OD003312) to Lisa Feldman Barrett.

MG and LFB designed the research; MG, DR and JMV performed research; MG analyzed the data; MG, LFB, and DR wrote the paper

Footnotes

§ While there is a numerical system in Herero, most individuals in the communities we visited did not keep track of their age according to this system. Instead, most of our participants either provided estimates of their age or we estimated age for the participants based on their reproductive history, number and age of children if applicable as well as general physical appearance. As a result, the age distribution we report for our Himba participants is not exact.

¶ We used the same translator as the original Sauter et al. [Sauter DA, Eisner F, Ekman P, & Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America* 107(6):2408-2412.] work, making our testing context and translations maximally comparable.

‖ One notable limitation to our translations is for longer responses that lacked a feeling label. Because of the novelty of the translation process, some participants would provide extremely long verbal responses that were not easily translated in real time. If the participant gave a long response, in some cases the translator provided only partial content for that response (i.e., leaving out speech repetitions). In those instances, the translator focused his efforts on recording any emotion terms that the participant provided. The experimenter (MG) was familiar with the most frequent discrete emotion labels in Otji-Herero and can attest that the translator provided adequate and consistent translations for those terms. A notable limitation, however, is that the data do not capture the narrative nature of some of the longer responses provided by participants. In addition to an immediate translation, the translator recorded the specific Otji-Herero word/words he translated on a given trial.

References

1. Scherer KR (1994) Affect bursts. *Emotions: Essays on emotion theory*, eds van Goozen SHM, Van de Poll NE, & Sergeant JA (Erlbaum, Hillsdale, NJ).
2. Izard CE (1994) Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychological bulletin* 115(2):288-299.
3. Izard CE (2007) Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm. *Perspectives on Psychological Science* 2(3):260-280.
4. Schyns PG, Petro LS, & Smith ML (2009) Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: behavioral and brain evidence. *PloS one* 4(5):e5625.
5. Hoehl S & Striano T (2010) Discrete Emotions in Infancy: Perception without Production? *Emotion Review* 2(2):132-133.
6. Hontz CR, Hartwig, M., Kleinman, S. M. & Meissner, C. A (2009) Credibility Assessment at Portals, Portals Committee Report.
7. Weinberger S (2010) Airport Security: Intent to deceive? *Nature* 465:412-415.
8. Elfenbein HA & Ambady N (2002) On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin* 128(2):203-235.
9. Norenzayan A & Heine SJ (2005) Psychological universals: what are they and how can we know? *Psychological bulletin* 131(5):763-784.
10. Sauter DA, Eisner F, Ekman P, & Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America* 107(6):2408-2412.
11. Barrett LF, Mesquita B, & Gendron M (2011) Context in emotion perception. *Current Directions in Psychological Science* 20(5):286-290.
12. Jack RE, Garrod OG, Yu H, Caldara R, & Schyns PG (2012) Facial expressions of emotion are not culturally universal. *Proc Natl Acad Sci U S A* 109(19):7241-7244.
13. Boucher JD & Carlson GE (1980) Recognition of facial expressions in three cultures. *Journal of Cross-Cultural Psychology* 11:263-280.
14. Ekman P & Rosenberg EL (1997) *What the face reveals : basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (Oxford University Press, New York) pp xvi, 495 p.
15. Widen SC, Christy AM, Hewett K, & Russell JA (2011) Do proposed facial expressions of contempt, shame, embarrassment, and compassion communicate the predicted emotion? *Cognition & emotion* 25(5):898-906.
16. Barrett LF, Lindquist KA, & Gendron M (2007) Language as context for the perception of emotion. *Trends in cognitive sciences* 11(8):327-332.
17. Tian X & Huber DE (2010) Testing an associative account of semantic satiation. *Cognitive psychology* 60(4):267-290.
18. Lindquist KA, Barrett LF, Bliss-Moreau E, & Russell JA (2006) Language and the perception of emotion. *Emotion* 6(1):125-138.
19. Fugate JM, Gouzoules H, & Barrett LF (2010) Reading chimpanzee faces: evidence for the role of verbal labels in categorical perception of emotion. *Emotion* 10(4):544-554.

20. Gendron M, Lindquist KA, Barsalou L, & Barrett LF (2012) Emotion words shape emotion percepts. *Emotion* 12(2):314-325.
21. Roberson D, Damjanovic L, & Pilling M (2007) Categorical perception of facial expressions: evidence for a "category adjustment" model. *Memory & cognition* 35(7):1814-1829.
22. Calabria M, Cotelli M, Adenzato M, Zanetti O, & Miniussi C (2009) Empathy and emotion recognition in semantic dementia: a case report. *Brain and cognition* 70(3):247-252.
23. Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, & Barrett LF (in press) The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*.
24. Roberson D, Davidoff J, & Braisby N (1999) Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition* 71(1):1-42.
25. Leppanen JM, Richmond J, Vogel-Farley VK, Moulson MC, & Nelson CA (2009) Categorical representation of facial expressions in the infant brain. *Infancy : the official journal of the International Society on Infant Studies* 14(3):346-362.
26. Caron RF, Caron AJ, & Myers RS (1985) Do infants see emotional expressions in static faces? *Child development* 56(6):1552-1560.
27. Widen SC & Russell JA (2010) Differentiation in preschooler's categories of emotion. *Emotion* 10(5):651-661.
28. Russell JA (1994) Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin* 115(1):102-141.
29. Rosenberg EL & Ekman P (1995) Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion* 19(2):111-138.
30. Izard CE (1971) *The face of emotion* (Appleton-Century-Crofts, New York) pp xii, 468 p.
31. Russell JA, Bachorowski JA, & Fernandez-Dols JM (2003) Facial and vocal expressions of emotion. *Annual review of psychology* 54:329-349.
32. Bachorowski JA & Owren MJ (2003) Sounds of emotion: production and perception of affect-related vocal acoustics. *Annals of the New York Academy of Sciences* 1000:244-265.
33. Russell JA (1991) Culture and the categorization of emotions. *Psychological bulletin* 110(3):426-450.
34. Simon-Thomas ER, Keltner DJ, Sauter D, Sinicropi-Yao L, & Abramson A (2009) The Voice Conveys Specific Emotions: Evidence From Vocal Burst Displays. *Emotion* 9(6):838-846.
35. Hawk ST, van Kleef GA, Fischer AH, & van der Schalk J (2009) "Worth a Thousand Words": Absolute and Relative Decoding of Nonlinguistic Affect Vocalizations. *Emotion* 9(3):293-305.
36. Ekman P & Friesen W (1971) Constants across cultures in the face and emotion. *Journal of personality and social psychology* 17(2):124-129.
37. Ekman P, Heider E, Friesen WV, & Heider K (1972) Facial expression in a preliterate culture.

38. Ekman P (1972) Universals And Cultural Differences In Facial Expressions Of Emotions. *Nebraska symposium on motivation, 1971*, ed Cole J (University of Nebraska Press, Lincoln, Nebraska), pp 207-283.
39. Sorensen ER (1975) Culture and the expression of emotion. *Psychological Anthropology*, ed Williams TR (Mouton Publishers, Paris), pp 361-372.
40. Bachorowski JA (1999) Vocal expression and perception of emotion. *Current Directions in Psychological Science* 8(2):53-57.
41. Barrett LF & Bliss-Moreau E (2009) Affect as a Psychological Primitive. *Advances in experimental social psychology* 41:167-218.
42. Kozak MN, Marsh AA, & Wegner DM (2006) What do I think you're doing? Action identification and mind attribution. *Journal of personality and social psychology* 90(4):543-555.
43. Wierzbicka A (1999) *Emotions across languages and cultres: diversity and universals* (Cambridge University Press, Cambridge).
44. de Silva P (1995) Theoretical Perspectives on Emotion is Early Buddhism. *Emotions in Asian Thought: A Dialogue in Comparative Philosophy*, eds Marks J & Ames RT (State University of New York Press, Albany), pp 109-120.
45. Matsumoto D, Keltner D, Shiota M, Frank M, & O'Sullivan M (2008) Facial expressions of emotions. *Handbook of emotion*, eds Lewis M, Haviland-Jones JM, & Barrett LF (Macmillan, New York), pp 211-234.
46. Scherer KR, Johnstone T, & Klasmeyer G (2003) Vocal expression of emotion. *Handbook of Affective Sciences*, eds Davidson RJ, Scherer KR, & Hill Goldsmith H (Oxford, New York), pp 433-456.
47. Scherer KR (2003) Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40:227-256.
48. Naab PJ & Russell JA (2007) Judgments of emotion from spontaneous facial expressions of new guineans. *Emotion* 7(4):736-744.
49. Owren MJ, Amoss RT, & Rendall D (2011) Two organizing principles of vocal production: Implications for nonhuman and human primates. *Am J Primatol* 73(6):530-544.
50. Scherer KR (1995) Expression of emotion in voice and music. *Journal of Voice* 9(3):235-248.
51. Ekman P (1992) Are there basic emotions? *Psychological review* 99(3):550-553.
52. Panksepp J (1998) Affective neuroscience: The foundations of human and animal emotions. in *Series in affective science* (Oxford University Press, New York), p 1 online resource.
53. Lench HC, Flores SA, & Bench SW (2011) Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. *Psychological bulletin* 137(5):834-855.
54. Sauter DA, Eisner F, Calder AJ, & Scott SK (2010) Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology* 63(11):2251-2272.
55. Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, & Barrett LF (2012) The brain basis of emotion: a meta-analytic review. *The Behavioral and brain sciences* 35(3):121-143.

56. Kreibig SD (2010) Autonomic nervous system activity in emotion: A review. *Biological psychology* 84(3):394-421.
57. Cacioppo JT, Berntson GG, Larsen JT, Poehlmann KM, & Ito TA (2000) The psychophysiology of emotion. *Handbook of emotion*, eds Lewis R & Haviland-Jones JM (Guilford Press, New York), 2nd Ed.
58. Russell JA (2003) Core affect and the psychological construction of emotion. *Psychological review* 110(1):145-172.
59. Barrett LF (2009) The Future of Psychology: Connecting Mind to Brain. *Perspectives on psychological science : a journal of the Association for Psychological Science* 4(4):326-339.
60. Boiger M & Mesquita B (in press) The construction of emotion in interactions, relationships, and cultures. *Emotion Review*.

Figure Legends

Figure 1. Cross-cultural comparisons for emotion. Himba and American participants' performance for perception of discrete content from the voice. Mean percentage (± SEM) of discrete emotion response agreement based on coding (by two independent raters) of the mental state content of responses only is presented on the y-axis. Data are presented by culture and discrete emotion portrayal type (x-axis).

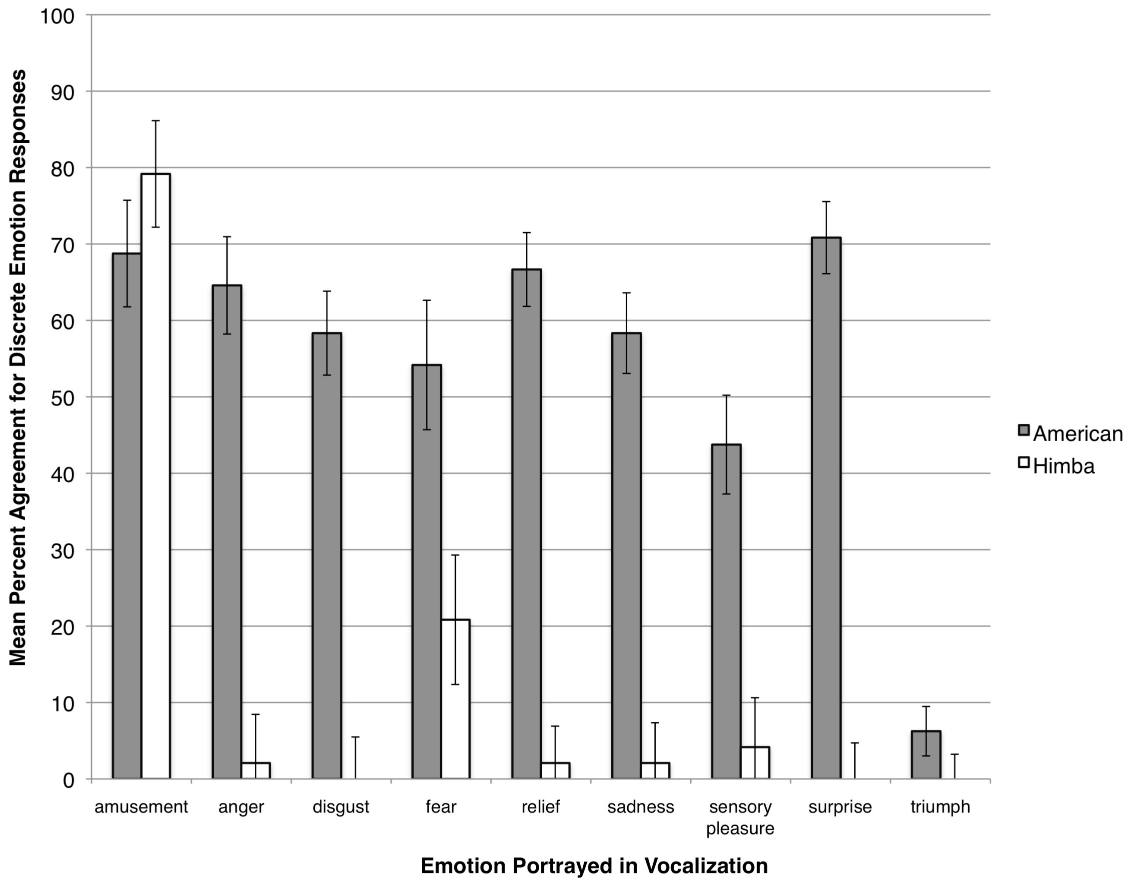
Figure 2. Cross-cultural comparisons for valence. Himba and American participants' performance for perception of valence (positive, negative and neutral) content from the voice. Mean percentage (± SEM) of valence response agreement based on coding (by two independent raters) of the mental state content of responses only 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. Himba and American participants' performance for perception of arousal/activation (high, mid, and low) content from the voice. Mean percentage (± SEM) of arousal response agreement based on subjective coding (of two independent raters) of the mental state content of responses only is presented on the y-axis. Data are presented by culture and discrete emotion portrayal type (x-axis).

Table 1. *Descriptions of Vocalizations by Emotion Portrayed*

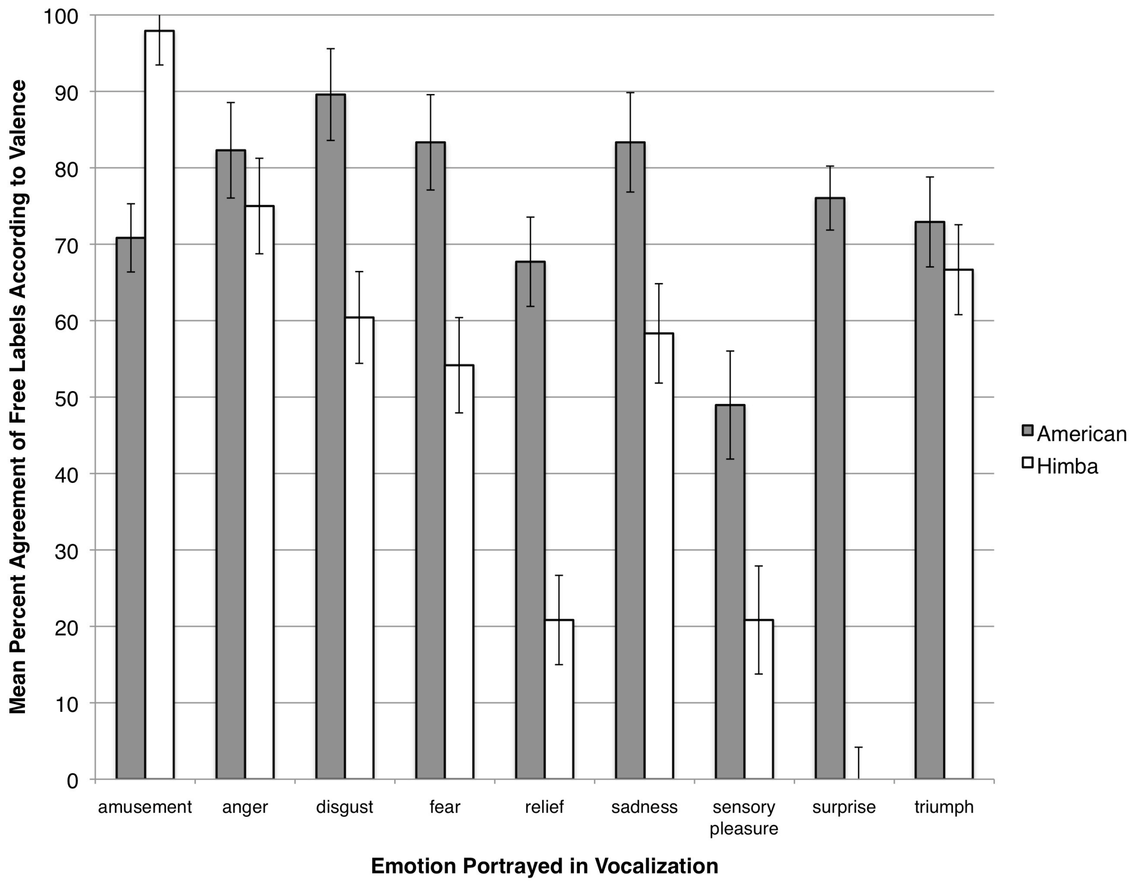
Discrete Emotion Portrayed	Vocalization
amusement	giggle/laughter
anger	guttural yell/growl
disgust	“ewww”
fear	scream
relief	sigh
sadness	cry
sensory pleasure	“mmm mmm”
surprise	“ahhh-ahhh”
triumph	“woohoo”

Figure 1.



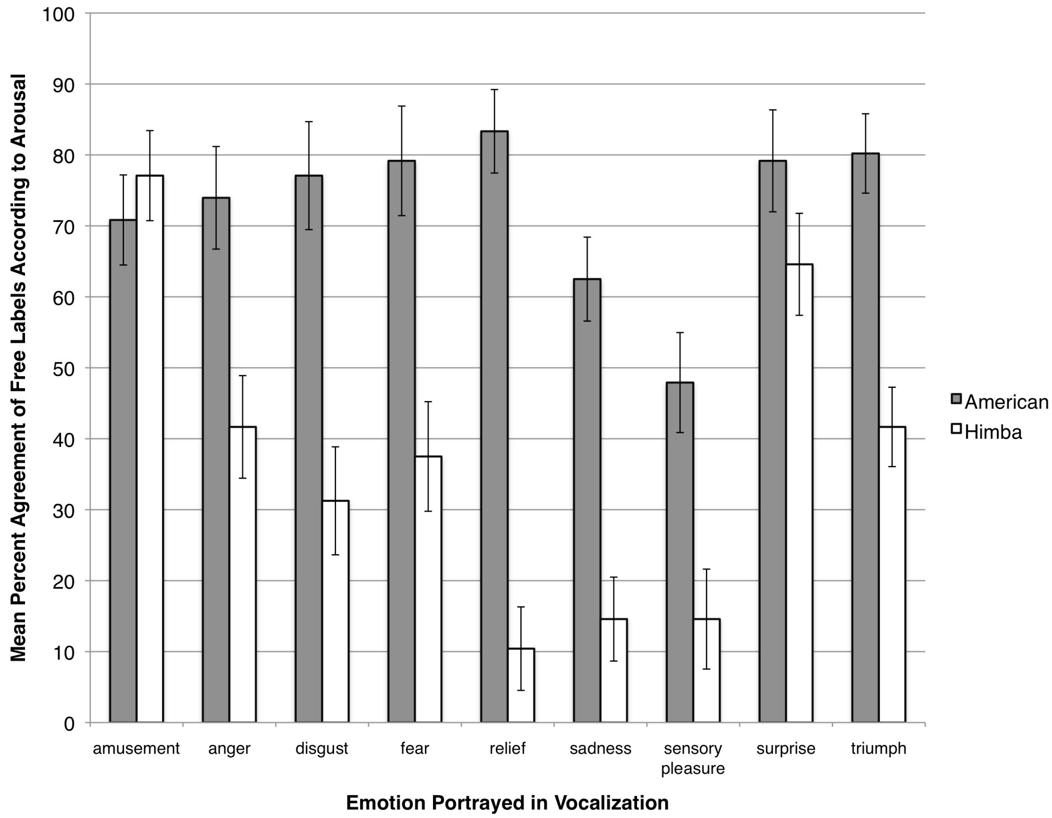
100
90
80
70
60
50
40
30
20
10
0
Mean Percent Agreement for Discrete Emotion Responses
amusement
anger
disgust
fear
relief
sadness
sensory pleasure
surprise
triumph
Emotion Portrayed in Vocalization
American
Himba

Figure 2.



Mean Percent Agreement of Free Labels According to Valence
100
90
80
70
60
50
40
30
20
10
0
amusement
anger
disgust
fear
relief
sadness
sensory pleasure
surprise
triumph
Emotion Portrayed in Vocalization
American
Himba

Figure 3.


Mean Percent Agreement of Free Labels According to Arousal
100
90
80
70
60
50
40
30
20
10
0
amusement
anger
disgust
fear
relief
sadness
sensory pleasure
surprise
triumph
Emotion Portrayed in Vocalization
American
Himba

Supplementary text:

Chance Level Response Analysis

We examined whether participants produced the "universally" correct labels more frequently than would be expected by chance (following Ekman and Rosenberg's analyses (6)) by comparing observed percentage agreements against a threshold of 1/9, given that there were nine different emotion categories portrayed by the vocalizations in our experiment. The results are presented in Table S5. For US participants, the overall pattern of results did not change from the more liberal test against 0. For Himba participants, perceptions of screams as portrayals of fear did not exceed chance levels in this analysis, in contrast to the significant effect in the more liberal test reported in the main text. This result cautions against the strong interpretation that fear is "universally" perceived from the voice at a group level, since only a few individuals from Himba culture labeled screams as fear. Further, Himba participants were significantly *less* accurate than chance, for a number of emotion portrayal types (i.e., anger, sadness and relief). Disgust, surprise and triumph portrayals elicited no correct responses from Himba participants (i.e., accuracy was at floor with a mean and standard deviation of 0) and thus did not undergo statistical testing.

In addition, we used a second set of more stringent thresholds for emotions that are similar in affective valence and arousal, allowing us to determine whether participants offered labels with any degree of specificity over and above what would be expected by their basic affective properties (again following Ekman and Rosenberg's analysis approach (6)). For example, vocalizations portraying anger, fear, and disgust are all unpleasant and highly aroused. If participants perceived those utterances in terms of affective properties of valence and arousal, then they would be equally to offer labels such as “anger” “fear” and “disgust” for screams. As such, to determine if participants’ labels communicated any information about discrete emotion over and above valence and arousal alone, we compared free-labeling responses for these three categories of vocalizations to what would be expected on the basis of affective perception, which was 1/3. Similarly, vocalizations portraying sensory pleasure and amusement are pleasant and moderate in their arousal. If participants perceived utterances in terms of affective perception, then they might be equally likely to offer labels such as “pleasure’ and “amusement” for laughs. As a consequence, to determine if participants were labeling discrete emotions (as compared to valence and arousal level), we compared free labeling responses for these categories to a criterion of ½. As can be seen in Table S6, participants did not perceive vocalizations as indicative of the same mental states across cultures, contrary to the universality hypothesis. The only emotion perceived by Himba participants significantly above chance was amusement from laughter vocalizations. This more stringent test also revealed less evidence in support of “universal” perception, such that agreement on sensory pleasure was not significantly above chance levels.

Action Identification vs Mental State Inference

One observation that was made when coding for Himba participants responses is that there were many instances where perceivers described the target person’s vocalizations physical (e.g., screaming) rather than mental (e.g., fearful) terms. The distinction between these two types of descriptions was the basis of Action Identification Theory (24), which posited that there is variation in how much mental state inference a person engaged in during action perception. For example, eating might be described as chewing and swallowing, or as getting nutrition, the latter implying an intention behind behavior. Perceiving emotion in a vocalization is an act of mental state inference, because an internal state is assumed to be responsible for the observable "expression"; this stands in contrast to perceiving the actions in more physical (and less psychological) terms. We tested whether adopting an action identification perspective might reveal more consistency in the responses of Himba participants. Two coders independently assessed for whether Himba and US participants labeled the vocalizations with action words (i.e., behaviors) and situational descriptions vs. mental state words (i.e., emotions). Reliability of codes (Cohen's Kappa) was high for both at .928 and .895, respectively. Overall, US participants generated more mental state content (83.102%) compared to Himba participants (71.296%), $F(1, 46)=7.116$, $p<.05$, $\eta_p^2=.134$, but this was qualified by an interaction between emotion portrayal type and culture, $F(8, 368)=5.375$, $p<.001$, $\eta_p^2=.105$ (Table S7). In contrast to the overall pattern of effects, Himba perceivers generated more mental state content than US perceivers in response to laughter vocalizations (poses of amusement). Further, several vocalization types (i.e., sensory pleasure, surprise and triumph) did not yield different amounts of mental state content across cultures. Interestingly, it was primarily the so-called "basic" emotion vocalizations that yielded more mental state content from US compared to Himba perceivers. This finding is consistent with the special status that emotions like anger, fear and sadness have in US culture.

In contrast to mental state content, action and situational content was generated more by Himba perceivers (68.750%) than by US perceivers (11.574%), $F(1, 46)=130.665$, $p<.001$, $\eta_p^2=.740$, but this was qualified by an interaction between emotion portrayal type and culture, $F(8, 368)=2.447$, $p<.05$, $\eta_p^2=.051$ (Table S8). Follow-up analyses within each culture revealed that Himba perceivers produce the same amount of action/situational content, regardless of the emotion portrayal type, $F(8, 184)=1.475$, $p=.169$, whereas US perceivers generate different amounts of action/situation content, depending on the emotion portrayal type, $F(8, 184)=4.113$, $p<.001$, $\eta_p^2=.152$. Bonferroni corrected pairwise comparisons revealed that US perceivers generate significantly less action/situational content in response to fear vocalizations compared to sensory pleasure vocalizations, $p<.05$, 2-tailed (other emotion portrayal means fell between these two extremes and did not differ significantly from one another).

Despite differences in the amount of content that US and Himba cultures produce, it is important to test whether the content produced was appropriately tailored to the stimulus. Action/situational content was then coded based on how well it agreed with the discrete emotion category portrayed in the vocalization. For example, if the vocalization "ewww" for disgust, was described as "someone is sick" this was coded as in agreement with the discrete emotion content. Simple action words such as "crying" could also be in agreement with the discrete emotion portrayed if they fit the US cultural assumptions about the relationship between behavior and emotion (e.g., crying is in agreement with sadness). Reliability of codes (Cohen's Kappa) for the two coders was high at .95. Discrepancies in coding were resolved by review and discussion among the coders and the first author in a way that was blind to culture. The results are presented in Table S9. Himba participants were more likely to choose action/situation labels that better described the perceptual information in the vocalizations than did American participants for vocalizations depicting amusement, anger, fear, sadness and sensory pleasure, *F*(8, 368)= 9.657, p<.001, η_p^2=.174. These data demonstrate that, for some of the vocalizations, Himba participants were able to perceive the vocalization and/or the situational context in which it might occur, even if they did not make mental state inferences about the vocalization as might be expected in a Western cultural context. While Himba participants do not perceive the intended emotion in the Western portrayals they were exposed to, some of the sounds were recognizable in terms of specific contexts in which they might occur, or based on a specific term for that type of vocalizations (e.g., "crying" for sadness).

Figure Legends

Figure S1. Cross-cultural comparisons for mean agreement between action and situation labels and the expected discrete emotion category. Himba and American participants' performance for contextual (situations or behaviors) response appropriateness given the discrete emotion posed by a given vocalization. Mean percentage (± SEM) of discrete emotion response agreement based on coding (by two independent raters) of the contextual content of responses only is presented on the y-axis. Data are presented by culture and discrete emotion portrayal type (x-axis).

Figure S2. Cross-cultural comparisons for mean agreement between action and situation labels and the valence portrayed. Himba and American participants' performance for contextual (situations or behaviors) response appropriateness given the valence (positive, negative and neutral) content posed by a given vocalization. Mean percentage (± SEM) of valence response agreement based on coding (by two independent raters) of the contextual content of responses only is presented on the y-axis. Data are presented by culture and discrete emotion portrayal type (x-axis).

Figure S3. Cross-cultural comparisons for mean agreement between action and situation labels and the arousal-level portrayed. Himba and American participants' performance for contextual (situations or behaviors) response appropriateness given the arousal/activation (high, mid, and low) content posed by a given vocalization. Mean percentage (± SEM) of arousal response agreement based on subjective coding (of two independent raters) of the contextual content of responses only is presented on the y-axis. Data are presented by culture and discrete emotion portrayal type (x-axis).

References

1. Ekman P (1972) Universals And Cultural Differences In Facial Expressions Of Emotions. *Nebraska symposium on motivation, 1971*, ed Cole J (University of Nebraska Press, Lincoln, Nebraska), pp 207-283.
2. Ekman P & Friesen W (1971) Constants across cultures in the face and emotion. *Journal of personality and social psychology* 17(2):124-129.
3. Ekman P, Heider E, Friesen WV, & Heider K (1972) Facial expression in a preliterate culture. Unpublished manuscript.
4. Sauter DA, Eisner F, Ekman P, & Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America* 107(6):2408-2412.
5. Sorensen ER (1975) Culture and the expression of emotion. *Psychological Anthropology*, ed Williams TR (Mouton Publishers, Paris), pp 361-372.
6. Rosenberg EL & Ekman P (1995) Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion* 19(2):111-138.

Supplementary Tables and Figures

Table S1. *Universality studies in remote cultural contexts*

	Study	Cultural Groups: Targets	Cultural Groups: Perceivers (n)	Conceptual Content	Affect Control	Response Format	Number of Response Options	Results (% agreement)
Studies in Support of Universality								
	Ekman & Friesen (1971)	U.S.	New Guinea Fore (319)	Y	Y	Select Face	2-3	71.8 %
	Ekman et al. (1972), unpublished; Partially described in Ekman (1972), Study 2	U.S.	New Guinea Dani (64)	Y	Y	Select Face	2-3	65.7 %
	Sauter et al., (2010)	English	Himba (58)	Y	N	Select Vocalization	2	64.5 %
Studies in Support of Relativity								
	Sorensen (1975)	U.S.	New Guinea Fore (n not reported)	N	n/a	Free label	n/a	25.78 %*
		U.S.	New Guinea Bahinemo (71)	N	n/a	Free label	n/a	"anger" response for all stimuli

Note. Adapted from Elfenbein & Ambady (2002). N = no; Y = yes; n/a = not applicable. "Conceptual Content" column indicates whether conceptual content for emotions (typically in the form of situation vignettes) was explicitly included within experimental trials. *Computed based on reported data (mean accuracy not weighted by sample sizes per condition because not reported).

Table S2. *Agreement between Freely-Produced and Expected Discrete Emotion Labels*

	Cultural Group			
	US		Himba	
Emotion Portrayed	Mean	SD	Mean	SD
Amusement	68.750***	35.547	79.167***	32.693
Anger	64.583***	42.934	2.083	10.206
Disgust	58.333***	38.069	.000	.000
Fear	54.167***	46.431	20.833**	35.864
Relief	66.667***	31.851	2.083	10.206
Sadness	58.333***	35.098	2.083	10.206
Sensory Pleasure	43.750***	39.871	4.167	20.412
Surprise	70.833***	32.693	.000	.000
Triumph	6.250	22.421	.000	.000

Note: Mean agreement (and standard deviation) between participants' free-labeling responses and the labels that were expected for each category of vocalizations (i.e., the Western emotion category that the vocalizations were intended to portray). Data are presented by cultural group and discrete emotion category. One-sample t-tests were conducted to examine whether perceivers agreement level was significantly greater than zero. ***p<.001, **p<.01

Table S3. *Agreement between Valence Portrayed and Freely-Produced Labels*

	Cultural Group			
	US		Himba	
Emotion Portrayed	Mean	SD	Mean	SD
Amusement	70.833***	29.180	97.917***	10.206
Anger	82.292***	23.865	75.000***	36.116
Disgust	89.583***	20.743	60.417***	36.053
Fear	83.333***	24.077	54.167***	35.864
Relief	67.708***	31.691	20.833***	25.181
Sadness	83.333***	31.851	58.333***	31.851
Sensory Pleasure	48.958***	39.342	20.833**	29.180
Surprise	76.042***	29.005	0.000	0.000
Triumph	72.917***	25.449	66.667***	31.851

Note: Mean agreement (and standard deviation) between participants' free-labeling responses and the labels that would be valence-congruent with each category of vocalization. Data are presented by cultural group and emotion category. One-sample t-tests were conducted to examine whether perceivers' agreement level was significantly greater than zero. ***p<.001, **p<.01

Table S4. *Agreement between Arousal Level Portrayed and Freely-Produced Labels*

	Cultural Group			
	US		Himba	
Emotion Portrayed	Mean	SD	Mean	SD
Amusement	70.833***	29.180	77.083***	32.900
Anger	73.958***	32.537	41.667***	38.069
Disgust	77.083***	32.900	31.250**	41.211
Fear	79.167***	35.864	37.500***	39.700
Relief	83.333***	31.851	10.417	25.449
Sadness	62.500***	33.783	14.583**	23.215
Sensory Pleasure	47.917***	40.323	14.583*	27.502
Surprise	79.167***	29.180	64.583***	40.323
Triumph	80.208***	26.559	41.667***	28.233

Note: Mean agreement (and standard deviation) between participants' free-labeling responses and the labels that would be arousal-congruent with each category of vocalization. Data are presented by cultural group and emotion category. One-sample t-tests were conducted to examine whether perceivers' agreement level was significantly greater than zero. ***p<.001, **p<.01, *p<.05

Table S5. *Overall Chance-Level One-Sample T-tests*

		Cultural Group			
		US		Himba	
Emotion Portrayed	Criterion	Mean	SD	Mean	SD
Anger	11.111	64.58***	42.93	2.08***†	10.21
Fear	11.111	54.17***	46.43	20.83	35.86
Disgust	11.111	58.33***	38.07	0	0
Sadness	11.111	58.33***	35.10	2.08***†	10.21
Amusement	11.111	68.75***	35.55	79.17***	32.69
Sensory Pleasure	11.111	43.75**	39.87	4.17	20.41
Relief	11.111	66.67***	31.85	2.08***†	10.21
Surprise	11.111	70.83***	32.69	0	0
Triumph	11.111	6.25	22.42	0	0

Note: Mean agreement (and standard deviation) between participants' responses and the portrayed discrete emotion category. Data are presented by emotion portrayal type (i.e., the Western emotion category that the vocalization was intended to portray) and cultural group. One-sample t-tests were conducted to test whether participant performance was above chance-level according to an overall criterion set based on the number of stimuli types included in the experiment. ****p*<.001, ***p*<.01, **p*<.05, † *t*-statistic is negative indicating mean is significantly less than criterion.

Table S6. *Within Valence and Arousal Chance-Level One-Sample T-tests*

		Cultural Group			
		US		Himba	
Emotion Portrayed	Criterion	Mean	SD	Mean	SD
Anger	33.33	64.58**	42.93	2.08***†	10.21
Fear	33.33	54.17**	46.43	20.83	35.86
Disgust	33.33	58.33*	38.07	0	0
Amusement	50	68.75*	35.55	79.17***	32.69
Sensory Pleasure	50	43.75	39.87	4.17***†	20.41

Note: Mean agreement (and standard deviation) between participants' responses and the portrayed discrete emotion category. Data are presented by emotion portrayal type (i.e., the Western emotion category that the vocalization was intended to portray) and cultural group. One-sample t-tests were conducted to test whether participant performance was above chance-level according to a criterion set based on the normative valence and arousal level associated with a given emotion in Western culture. ***$p<.001$, **$p<.01$, *$p<.05$, † *t*-statistic is negative indicating mean is significantly less than criterion.

Table S7. *Mean Percent Production of Mental State Labels*

	Cultural Group					
	US (N=24)		Himba (N=24)			
Emotion Portrayed	Mean	SD	Mean	SD	*t*	df
Amusement	77.083	29.4115	97.917	10.2062	3.278**	28.460†
Anger	91.667	19.0347	64.583	42.9336	-2.825**	31.705†
Disgust	87.500	22.1163	72.917	25.4489	-2.119*	45.123†
Fear	93.750	16.8916	62.500	36.8605	-3.776***	32.252†
Relief	89.583	25.4489	66.667	35.0982	-2.590*	41.947†
Sadness	83.333	28.2330	52.083	42.9336	-2.979**	39.758†
Sensory Pleasure	60.417	36.0530	68.750	32.3449	.843	45.469†
Surprise	85.417	27.5016	79.167	32.6931	-.717	44.690†
Triumph	79.167	25.1805	77.083	32.9003	-.246	43.062†

Note: Mean percentage (and standard deviation) of responses containing mental state content based on the portrayed discrete emotion type (i.e., the Western emotion category that the vocalization was intended to portray) and culture. Independent-sample t-tests were conducted to examine whether perceivers from US and Himba culture differed in the amount of mental state labels they produced. ***p<.001, **p<.01, *p<.05; †Equal variances not assumed.

Table S8. *Mean Percent Production of Action and Situational Descriptions*

	Cultural Group					
	US (N=24)		Himba (N=24)			
Emotion Portrayed	Mean	SD	Mean	SD	*t*	df
Amusement	18.750	35.5470	66.667	43.4057	4.184***	46
Anger	4.167	14.1165	68.750	24.7268	11.112***	36.553†
Disgust	8.333	19.0347	68.750	32.3449	7.886***	37.225†
Fear	2.083	10.2062	70.833	35.8641	9.033***	26.701†
Relief	8.333	24.0772	66.667	31.8511	7.157***	42.815†
Sadness	8.333	19.0347	85.417	23.2153	12.579***	44.299†
Sensory Pleasure	29.167	32.6931	70.833	35.8641	4.206***	46
Surprise	10.417	25.4489	56.250	42.5096	4.532***	37.610†
Triumph	14.583	23.2153	64.583	40.3225	5.265***	36.738†

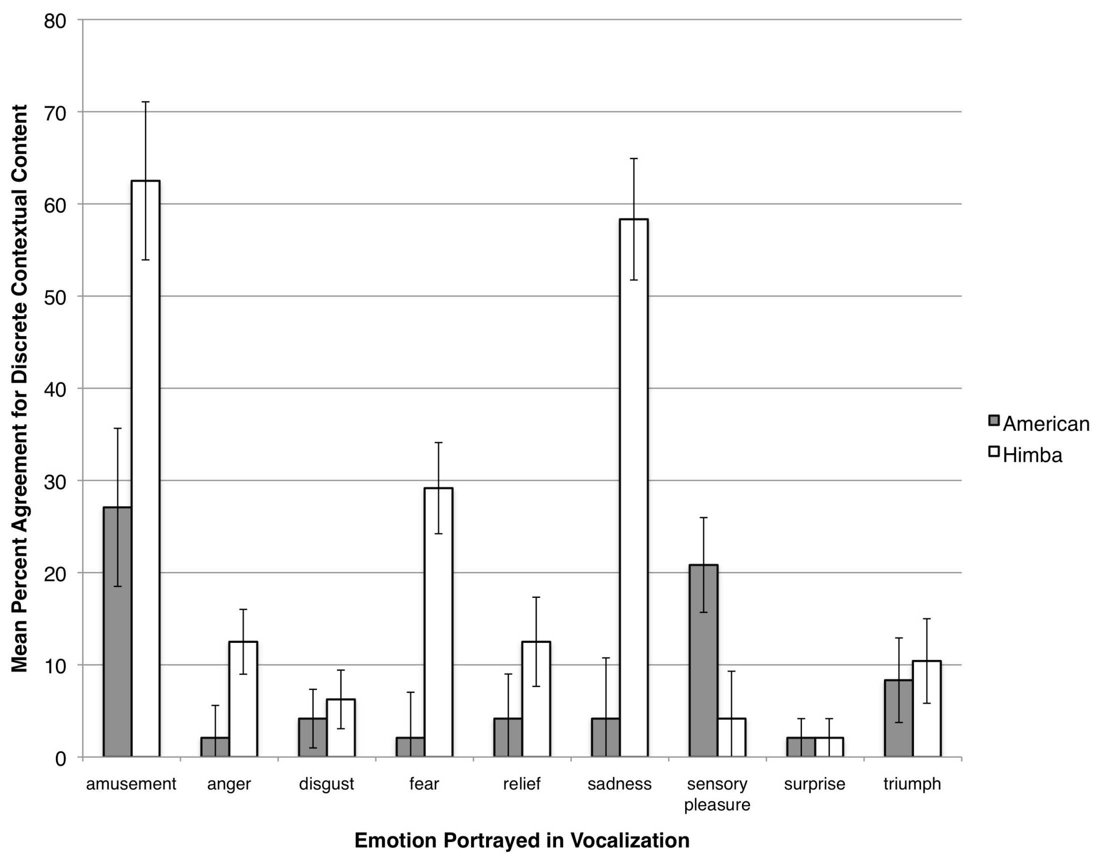
Note: Mean percentage (and standard deviation) of responses containing action and situational content based on the portrayed discrete emotion type (i.e., the Western emotion category that the vocalization was intended to portray) and culture. Independent-sample t-tests were conducted to examine whether perceivers from US and Himba culture differed in the amount of action/situational they produced. ****p*<.001; †Equal variances not assumed

Table S9. *Mean Agreement for Action and Situation Labels and the Expected Discrete Emotion Category*

	Cultural Group			
	US		Himba	
Emotion Portrayed	Mean	SD	Mean	SD
Amusement	25.000**	39.010	62.500***	44.843
Anger	2.083	10.206	12.500*	22.116
Disgust	4.167	14.117	6.250	16.892
Fear	2.083	10.206	29.167***	32.693
Relief	4.167	20.412	12.500*	26.581
Sadness	4.167	14.117	58.333***	43.406
Sensory Pleasure	18.75**	32.345	4.167	14.117
Surprise	2.083	10.206	2.083	10.206
Triumph	8.333*	19.035	10.417	25.449

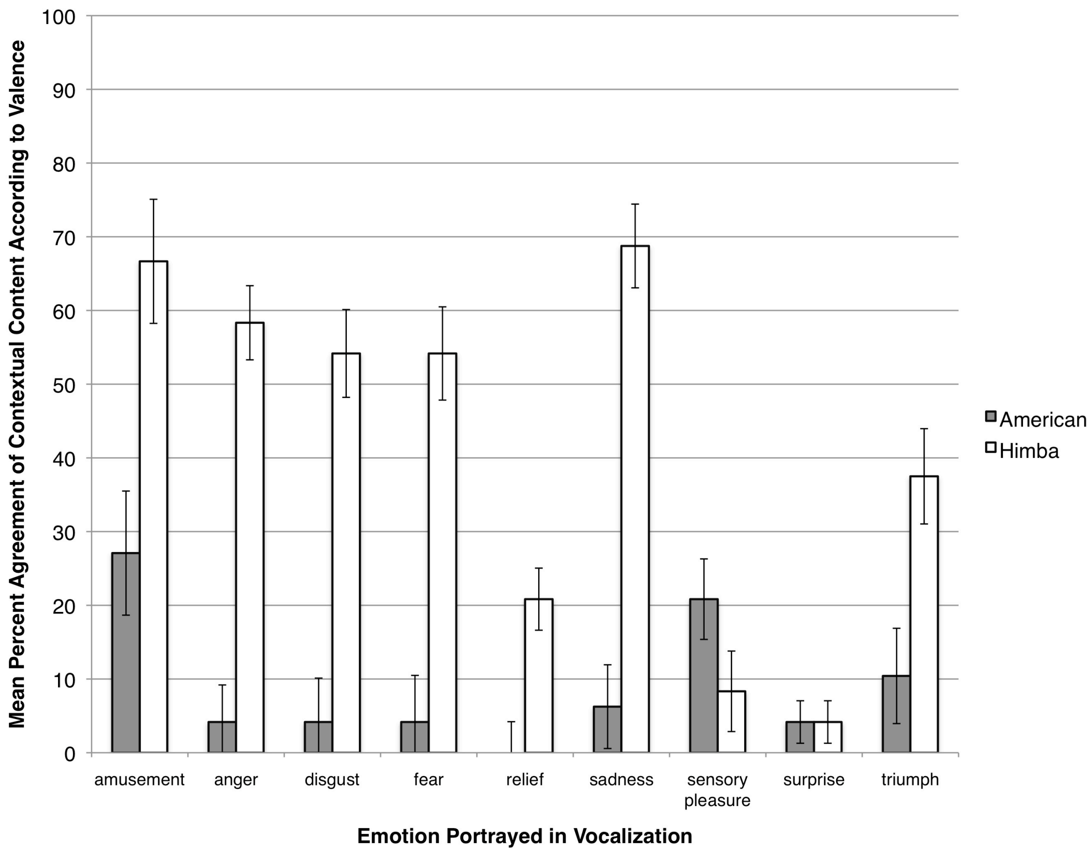
Note: Mean agreement (and standard deviation) between participant's contextual response content and the portrayed discrete emotion category. Data are presented by cultural group and emotion portrayal type (i.e., the Western emotion category that the vocalization was intended to portray). One-sample t-tests were conducted to examine whether perceivers performed at significantly greater than zero. ****p*<.001, ***p*<.01

Figure S1.



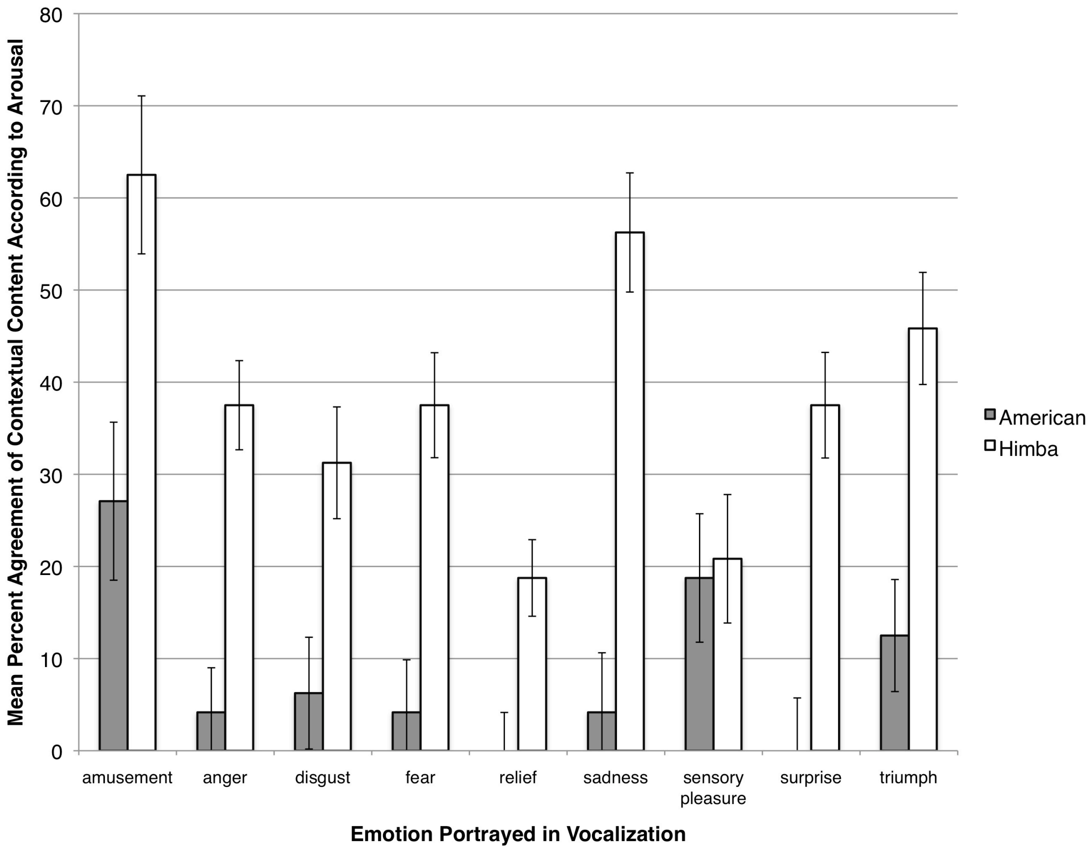
80
70
60
50
40
30
20
10
0
Mean Percent Agreement for Discrete Contextual Content
amusement
anger
disgust
fear
relief
sadness
sensory pleasure
surprise
triumph
Emotion Portrayed in Vocalization
American
Himba

Figure S2.



100
90
80
70
60
50
40
30
20
10
0
Mean Percent Agreement of Contextual Content According to Valence
amusement
anger
disgust
fear
relief
sadness
sensory pleasure
surprise
triumph
Emotion Portrayed in Vocalization
American
Himba

Figure S3.



80
70
60
50
40
30
20
10
0
Mean Percent Agreement of Contextual Content According to Arousal
amusement
anger
disgust
fear
relief
sadness
sensory pleasure
surprise
triumph
Emotion Portrayed in Vocalization
American
Himba

Chapter III

Cultural and Linguistic Relativity in Perceiving Facial Expressions for Emotion

A test of the universality of attribution hypothesis for facial cues. This study replicated and extended the universality work on facial expressions in order to examine whether facial muscle poses are cross-culturally perceived as discrete emotion mental states. Further, we manipulated the presence of emotion concepts in the task in order to assess the impact on discrete emotion perception.

Title: Cultural and Linguistic Relativity in Perceiving Facial Expressions for Emotion

Authors: Maria Gendron[1,2*], Debi Roberson[3], Jacoba Marietta van der Vyver[4], Lisa Feldman Barrett[2,5,6]

Affiliations:

1Department of Psychology, Boston College

2Affective Science Institute and Department of Psychology, Northeastern University

3Department of Psychology, University of Essex

4 School of Nursing and Public Health, University of Namibia

5 Martinos Center for Biomedical Imaging, Massachusetts General Hospital

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

* To whom correspondence should be addressed. E-mail: gendroma@bc.edu

Abstract

It is widely believed that certain emotions are universally recognized in facial expressions. Recent evidence indicates that Western perceptions (e.g., scowls as anger) depend on emotion concepts embedded in experiments. Since concepts are standard in cross-cultural experiments, we hypothesized that evidence of universality depends on this context. Participants from the US and the Himba ethnic group sorted images of facial expressions into piles by emotion. Himba participants did not show the "universal" pattern. To manipulate concepts, half of the participants were instructed with emotion words. Concepts impacted sorting in both cultures; participants produced sorts closer to the "universal" pattern, but substantial cultural variation was evident for Himba participants. Our findings indicate that perceptions of emotion are not universal, but depend on cultural and conceptual contexts.

One Sentence Summary: Emotional expressions are not universally recognized, but perceiving emotion in facial expressions depends on cultural context and momentarily accessible emotion concepts.

Since the 1970s, it has been widely accepted that all humans around the world (barring illness) reliably express particular emotional states (such as *anger*, *sadness*, *fear*, *disgust*, and *happiness*) with characteristic facial movements, called "facial expressions" (Fig. 1). It is further assumed that people all around the world automatically recognize emotions in these expressions (e.g., scowls are recognized as anger, pouts as sadness, and so on) in a way that is language and concept free (*1-3*). We refer to this as the *universality hypothesis*. The idea that emotions have universally recognized facial expressions has been absorbed into Western popular culture and undergraduate science curriculums. Each year the government spends on security training based, in part, on the assumption that it is possible to read a person's intent and predict his or her behavior from facial expressions, regardless of cultural background or life experiences (*4, 5*).

A strong test of the universality hypothesis is achieved by comparing emotion perceptions in individuals with limited exposure to Western style "expressions" to perceptions in individuals from Western cultural contexts (such as the US) (*6*). Of the hundreds of experiments cited in support of the universality hypothesis (*7*), only five provide a strong test because they sampled participants from groups culturally isolated from Western influence (*1, 8-11*) (see Table 1). Of those five, three found support for the universality hypothesis. In these three experiments, participants performed a task of matching posed expressions (usually static, unmoving faces) to conceptual content for emotions. For example, participants might be presented with an orally translated story, such as "He is looking at something which smells bad" (describing a disgust scenario) and asked to select the matching expression from two or three options (*8*). In the two experiments that did not ask participants to match a facial expression to preselected conceptual content, the universality hypothesis was not supported (*12*) (see Table 1). Evidence for the universality hypothesis does not come from people merely "recognizing" emotion in the perceptual features of facial expressions alone; instead, participants typically match faces to provided emotion concepts (either primed directly with a word or indirectly with a story) that represent a Western model for perceiving emotion.

One of the most efficient ways of making emotion concepts available to perceivers is to embed emotion words themselves into the task. In fact, there is emerging evidence from laboratory experiments, lesion studies, and brain imaging research that the presumed "universal" (i.e., Western) pattern of emotion perception necessarily depends on being able to process the emotion words that are routinely explicitly offered as part of the experimental design (*13, 14*). Experiments that do not offer US participants a small, preselected set of emotion words to match posed Western-style facial expressions do not find evidence that people are perceiving emotion according to the presumed universal pattern (*15*). Taken together, there is an emerging picture that challenges one of the most cherished facts in the science of psychology: the "universal" pattern of emotion perception is achieved not on the basis of the perceptual features of faces alone, but on how those features are understood in the conceptual context of the experimental task.

New evidence of cultural relativity rather than universality in emotion perception

To re-examine the universality hypothesis, we tested emotion perception in non-Western participants using an experimental procedure that is relatively freer from the conceptual content typically embedded in the task. We tested members of the Himba ethnic group who live in the remote villages within the Keunene region of Northwestern Namibia and have relatively little contact with people outside of their immediate communities (and therefore limited exposure to emotional expressions outside their own cultural context), making them good candidates for testing the universality hypothesis. The Himba speak a dialect of Otji-Herero that includes translations for the English words “anger”, “sadness”, “fear”, disgust”, “happy” and “neutral” (*10*). We presented 26 Himba adults (*16*) and 31 US adults with photos of six target individuals posing facial expressions of the type that are typical of emotion perception experiments (each smiling, scowling, pouting, nose wrinkling, wide eyed to depict happiness, anger, sadness, disgust, and fear, respectively), as well as each in a neutral expression (see Fig. 1). Participants sorted the 36 faces into piles, where targets in each pile were experiencing the same emotional state. Following their free sorting, participants labeled their piles. Labels were directly translated into English by an interpreter who worked on a prior study with Himba participants supporting the universality hypothesis (Table 1, Row 4; *10*). If scowling faces are universally perceived as angry, pouting faces as sad, smiling faces as happy, and so on, then participants in both cultures would freely sort the 36 stimulus faces into six piles (based on the perceptual regularities in the stimulus set, with all the scowling faces in one pile, all the pouting faces in another, and so on), and would label those piles using the expected emotion words.

A hierarchical cluster analysis of the free sort data, and subsequent examination of the six-cluster solution (corresponding to the presumptive “universal” solution), indicated that the Himba and US participants did not perceive the facial expressions similarly (Fig. 2, A, B; dendrograms in Fig. S1). As expected, the US participants grouped together smiling (happy), scowling (angry), wide-eyed (fearful) and neutral faces into distinct piles, although there was less clear discrimination of the pouting (sad) and nose-wrinkled (disgusted) faces. Himba participants, in contrast, (Fig. 2, B) grouped smiling (happy) faces together, wide-eyed (fearful) faces together, and neutral faces together, with no clear grouping of the scowling (angry), pouting (sad), and nose-wrinkled (disgusted) faces. The words that participants freely offered to name their piles confirmed our observation that Himba participants did not perceive the facial expressions according to the "universal" pattern (Fig. 3). As expected, US participants tended to use discrete emotion words such as "sadness" or "disgust" to label their piles (M = 4.45, SD = 1.091) more than did Himba participants (M = .81, SD = .939), t (55) = 2.644, $P < .001$. Interestingly, US participants more generally named their piles with a wide array of additional mental state words such as "surprise" or "concern" (M = 2.32, SD = 1.447) at greater frequency than did Himba participants (M = .69, SD = .838), t (55) = 1.630, $P < .001$. In contrast, Himba participants were more likely to label face piles with descriptions of physical actions such as "laughing" or "looking at something" (M = 2.38, SD = 1.098) when compared to US participants (M = .84, SD = 1.098), t (55) = -1.546, $P < .001$, suggesting that individuals in Himba culture used *behavioral descriptors* more frequently than *mental state descriptors* to understand the meaning of facial actions. Indeed, extensive research on Action Identification Theory (*17*) indicates that it is possible to describe another person's actions in either physical or mental terms (*18*). For example, perceiving a face as "fearful" requires mental state inference because an internal state is assumed to be responsible for the observable "expression"; this stands in contrast to perceiving actions on the face in more physical (and less psychological) terms (e.g., looking) that might be observed in any number of emotional or non-emotional instances (e.g., see Table S1).

Multidimensional scaling (MDS) analyses of the free sort data (Fig. 4) confirmed that Himba participants possessed a cognitive map of the facial expressions that was anchored in action identification, whereas the US participants' cognitive map also contained information about mental state inferences of discrete emotions. MDS analyses produce a cognitive map of the sorting pattern for each group of participants, along with the underlying dimensions that best represent how participants perceived the similarities and differences among the face stimuli (see

SOM for more detail; *19*). Rather than subjectively interpreting the dimension weights (as most experiments do), we opted for an empirical approach using hierarchical regression analysis (as in *20*). Following Action Identification Theory, we first examined the extent to which the MDS dimensions reflected "behavioral" information in the faces (see Table S2), followed by the extent to which the dimensions reflected "mental state" inferences supported by the faces (see Table S3) over and above behavior (see SI for details). We predicted, and found, that the MDS dimensions for the Himba solution would reflect behavioral distinctions, whereas the dimensions for the US solution would reflect mental state distinctions (over and above behavior; see Table 2, Tables S4 & S5 for F-tests).

Taken together, comparing Himba and US free sort data indicated strong cultural differences in emotion perception. It is unlikely that these findings were simply due to "poor" task-related performance in Himba participants, who have performed well on other psychological testing with significant demands (*21*). Furthermore, the R2 values for identification of the MDS dimensions were large for Himba participants' data, indicating that there was robust and meaningful group-level consistency in how Himba participants sorted the face stimuli.

Emotion words impact emotion perception in a culturally relative manner

To explicitly examine the effect of emotion words on emotion perception, 28 additional Himba participants (i.e., separate participants drawn from same two villages where we ran the free sort task) and 36 additional US participants sorted the 36 stimulus faces into piles but were provided with verbal anchors of the six relevant emotion words ("anger", "fear", "disgust", "sadness", "happy", and "neutral", or their translation in the Himba dialect of Herero as "okupindika", "okutira", "okujaukwa", "oruhoze", "ohange" and "nguri nawa", respectively). We hypothesized that if these emotion words encourage a certain pattern of emotion perception, as suggested by recent research (*15, 22, 23*), then both Himba and US participants would differ in their sorting compared to the free sort samples where the emotion words were not introduced. Specifically, we expected that US anchored sort responses would better resemble the expected "universal" solution when compared to US free sort responses. Furthermore, if emotion concept words are linked to universal perceptual representations (e.g., furrowed brow and pressed lips in *anger*) across distinct cultural contexts, then Himba participants would also produce a more "universal" (i.e., Western) looking solution in the anchored sort procedure. If the perceptual representations linked to emotion concept words vary across distinct cultural contexts (as recent evidence suggests, *24*), however, then Himba participants' sorting of facial expressions might differ with the introduction of emotion words, but would not be clearly in line with the presumed universal (i.e., Western) cultural model.

As predicted, the introduction of emotion word anchors prior to and during sorting influenced both US and Himba participants' perceptions of the facial expressions. Participants from the US (and to a limited extent, Himba participants) conformed more to the "universal" emotion perception solution. The hierarchical cluster analysis produced a much clearer "universal" pattern for the US participants in the conceptually anchored condition, with nose-wrinkled (disgust) and pouting (sadness) expressions in separate clusters (Fig. 2C, clusters 4 and 6, respectively) compared to the free sorting solution where those expressions ended up in the same cluster (Fig. 2A, cluster 5). For Himba participants, conceptually anchored sorting did not yield a dramatically more universal cluster solution. Nonetheless, the MDS analyses indicated that the emotions words did influence the sorting behavior of Himba participants in more subtle ways. Specifically, we found that facial expressions depicting emotion categories were more tightly clustered together in multidimensional space for Himba participants exposed to emotion word anchors compared to Himba participants who freely sorted the faces into piles (Fig. 5A, C); this increase in clustering was most apparent for smiling and neutral faces. For US participants (Fig. 5B, C), there was little change in the distances between within-category facial expressions.

As in the free sorting data, the MDS dimensions anchoring the US cognitive map were identified in emotion terms, reflecting that US participants were making mental state inferences about the facial expressions (see Table 2, Tables S10 & S11 for F-tests). The Himba cognitive map for the anchored sort data was described more in terms of mental states when compared to the map representing the freely sorted faces. One of the three MDS dimensions was empirically identified in emotion terms, and two remained in behavioral terms (although, one of these dimensions was trending for emotion; $P = .064$). Taken together, these data indicate that emotion words have a powerful effect on emotion perception, even under relatively unconstrained task conditions.

Discussion

The Western cultural script for what it means to be human includes the ability to "recognize" emotional expressions in a universal way. Yet, as we have shown, small changes in experimental procedure disrupt evidence for universal emotion perception. By comparing emotion perception in participants from maximally distinct cultural backgrounds, US participants and Himba participants from remote regions of Namibia with limited exposure to Western culture, we demonstrated that facial expressions are not universally "recognized" in emotional terms. Unlike prior experiments, we used a face-sorting task that allowed us to manipulate the influence of emotion concepts on how the faces were perceived. Without emotion concepts to structure perception, Himba individuals perceived the facial expressions as behaviors that do not have a necessary relationship to emotions, whereas US participants were more likely to perceive the expressions in mental terms, as the presumed universal emotion categories. When words for emotion concepts were introduced into the perception task, participants from both cultures grouped the facial expressions in way that better resembled the presumed "universal" solution, although this solution was more in evidence for US participants than for Himba participants. For Himba participants, the difference between free and conceptually anchored solutions was more subtle, and consistent with recent evidence demonstrating that people from different cultures vary in their internal, mental representations of emotional expressions (*25*).

Although we report only one study in this paper, our findings do not stand alone. They are consistent with a growing body of evidence that emotions are not "recognized" but are perceived via a complex set of processes (*14*) that involve the interplay of different brain networks, such as those for action identification and mental state inference (*26, 27*). Our finding that the "universal" pattern of emotion perception appears to be linguistically relative is consistent with the pattern of published results (see Table 1), as well as our own laboratory research showing that experimentally decreasing the accessibility of emotion words' semantic meaning, using a procedure called "semantic satiation"(*28*), reduces the accuracy with which produce the presumed universal pattern of emotion perception (*23*) because words help to shape the underlying perceptual representation of those faces (*22*). Our current findings are consistent with research on patients with semantic deficits due to progressive neurodegeneration (i.e., semantic dementia) or brain injury (i.e., semantic aphasia), who do not perceive emotions in scowls, pouts, smiles and so on (*29, 30*). Even research in young children points to the importance of emotion words in emotion perception, because the universal pattern of emotion perception emerges in young children as they acquire conceptual categories for emotions that are anchored by emotion words (*31*). Taken together, these findings challenge the assumption that facial expressions are evolved, universal “signals” of emotion, and instead suggest that facial expressions are culturally-sensitive “signs” (*14, 32*). To properly understand how humans perceive emotion, we must attempt to move past the Western cultural model as the assumed model for people everywhere.

The present findings are not readily explained by a universality account of emotion recognition, even those that admit some minor cultural variability. It has been suggested, for example, that in addition to universal recognition abilities cultures have “display rules” (*1*) that allow people to regulate their expressions. Cultures are also thought have “decoding rules” (*33*) that govern how people report on their perceptions of emotion to maintain a culturally appropriate response (e.g., in Japanese culture, individuals will discount the intensity of emotion perceived and report that another person is feeling less intensely). Neither display nor decoding rule accounts have the predictive power in explaining the language-based effects we observed in the present study, or in any of the publishing findings demonstrating that, in US samples, the presumed universal solution can be disrupted by interfering with emotion word processing. Display and decoding rules, as instantiations of culture, are relatively rigid and would not predict that emotion perceptions vary based on momentarily accessible concepts.

Our own study is not without limitations. Although our stimuli were posed by individuals of African origin, the expressions may not be isomorphic with the expressions that Himba individuals typically make in everyday life. Future work should explore this possibility. Furthermore, additional research on Himba lexical categories is needed. It is possible that the Herero translation for emotion words used in our study are recently borrowed, as appears to be the case for the color term "burou", borrowed from the German '*blau'* also meaning blue. This might explain why Himba participants infrequently used discrete emotion words to label their piles, as well as the reduced potency of words to shape the perception of facial expressions. To that end, it would also be important to know the actual frequency with which Himba speakers use emotion words in everyday discourse (which could speak to how functional those words are for shaping perception), and the extent to Himba individuals engage in action identification vs mental state inference more broadly. Our results suggest that Himba speakers might effectively use behavior categories for making predictions about future actions and their consequences, whereas US individuals rely on mental state inferences for those purposes.

References

1. P. Ekman, in *Nebraska symposium on motivation, 1971,* J. Cole, Ed. (University of Nebraska Press, Lincoln, Nebraska, 1972), pp. 207-283.
2. C. E. Izard, *The face of emotion*. Century psychology series (Appleton-Century-Crofts, New York, 1971), pp. xii, 468 p.
3. S. S. Tomkins, *Affect, Imagery, Consciousness*. (Springer, New York, 1963), vol. 2.
4. C. R. Hontz, Hartwig, M., Kleinman, S. M. & Meissner, C. A, "Credibility Assessment at Portals, Portals Committee Report" (2009).
5. S. Weinberger, Airport Security: Intent to deceive? *Nature* **465**, 412 (2010).
6. A. Norenzayan, S. J. Heine, Psychological universals: what are they and how can we know? *Psychological bulletin* **131**, 763 (Sep, 2005).
7. H. A. Elfenbein, N. Ambady, On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin* **128**, 203 (Mar, 2002).
8. P. Ekman, W. V. Friesen, Constants across cultures in the face and emotion. *Journal of personality and social psychology* **17**, 124 (Feb, 1971).
9. P. Ekman, E. Heider, W. V. Friesen, K. Heider. (1972).
10. D. A. Sauter, F. Eisner, P. Ekman, S. K. Scott, Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 2408 (Feb, 2010).
11. E. R. Sorensen, in *Psychological Anthropology,* T. R. Williams, Ed. (Mouton Publishers, Paris, 1975), pp. 361-372.
12. Individuals from the Bahinemo tribe in Papau New Guinea (perhaps the most remote of groups tested) free-labeled all facial expressions (Western poses of happiness, sadness, disgust and so on) as "angry."
13. L. F. Barrett, K. A. Lindquist, M. Gendron, Language as context for the perception of emotion. *Trends in cognitive sciences* **11**, 327 (Aug, 2007).
14. L. F. Barrett, B. Mesquita, M. Gendron, Context in emotion perception. *Current Directions in Psychological Science* **20**, 286 (2011).
15. S. C. Widen, A. M. Christy, K. Hewett, J. A. Russell, Do proposed facial expressions of contempt, shame, embarrassment, and compassion communicate the predicted emotion? *Cognition & emotion* **25**, 898 (Aug, 2011).
16. Participants in our study were from different villages than those tested in prior research on emotion perception with the Himba.
17. R. R. Vallacher, D. M. Wegner, What Do People Think Theyre Doing - Action Identification and Human-Behavior. *Psychol Rev* **94**, 3 (Jan, 1987).
18. M. N. Kozak, A. A. Marsh, D. M. Wegner, What do I think you're doing? Action identification and mind attribution. *Journal of personality and social psychology* **90**, 543 (Apr, 2006).
19. J. B. Kruskal, M. Wish, *Multidimensional scaling*. A Sage university paper: Quantitative applications in the social sciences no 07-011 (Sage Publications, Beverly Hills, Calif., 1978), pp. 93 p.
20. L. Feldman, Valence focus and arousal focus: individual differences in the structure of affective experience. *Journal of Personality & Social Psychology* **69**, 153 (1995).
21. J. W. de Fockert, S. Caparos, K. J. Linnell, J. Davidoff, Reduced Distractibility in a Remote Culture. *PloS one* **6**, e26337 (2011).

22. M. Gendron, K. A. Lindquist, L. Barsalou, L. F. Barrett, Emotion words shape emotion percepts. *Emotion* **12**, 314 (Apr, 2012).
23. K. A. Lindquist, L. F. Barrett, E. Bliss-Moreau, J. A. Russell, Language and the perception of emotion. *Emotion* **6**, 125 (Feb, 2006).
24. R. E. Jack, O. G. B. Garrod, H. Yu, R. Caldara, P. G. Schyns, Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 7241 (May 8, 2012).
25. R. E. Jack, O. G. Garrod, H. Yu, R. Caldara, P. G. Schyns, Facial expressions of emotion are not culturally universal. *Proc Natl Acad Sci U S A* **109**, 7241 (May 8, 2012).
26. R. P. Spunt, M. D. Lieberman, An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage* **59**, 3050 (Feb 1, 2012).
27. J. Zaki, K. Hennigan, J. Weber, K. N. Ochsner, Social cognitive conflict resolution: contributions of domain-general and domain-specific neural systems. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **30**, 8481 (Jun 23, 2010).
28. X. Tian, D. E. Huber, Testing an associative account of semantic satiation. *Cognitive psychology* **60**, 267 (Jun, 2010).
29. M. Calabria, M. Cotelli, M. Adenzato, O. Zanetti, C. Miniussi, Empathy and emotion recognition in semantic dementia: a case report. *Brain and cognition* **70**, 247 (Aug, 2009).
30. D. Roberson, J. Davidoff, N. Braisby, Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition* **71**, 1 (May 3, 1999).
31. S. C. Widen, J. A. Russell, Differentiation in preschooler's categories of emotion. *Emotion* **10**, 651 (Oct, 2010).
32. L. F. Barrett, Emotions are real. *Emotion* **12**, 413 (Jun, 2012).
33. D. Matsumoto, P. Ekman, American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion* **13**, 143 (1989).
34. R. C. Gur *et al.*, A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of neuroscience methods* **115**, 137 (Apr 15, 2002).
35. R. R. Sokal, C. D. Michener, A statistical method for evaluating systematic relationships. *The University of Kansas Scientific Bulletin* **38**, 1409 (1958).
36. J. A. Hartigan, M. A. Wong, Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **28**, 100 (1979).
37. M. S. Aldenderfer, R. K. Blashfield, *Cluster analysis*. Sage university papers series Quantitative applications in the social sciences (Sage Publications, Newbury Park, 1984), pp. 88 p.
38. A. P. M. Coxon, P. Davies, *The user's guide to multidimensional scaling : with special reference to the MDS(X) library of computer programs*. (Heinemann Educational Books, Exeter, NH, 1982), pp. xii, 271 p.
39. F. W. Young, R. Lewyckyj, in *Data Analysis and Theory Associates, 3rd. Ed.* (1979).
40. M. Buhrmester, T. Kwang, S. D. Gosling, Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspect Psychol Sci* **6**, 3 (Jan, 2011).
41. H. Aviezer, R. R. Hassin, S. Bentin, Y. Trope, in *First Impressions,* N. Ambady, J. J. Skowronsky, Eds. (The Guilford Press, New York, 2008).

Acknowledgments: Preparation of this manuscript was supported by the NIH Director's Pioneer Award (DP1OD003312) to LFB. We acknowledge the translation work of Kemuu Jakara. We thank J. Vlassakis, A. Matos, J. Fischer, S. Ganley, P. Aidasani, O. Bradeen and E. Muller for their help with data collection, entry or coding. We also thank Serge Caparos and Jules Davidoff for use of their supplies and support regarding data collection in Namibia. The data reported in the paper are presented in the Supplementary Materials and can be supplied in database form from the corresponding author (MG) by request. MG, DR, JMV & LFB contributed to design, MG, DR & JMV contributed to data collection, MG & LFB analyzed the data, MG, DR & LFB contributed to writing.

Table 1. Summary of Prior Universality Studies that Tested Individuals from Remote Cultural Contexts.

Study	Cultural Groups: Targets	Cultural Groups: Perceivers (n)	Conceptual Content	Response Format	Number of Response Options	Results (% accuracy)
Ekman & Friesen (1971)	U.S.	New Guinea Fore (319)	Y	Select Face	2-3	71.8 %
Ekman et al. (1972), unpublished	U.S.	New Guinea Dani (64)	Y	Select Face	2-3	65.7 %
Sauter et al. (2010)	English Speakers	Himba (58)	Y	Select Vocalization	2	64.5%
Sorensen (1975)	U.S.	New Guinea Fore (n not reported)	N	Free label Face	n/a	25.78%*
	U.S.	New Guinea Bahinemo (71)	N	Free label Face	n/a	Labeled all faces "angry"

Note. Adapted from Elfenbein & Ambady (2002). N = no; Y = yes; n/a = not applicable. "Conceptual Content" column indicates whether conceptual content for emotions (typically in the form of situation vignettes) was explicitly included within experimental trials. *Computed based on reported data (mean accuracy not weighted by sample sizes per condition because not reported).

Table 2. Multidimensional Scaling (MDS) Dimensions Identified using Hierarchical Multiple Regressions

	Dimension 1		Dimension 2		Dimension 3		Dimension 4	
Sort Solution	Behavior	Emotion	Behavior	Emotion	Behavior	Emotion	Behavior	Emotion
Himba Free	**.826*****	.064	**.732*****	.060	**.597*****	.122		
US Free	**.938*****	.021	**.885*****	**.078*****	**.913*****	.028	**.776*****	**.105***
Himba Anchored	**.865*****	.055	**.803*****	**.100***	**.704*****	.091		
US Anchored	**.854*****	**.098*****	**.926*****	**.032***	**.786*****	**.137****	**.851*****	**.080****

Note. The first block of regressors was comprised of eight sets of independent ratings where the 36 face stimuli were rated for the extent to which they depicted specific behaviors (e.g., smelling, crying; see Tables S1, S2; SOM for details); R^2 values are presented in the Behavior column under each dimension. The second block of regressors was comprised of six sets of ratings where each of the 36 faces was rated for the extent to which they represented emotions (e.g., disgust, sadness); the change in R^2 due to these ratings is presented under each dimension labeled Emotion. The change in R^2 reflects the extent to which the dimension reflects mental state inferences about emotion over and above the variance accounted for by mere action identification. Himba sorting data was captured by a lower dimensionality solution compared to US sorting data (3 dimensions instead of 4). As a result, a fourth dimension was only identified for US participants' data. Significant effects are bolded; * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$.

Fig. 1. Faces posed in typical Western-style expressions for emotions.

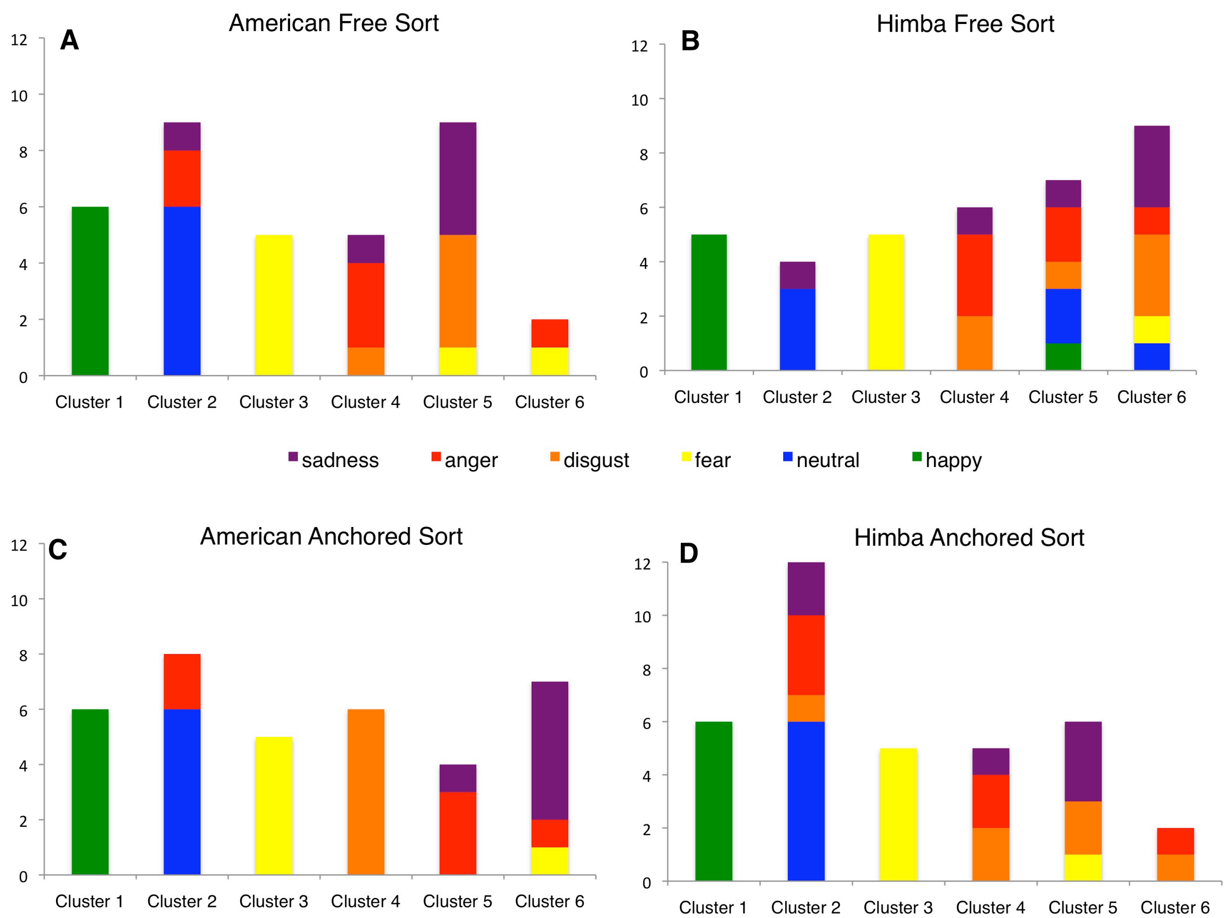


Fig. 2. The results of the six-cluster solutions from the hierarchical cluster analyses. Data are plotted in A-D with cluster on the x-axis. The y-axis represents the number of items grouped into a given cluster, with contents stacked by the emotion portrayed in each posed facial expression. Stacked bars containing several different colors indicate that faces portraying different discrete emotions were clustered together. Bars with a single color (or predominance of a given color) indicate relatively clean clustering of faces depicting one emotion category. The US free sort (**A**) cluster solution contains relatively "universal" clustering with the exception of cluster five, which appears to contain portrayals of both disgust and sadness. The Himba free sort (**B**) has three clear clusters (one through three). Both US (**C**) and Himba (**D**) conceptually anchored sorting appears to yield relatively distinct cluster solutions compared to free sorting.

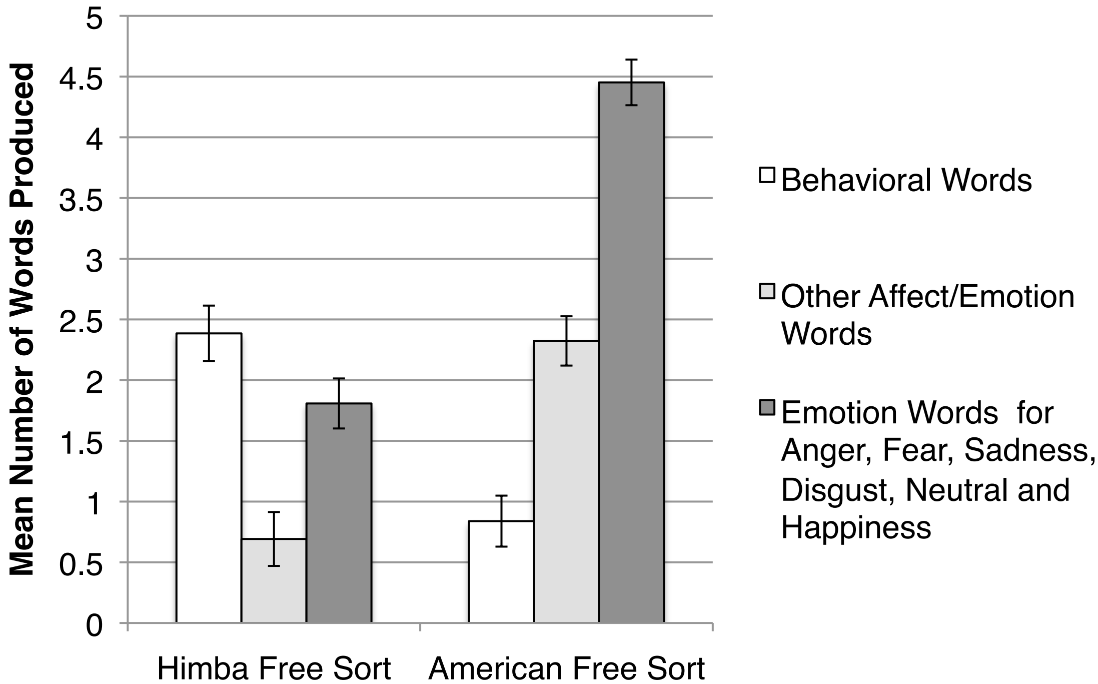


Fig. 3. Pile labels produced by Himba and US participants. Mean number of words produced by each group (± standard error) is plotted on the y-axis broken down by word type. We observed cross-cultural differences in label use when participants were asked to freely sort facial expressions, ($F_{1,55}$=24.952, P<.001, η_p^2 =.312), and this effect was qualified by the type of label produced, ($F_{2,\,110}$=56.719, P<.001, η_p^2 =.508).

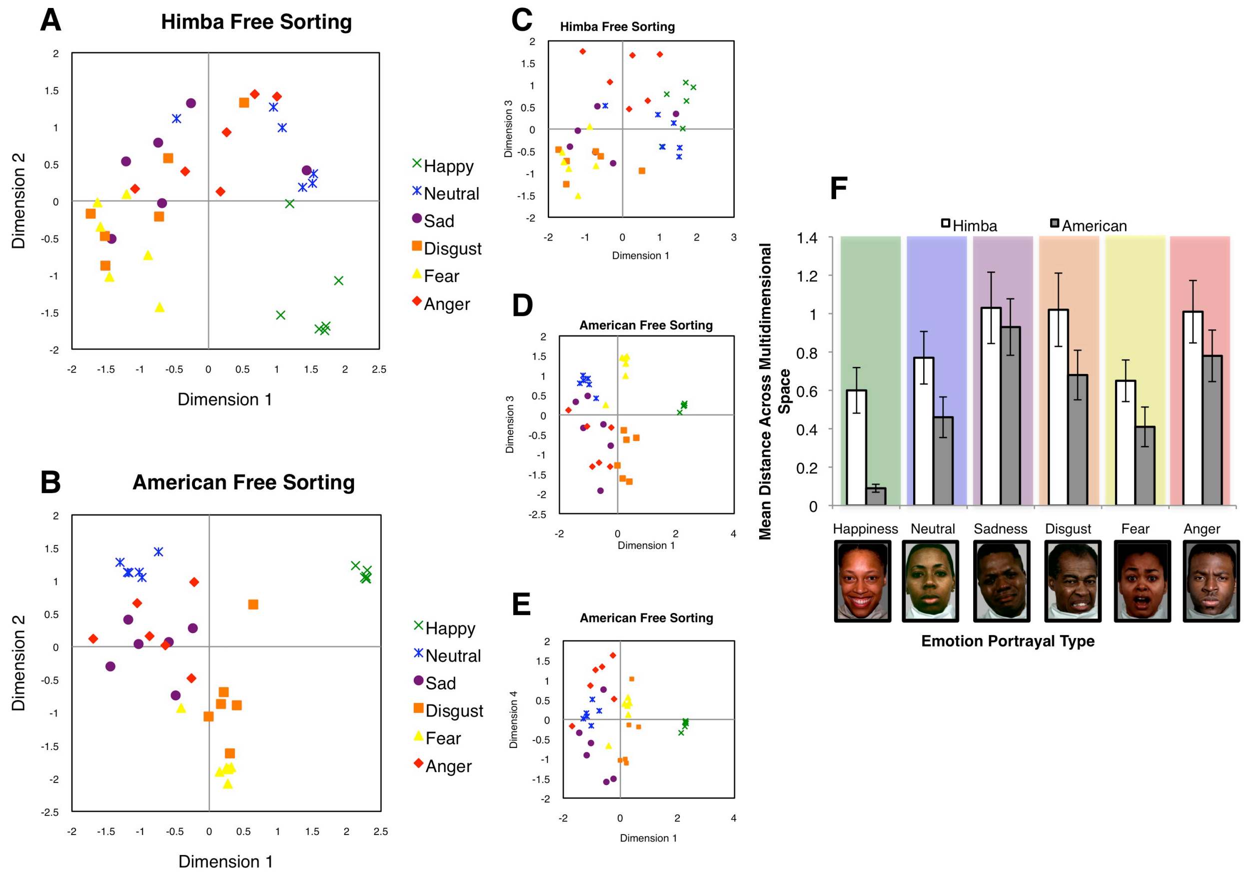


Fig. 4. Multidimensional scaling (MDS) solutions for free sorting of facial expressions. Data are plotted in two-dimensional space for Himba (**A**) and US (**B**) participants. Items are plotted by emotion type. Clearer evidence of the "universal solution" (closer clustering of facial expressions within the same emotion category) is more evident in the US solution than in the Himba solution. Additional dimensions are plotted for Himba (**C**) and US participants (**D**, **E**). For Himba participants, scowls ("anger" faces) were perceived differently than other negative expressions (**C**), but are not themselves tightly clustered together in multidimensional space. For US participants, wide-eyed ("fear") and wrinkle-nosed ("disgust") expressions are more clearly separated (**D**) as well as pouts ("sadness") and scowls ("anger") (**E**) in these additional plots. We

quantified the clustering of items within a category across all dimensions and plotted these mean distances (+/- standard error) for each cultural group (**F**).

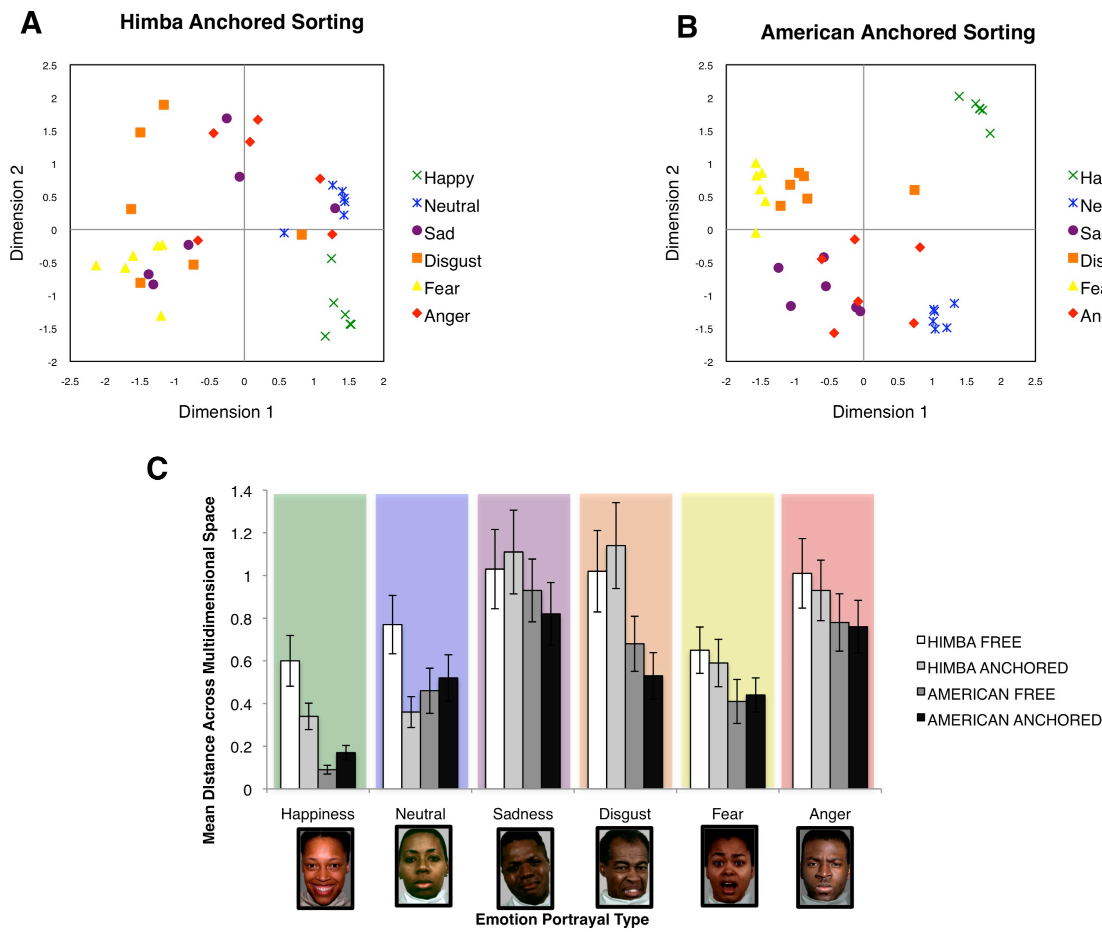


Fig. 5. MDS solutions for conceptually anchored sorting of facial expressions. Data are plotted in two-dimensional space for Himba (**A**) and US (**B**) participants. Interface distances (**C**) were quantified across all dimensions and plotted for each cultural group and each task version (anchored or free).

Supplementary Materials:
Materials and Methods
Figures S1-S3
Tables S1-S9
References (*34-41*)

Supplementary Materials:

Materials and Methods:

Participants. Participants were 65 individuals (32 male, 33 female; mean age= 30.84, SD=13.04) from the Himba ethnic group of Otji-Herero speakers. 11 Himba participants were excluded from analysis (resulting in a final sample of 54 participants) based on compliance (i.e., 4 participants decided not to finish sorting the stimuli), failure to sort into piles by emotion (2 participants) and feasibility (3 participants—one with low visual acuity, two with an inability to sort the stimuli at all and low forward span in a separate experiment). In addition, 2 participants' data were dropped from further analysis due to data recording issues that were discovered during data entry (i.e., not all image placements were recorded).

Individuals from the Himba ethnic group constitute a strong sample for the two-culture approach (*6*) to test the universality of emotion perception. Most of the Himba ethnic group lives in an ancestrally "traditional" culture, such that they do not take part in their country's political or economic systems, but rather live as semi-nomadic pastorialists, tending to herds of goats and cattle in remote regions of the Kaokoland (or Kunene region). Studying emotion perception in the Himba is additionally important given a recent paper concluding that Himba and British individuals have universal emotion perception in non-word human vocalizations (*10*). In our study, Himba participant data was collected in two remote villages, both located in the Mountainous Northwest region of Namibia, Kunene. The first village contained a mobile school as well as a missionary church; all members of the village otherwise lived traditionally and did not show significant signs of Westernization. The second village did not contain an established school or outside presence in the community. Both locations were far from the regional towns, and there was no evidence of tourism to these communities as is typical of "show" villages closer to the main regional town of Opuwo. While several of the participants (particularly in location 1) had some conversational English (greetings), none of the participants were English language speakers. All participants in this sample were native speakers of the Otji-Herero language.

Our US participants were 68 individuals (30 male, 38, female; mean age=38.27; sd=12.77) from an American community sample tested at the Boston Museum of Science in the Living Laboratory environment (http://www.mos.org/discoverycenter/livinglab). One participant's data were dropped from further analysis due to data recording issues that were discovered during data entry (i.e., not all image placements were recorded). The Living Laboratory is a well-suited site for the collection of control data given the hectic environment that mirrors the busy and often social context of testing out in the field. In addition, the Living Laboratory affords the collection of data from a community sample, such that individuals from a range of ages and backgrounds can be included.

Stimuli. Stimuli were 4x6 cards containing photographs of facial expressions of emotion posed by African American individuals. Stimuli were selected based on the following criteria. (1) The facial structure of the identities included in the stimulus set were rated as closest to that of several example with-in gender identities from the Himba ethnic group (as judged by a set of 120 perceivers from a largely Western cultural context using Amazon's Mechanical Turk system); the final set of face stimuli (three male and three female identities) were selected from a larger set of 23 male and 20 female identities from Gur (*34*), IASLab (www.affective-science.org), and NimStim (http://www.macbrain.org/resources.htm) face sets. (2) A given identity posed a facial expression depicting anger, fear, sadness, disgust, neutral affect, and happiness, with a final stimulus set of 36 faces. Each face stimulus was edited in Adobe Photoshop such that visual background information was removed. A uniform white collar (as is found on NimStim identities) was added to all photographs in order to make the stimuli uniform and remove variation that might distract from facial actions. (3) Additional norming conducted on Amazon's Mechanical Turk confirmed that the face stimuli were indeed rated as depicting the expected emotion (Table S2).

Procedure. All participants were consented prior to performing experiment. Participants from the Himba ethnic group were verbally consented with the use of a translator. Participants within each culture were randomly assigned to either the free sorting or anchored sorting condition. In the free sorting condition, participants were instructed as follows (through a translator for Himba participants): "I am going to hand you a pile of cards. The people on these cards are feeling some different types of emotions. I want you to put people who feel the same way into a pile together. So, if you see two people who both feel the same emotion, they should end up together in the same pile. People who feel different emotions should be put in different piles. You can create as many piles as you need to. In the end, everyone in a pile should feel exactly the same way. You can look at all of the pictures first, before you start sorting, if you want to see the different types of emotions that the people are feeling. You can always move people into a different pile if you change your mind. This is not timed, so you can take as long as you need. Do you have any questions?" The participant was then allowed to freely sort the images into piles. Participants were reinstructed if initial sorting appeared to be based on identity, rather than emotion (this took place in less than 25% of Himba participants). All but 2 participants were able to sort by emotion following reinstruction.

Participants in the word-anchored sort condition heard slightly modified instructions (again, via the translator for Himba participants). Differences from the unanchored instructions are italicized: "I am going to hand you a pile of cards. *We think you will find that people on these* *cards are feeling one of six different emotions. Those emotions are: anger, fear, sadness, disgust, neutral and happy*. I want you to put people who feel the same way into a pile together. So, if you see two people who both feel the same thing, they should end up together in the same pile. People who feel different emotions should be put in different piles. You can create as many piles as you need to. In the end, everyone in a pile should feel exactly the same way. You can look at all of the pictures first, before you start sorting, if you want to see the different types of emotions that the people are feeling. You can always move people into a different pile if you change your mind. Do you have any questions? *We are going to remind you of the six emotions as you sort the faces. When we remind you, it doesn't mean that you are doing well or doing poorly at sorting the faces. Instead we just give a reminder every so often. If you want a reminder of the six emotions at any point, just tell us and we will give you the words again. I'm going to give you the emotions to look for again and then you can start sorting. Anger, fear, sadness, disgust, neutral and happiness.*" Anchors were delivered verbally to all participants, rather than written, because the Himba participants we tested are from a pre-literate culture.

To ensure that the emotion concepts remained accessible throughout the task, participants were reminded of the six emotions every 6 cards that they placed down. If the participant initially looked through the stimuli before beginning to sort, they were also reminded every six stimuli that were closely examined by the participant. In total, each participant was reminded of the emotion labels at least 6 times (including the initial instruction) and in some cases more depending on how much they looked through the stimuli initially, modified their piles and asked for reminders. For American participants, a few modifications to the anchoring protocol were set in place based on piloting. American participants were given the option of waiving further repetition of the list of emotion terms if they 1) indicated that they did not wish to hear it again and 2) were able to repeat the list back to the experimenter without error.

Once sorting was complete, all participants (regardless of culture or task version) were asked provide a label to describe the content of each pile. This question was initially phrased in an open-ended manner (i.e., "What is in this pile?") in order to minimize expectancy effects regarding what type of content to produce. The experimenter recorded any behavioral descriptors or mental state descriptors that the participant produced. All participants were also asked specifically for facial expression and emotion labels that were not spontaneously provided by participants during open-ended questioning. Since many Himba participants failed to provide additional information in response to these prompts, prompted responses (both for American and Himba samples) were not subjected to further analysis.

Details of Analysis Approaches Reported in Main Text:

Cluster Analysis. A hierarchical cluster analysis (*35*) produces a set of nested clusters organized in a hierarchical tree. Unlike other clustering procedures (e.g., k-means; *36*), the number of clusters can be discovered rather than being pre-specified. We used an agglomerative approach, starting with each item as its own cluster and progressively linking those items together based on an estimate of their distance from one another (computed from the number of times face stimuli appeared in the same vs separate piles; see second paragraph in this section). We employed a cluster distance measure of average linkage because it balances the limitations of single and complete linkage methods, which can be influenced by noise and outliers or force clusters with similar diameters, respectively (*37*). The average linkage clustering method uses information about all pairs of distances to assign cluster membership, not just the nearest or the furthest item pairs. We performed a cluster analysis on the data from each cultural group separately, further broken down by sorting condition, resulting in four different MDS solutions: Himba Free Sort, Himba Anchored Sort, US Free Sort, and US Anchored Sort.

To accomplish the cluster analysis, for each of the four conditions, we computed a co-occurrence matrix (*38*). Each co-occurance matrix contained a row and column for each of the 36 items in the set, resulting in a 36 x 36 symmetric matrix. Each cell in the matrix represented the number of times a given pair of face items was grouped by participants into the same pile (i.e., across participants). The larger the number in a cell, the more frequently those two items co-occurred, and thus the higher perceived similarity between those items at a group level. To perform the cluster analysis, we then converted this co-occurrence similarity matrix into a distance matrix, where a higher cell value was an indication of less similarity between items. The cluster analysis was then performed on each dissimilarity matrix and we examined the resulting dendrogram for each (Fig. S1, S2). Based on the observed increases in within cluster average item distance when the solution contained less than six clusters, (i.e., as items were grouped into larger, increasingly inclusive clusters, the clusters became less coherent), and for theoretical reasons (since 6 discrete expression portrayals were included in the set), we report the six-cluster solutions in the main text.

Multidimensional Scaling. Each co-occurrence matrix (i.e., the similarity matrix) was also subjected to an ALSCAL multidimensional scaling procedure (*39*). MDS provides an n-dimensional map that represents the similarities and dissimilarities between the face stimuli. Based on the stress-by-dimensionality plots for solutions between one and six dimensions (Fig. S3), a four dimensional solution was selected as the best fit for the two US data (free sorting and anchored) and a three dimensional solution was selected as the best fit for the two Himba data (free sorting and anchored). The multidimensional plots are presented in Figs. 4 and 5 within the main text.

To empirically identify the MDS dimensions (rather than to label them subjectively), 14 independent groups of participants rated the 36 images in the stimulus set to estimate the extent to which each face depicted a given behavior (crying, frowning, laughing, looking, pouting, scowling, smelling, and smiling; Table S2) or a given emotion as a mental state (anger, fear, disgust, sadness, neutral and happiness; Table S3). Each group of 40 participants rated the 36 faces on one property only, resulting in a total of 560 participants. Ratings for each property were on a Likert scale ranging from 1 "not at all" to 5 "extremely". The images were rated by participants recruited on Amazon's Mechanical Turk (*40*) and restricted to IP addresses located in North America. Since our experiments on Mechanical Turk would not allow for us to collect responses by Himba individuals, we decided to limit our normative ratings to the US cultural context. This allows us to test the applicability of the US cultural model of emotions (captured by these ratings) to both US and Himba cultural contexts. We computed the mean property rating for each image resulting in 14 attribute vectors with 36 means, which we then used to empirically identify the MDS dimensions (following *20*). In a series of hierarchical multiple regression analyses, we examined the extent to which behavior and emotion vectors explained variance in the MDS dimension coordinates for each dimension. By entering behavior vectors as the first block of predictors, and the emotion vectors as the second block, we were able to estimate the extent to which each dimension captured mental state (emotion) inferences (e.g., sadness) over and above merely describing the physical action of a face (e.g., pouting). Following action identification theory (*17, 18*), behavior vectors were always entered as the first block of predictors, and emotion vectors as the second. We performed a series of F-tests (Table

S4, S5, S6, S7) examining whether the behavior vectors sufficiently explained a given MDS dimension, and whether discrete emotion vectors accounted for that dimension over and above behaviors. A summary of the dimension variance accounted for by the behavioral and emotion vectors is presented in Table 2 in the main text.

Pile Label Analyses. We tallied the number of discrete emotional words that a given participant generated when anchoring their face piles. Synonyms (e.g., frustrated) were assigned to the closest discrete emotion category (e.g., anger) and included in the tally. This was done for both US and Himba data in the free sorting condition in order to examine how participants were spontaneously structuring their piles in the absence of emotion concept words (imposed in the conceptually anchored condition). Himba responses were recorded based on exact translations (rather than a summary) provided by the translator. In addition, we tallied the number of behavioral descriptors (e.g., looking, smelling, laughing, crying, smiling) used to anchor each pile. Finally, we also tallied the number of other "mental state" words were used by participants from a given cultural context. For the Himba participants, this count included a number of words that do not directly translate as "emotion" terms, but were produced by a large majority of participants in response to prompting to describe the emotional state of the individuals grouped into a given pile. These terms translated to "cool", "death", "quiet", and "wonder". For participants in the free sorting condition frequency counts for each type of word label used was subjected to a mixed model analysis of variance (ANOVA) of cultural group (US, Himba) by word type (discrete emotion, behavioral, culture-specific terms) with word type as a within-subject variable. The results of this ANOVA are summarized in the main text and Fig. 3.

Emotion Accuracy and Discrimination Analyses:

In typical studies of the universality hypothesis, researchers compute the extent to which each facial expression is "accurately" perceived, meaning the extent to which participants perceived the posed face in the way agrees with the experimenter's expectations. "Accuracy" is defined in a number of ways, most typically do participant responses match the intended emotion portrayed by the individual emitting a cue. In our experiment, we defined "accuracy" as the number of faces from a given emotion category that comprised the "dominant" content in a pile. This score allowed for items from the same category to be broken up into multiple piles and also allowed singletons (i.e., images that are placed alone and not grouped with other items) to be counted as "accurate". Accuracy was defined as the percentage of items that were grouped together by discrete emotion category and made up the majority of the content in a given pile. Himba participants had lower accuracy for sorting discrete emotion faces together with a mean accuracy of 49.615% (SD=33.559) compared to American participants, 67.101 % (SD=22.350), ($F_{1,56}$=53.412, $P < 0.001$; for means see Table S8). Lower discrete emotion sorting accuracy was found regardless of emotion portrayal type. A mixed-model analysis of variance, with accuracy as the dependent variable, emotion portrayal type as the repeated measure (portrayals of anger, fear, sadness, disgust, neutral and happiness) and cultural group as the between participants factor (American or Himba) did not yield a significant interaction, ($F_{5,280}$=.816, $P = 0.510$, Greenhouse-Geisser corrected), indicating that any impact of exposure to words did not differ by culture. In both Himba and US cultures, discrete sorting accuracy was marginally higher for participants exposed to words (M=58.576, SD=24.669) than participants who were not (M=54.223, SD=25.158), ($F_{1,118} = 3.422$, $P = 0.067$; for means see Table S8).

Discrimination scores for each emotion category were computed as the maximum number of faces from a given category that were grouped together into a pile, divided by the total number of faces within that pile. Discrimination scores represent how cleanly participants groups items from a given emotion category into a single pile. Discrimination scores more clearly reflect grouping of faces by emotion category; accuracy can be driven by more simplistic perceptual matching of faces. For example, some sadness portrayals included in the set had open mouth frowns, whereas others had closed mouth frowns. If simple feature matching occurred, the open and closed mouth expressions would be grouped into different piles. This sorting strategy would negatively impact a discrimination score (because less sadness portrayals would ultimately be grouped together) but not an accuracy score (because both piles still contained sad dominant images and thus all of those items would count as accurate). Himba participants were less able to discriminate between discrete emotions, with a mean discrimination proportion of .644 (SD=0.276) compared to US participants .740 (SD=0.196), ($F_{1,56}$=30.766, P < 0.001; for means see Table S9). Lower discrimination for Himba participants was found regardless of emotion portrayal type. A mixed-model analysis of variance, with discrimination as the dependent variable, emotion type as the repeated measure (portrayals of anger, fear, sadness, disgust, neutral and happiness) and cultural group as the between participants factor (American or Himba) did not yield a significant interaction, ($F_{5,280}$=1.373, P = 0.248, Greenhouse-Geisser corrected). In both Himba and US cultures, emotion words did not help participants to sort into relatively cleaner piles (i.e., with less error items), ($F_{1,118}$ = 1.921, P = 0.168; for means see Table S9).

Figures and Legends:

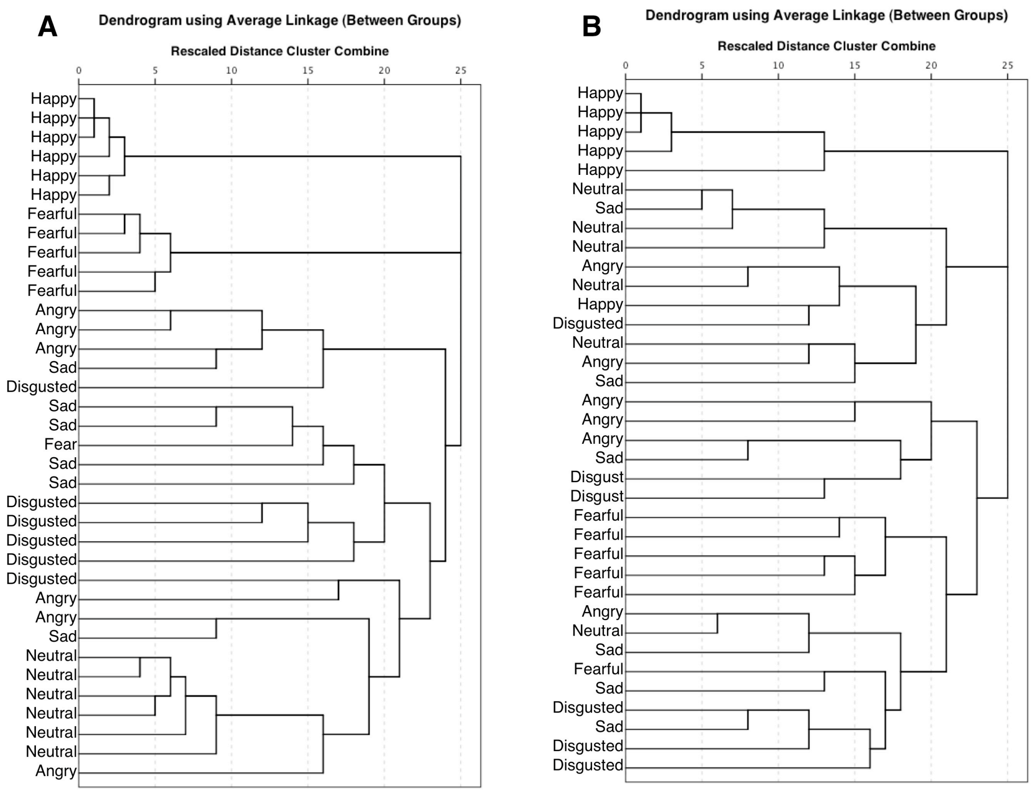


Fig. S1. Cluster analysis dendrograms for US (**A**) and Himba (**B**) participants' free sorting data. As an increasing number of items are included in a cluster, the more inclusive cluster also increases the average distance within the cluster. As each cluster (which starts out as a single item) is united with another cluster (i.e., agglomerated), this is indicated in the dendrogram by a bracket. The longer the bracket between two clusters (i.e., the bigger the distance before two clusters are joined), the more spread out the newly formed cluster is (i.e., the bigger the difference between the items joining the new cluster). The average distance within a cluster is generally lower for US participants than for Himba participants (evidenced by the longer brackets early even at the first stage of clustering in the Himba data), and this is particularly evident in clusters of happy, fearful and neutral items, which all have relatively small average cluster distances for US participants. In both cluster solutions, the happy portrayals are most tightly clustered together (indicated by the very short brackets between happy items) indicating that across both cultures the perceived similarity between happy items is greater than within other discrete emotion portrayal types.

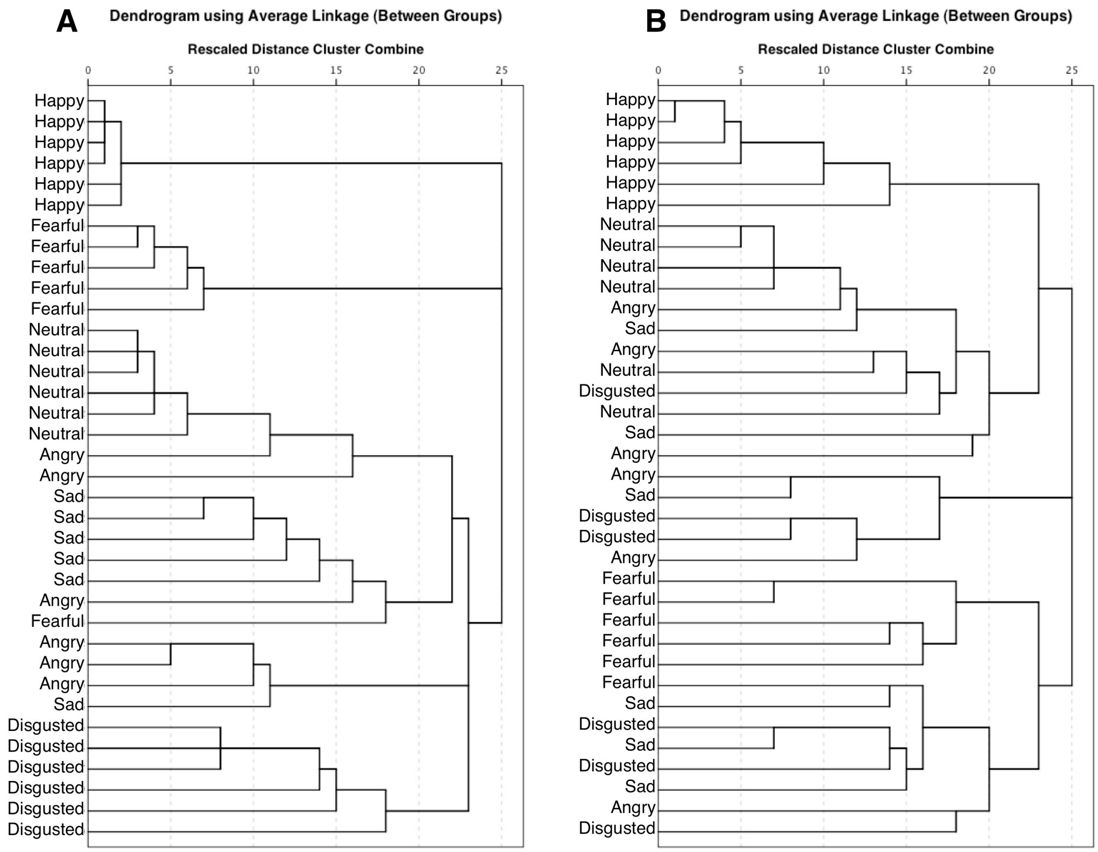


Fig. S2. Cluster analysis dendrograms for US (**A**) and Himba (**B**) participants' anchored sorting data. See Fig. S1 for an overview of how to read these dendrograms. For the US participants, a relatively clear 6-cluster solution emerged, such that an average cluster distance below 20 is maintained for all clusters.

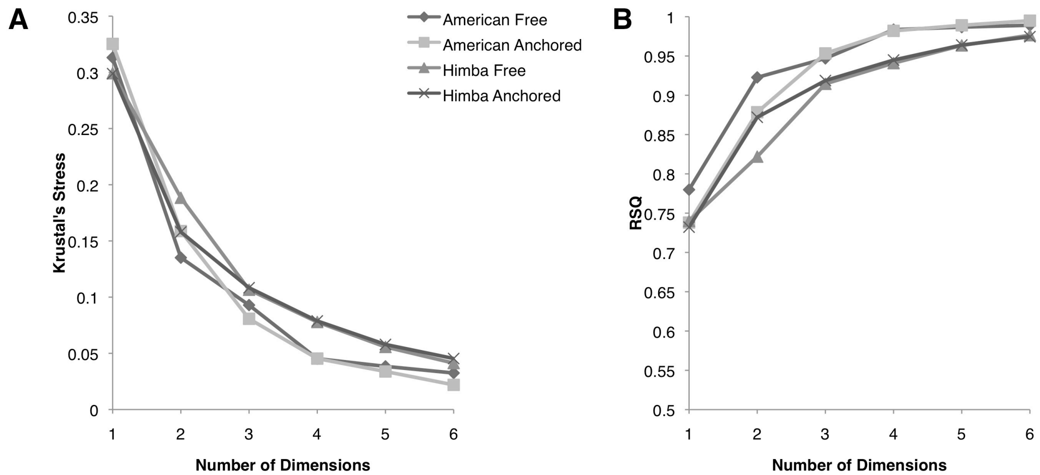


Fig. S3. Stress-by-Dimension (**A**) and *RSQ*-by-dimension (**B**) plots for multidimensional scaling (MDS) solutions for the four sorting conditions. The number of dimensions in a given solution is plotted on the x-axis. A clear elbow in the plots is visible for US data at four dimensions, indicating that a 4D solution optimally fits the US data. Himba data have an earlier, although less well defined elbow at three dimensions, indicating that a 3D solution optimally fits the Himba data.

Tables:

Table S1. Correlations between Emotion and Behavior Ratings across the 36 Face Stimuli.

Attribute Rated	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Anger		0.121	.750**	0.261	-0.238	-.685**	0.078	.812**	-.622**	0.123	.746**	.945**	.512**	-.653**
2. Fear			.473**	.519**	-.402*	-.563**	.398*	.388*	-.508**	.336*	0.204	0.124	0.168	-.515**
3. Disgust				.444**	-.489**	-.690**	0.283	.853**	-.588**	0.165	.675**	.839**	.782**	-.614**
4. Sadness					-0.122	-.652**	.878**	.672**	-.581**	-0.307	.671**	.333*	0.072	-.588**
5. Neutral						-0.064	-0.227	-.365*	-0.211	-.684**	-0.276	-0.268	-.639**	-0.178
6. Happiness							-.376*	-.781**	.979**	0.197	-.663**	-.687**	-0.195	.986**
7. Crying								.491**	-0.295	-0.307	.505**	0.126	0.086	-0.297
8. Frowning									-.679**	-0.018	.928**	.878**	.532**	-.703**
9. Laughing										0.26	-.577**	-.610**	-0.098	.991**
10. Looking											-0.109	0.037	0.313	0.226
11. Pouting												.817**	.452**	-.598**
12. Scowling													.614**	-.640**
13. Smelling														-0.124
14. Smiling														

Note. Correlations among mean emotion ratings (not bolded) are consistent with physical and affective similarities between face stimuli (e.g., anger and disgust are both high arousal negative emotions and both are portrayed with a furrowed brow). Correlations among mean behavior ratings (not bolded) are consistent with similarities between the physical manifestations of those behaviors on the face (e.g., crying is correlated with both frowns and pouts). Correlations between behavior and emotion vectors (bolded) indicate that behaviors (action identification) are not redundant with emotions (mental state inference) and that a single behavior is associated with more than one emotion. For example, both anger and disgust ratings were correlated with ratings of smelling. ** = $p \le .05$, ** = $p \le .01$. Findings like this call into question the routine practice of equating action identification with emotion perception in cross-cultural studies that claim support for the universality hypothesis. For example, most experiments conducted in remote cultural samples (*1, 8-10*) (Table 1), participants were provided with a situational frame including action identification information such as "He (she) is looking at something which smells bad". This situational frame was presented along with two to three facial expressions, for example a face with a nose-wrinkle portraying disgust and a pouting face portraying sadness; participants, on average, choose the disgust face. It is typically assumed that selecting the nose-wrinkled face supports the conclusion that the perceiver correctly "recognized" the expression of a mental state of "disgust". These findings, along with our Himba data, suggest another interpretation: participants might have engaged in action identification rather than in mental state inference and selected a face that appeared to be engaged in a situationally-tuned behavior (e.g. smelling). That is, action identification by individuals from distinct cultural contexts might be mistakenly treated as evidence of universal emotion recognition.

Table S2. Mean Behavior Ratings for Face Stimuli

Emotion Portrayed		Crying N=40	Frowning N=40	Laughing N=40	Looking N=40	Pouting N=40	Scowling N=40	Smelling N=40	Smiling N=40
		Behavior Ratings							
Anger	Mean	1.24[a]	**3.12[b]**	1.07[c]	**3.33[b]**	2.74[d]	**3.34[b]**	2.26[e]	1.14[c]
	(SD)	(.12)	(.52)	(.12)	(.48)	(.61)	(.78)	(.29)	(.21)
Fear	Mean	1.43[a]	2.53[b]	1.02[c]	**3.98[d]**	1.86[e]	2.25[f]	2.39[g]	1.10[h]
	(SD)	(.52)	(.47)	(.03)	(.25)	(.31)	(.28)	(.12)	(.06)
Disgust	Mean	1.67[a]	**3.32[b]**	1.15[c]	**3.24[b]**	2.75[d]	**3.39[b]**	**3.21[b]**	1.29[e]
	(SD)	(.49)	(.62)	(.19)	(.35)	(.73)	(.55)	(.48)	(.34)
Sadness	Mean	2.31[ad]	**3.11[b]**	1.05[c]	**3.07[b]**	**2.98[b]**	2.43[a]	2.08[d]	1.14[e]
	(SD)	(1.11)	(.48)	(.10)	(.19)	(.70)	(.64)	(.22)	(.17)
Neutral	Mean	1.17[a]	1.61[b]	1.10[c]	**2.87[d]**	1.44[e]	1.81[f]	1.79[f]	1.29[g]
	(SD)	(.07)	(.24)	(.07)	(.17)	(.11)	(.21)	(.12)	(.25)
Happiness	Mean	1.02[a]	1.01[b]	3.28[c]	3.66[d]	1.01[b]	1.11[e]	2.19[f]	**4.24[g]**
	(SD)	(.02)	(.02)	(.38)	(.18)	(.01)	(.09)	(.11)	(.48)

Note. Independent-samples t-tests were conducted to compare multiple behaviors within a given emotion portrayal type. Items with the same superscript did not differ statistically from one another (i.e., p >.05). Bolded items are the highest rated behavior (or set of behaviors) for a given emotion portrayal type. For example, anger portrayals were equivalently and most strongly associated with frowning, looking and scowling behaviors.

Table S3. Mean Emotion Ratings of Face Stimuli

		Emotion Rating Scale					
Emotion Portrayed		Anger N=40	Fear N=40	Disgust N=40	Sadness N=40	Neutral N=40	Happiness N=40
Anger	Mean	**3.23**	1.59***	2.54**	2.13***	2.20***	1.29***
	(SD)	(.80)	(.91)	(1.16)	(1.11)	(.87)	(.50)
Fear	Mean	2.21***	**3.87**	2.76***	2.38***	1.22***	1.22***
	(SD)	(1.10)	(.95)	(1.31)	(1.08)	(.60)	(.48)
Disgust	Mean	2.92**	2.12**	**3.70**	2.40***	1.36***	1.30***
	(SD)	(1.12)	(1.06)	(.99)	(1.11)	(.67)	(.52)
Sadness	Mean	2.19***	2.48***	2.22***	**3.40**	2.05***	1.23***
	(SD)	(.97)	(1.09)	(1.05)	(.93)	(.94)	(.43)
Neutral	Mean	1.81***	1.49***	1.46***	1.81***	**4.04**	1.71***
	(SD)	(.78)	(.68)	(.66)	(.80)	(.97)	(.66)
Happiness	Mean	1.04***	1.00***	1.05***	1.06***	1.56***	**4.17**
	(SD)	(.22)	(.0)	(.22)	(.29)	(.97)	(.68)

Note. Ratings on the diagonal (bolded) are ratings on the scale that a given facial pose was meant to portray. Independent samples *t*-tests (2-tailed) were conducted comparing the ratings of each emotion portrayal type on the target emotion scale (bolded) against each of the other rating scales. All comparisons were statistically significant, indicating that perceiver-based ratings were consistent with the intended portrayal type (across all categories). Non-zero ratings on other scales aside from the intended portrayal, are consistent with prior normative data suggesting that canonical facial expressions can still be perceived in line with other emotions, particularly those that are perceptually (*41*) and affectively similar to one another. $^{**}p \leq .01$, $^{***}p \leq .001$

Table S4. Dimension Identification: US Free Sort Data

Dimension	Variable	Model 1 (Behavior)			Model 2 (Behavior + Emotion)		
		B	*SE B*	β	*B*	*SE B*	β
1	Crying	0.067	0.166	0.037	0.378	0.268	0.211
	Frowning	0.107	0.277	0.086	0.043	0.404	0.035
	Laughing	-0.467	0.546	-0.334	-0.581	0.562	-0.416
	Looking	0.679	0.16	0.264***	0.393	0.329	0.153
	Pouting	-0.076	0.195	-0.056	0.075	0.334	0.056
	Scowling	-0.263	0.217	-0.208	0.16	0.364	0.126
	Smelling	0.905	0.176	0.386***	0.538	0.373	0.229
	Smiling	1.081	0.403	1.079*	0.418	0.539	0.417
	Anger				-0.466	0.308	-0.344
	Disgust				-0.011	0.376	-0.009
	Fear				0.05	0.226	0.043
	Sad				-0.491	0.267	-0.363
	Neutral				-0.306	0.197	-0.272
	Happiness				0.629	0.714	0.582
	R^2		.938			.958	
	F for change in R^2		50.700***			1.716	
2	Crying	-0.236	0.211	-0.14	0.176	0.238	0.104
	Frowning	-1.174	0.353**	-1.011	-0.148	0.358	-0.127
	Laughing	0.03	0.697	0.023	-0.332	0.498	-0.253
	Looking	-1.245	0.204***	-0.514	0.039	0.292	0.016

	Pouting	0.883	0.249**	0.697	0.25	0.296	0.197
	Scowling	0.649	0.277*	0.544	0.371	0.322	0.311
	Smelling	-0.863	0.224**	-0.391	-0.166	0.331	-0.075
	Smiling	0.516	0.513	0.548	0.46	0.478	0.489
	Anger				-0.276	0.273	-0.217
	Disgust				-0.327	0.333	-0.274
	Fear				-0.573	0.201	-0.521**
	Sad				0.032	0.237	0.025
	Neutral				0.377	0.175	0.357
	Happiness				-0.002	0.633	-0.002*
	R^2		.885			.963	
	F for change in R^2		26.058***			7.300***	
3	Crying	-0.094	0.156	-0.066	-0.388	0.255	-0.271
	Frowning	0.255	0.261	0.26	-0.184	0.383	-0.187
	Laughing	-0.475	0.515	-0.427	-0.067	0.533	-0.061
	Looking	0.522	0.151	0.255**	0.179	0.312	0.088
	Pouting	-0.663	0.184	-0.617**	-0.418	0.317	-0.389
	Scowling	-0.852	0.205	-0.842***	-0.857	0.345	-0.848*
	Smelling	-0.118	0.166	-0.063	-0.154	0.354	-0.082
	Smiling	-0.215	0.38	-0.269	0.474	0.511	0.594
	Anger				0.428	0.292	0.397
	Disgust				0.133	0.357	0.132
	Fear				0.397	0.215	0.426

	Sad				0.214	0.253	0.199
	Neutral				0.228	0.187	0.255
	Happiness				-0.621	0.678	-0.721
	R^2		.913			.941	
	F for change in R^2		35.236***			1.651	
4	Crying	-0.59	0.2	-0.517**	-0.107	0.288	-0.093
	Frowning	-0.366	0.335	-0.466	0.648	0.434	0.824
	Laughing	-0.161	0.66	-0.181	-0.406	0.604	-0.457
	Looking	0.58	0.193	0.354**	0.511	0.354	0.312
	Pouting	0.124	0.235	0.145	-0.434	0.359	-0.505
	Scowling	0.717	0.262	0.887	1.232	0.391	1.524**
	Smelling	-0.666	0.212	-0.446**	-0.045	0.401	-0.03
	Smiling	0.062	0.486	0.097*	-0.534	0.579	-0.837
	Anger				-0.226	0.331	-0.263
	Disgust				-0.977	0.404	-1.21*
	Fear				0.257	0.243	0.345
	Sad				-0.479	0.287	-0.556
	Neutral				0.057	0.212	0.08
	Happiness				0.887	0.768	1.288
	R^2		.776			.881	
	F for change in R^2		11.669***			3.09*	

Note. Hierarchical multiple regressions were conducted to identify the derived dimensions from the MDS analyses. Predictors were entered in blocks: Model 1 contained only behavior vectors as predictors; Model 2 contained both behavior and emotion vectors. Coefficients for each predictor within each model are presented as well as the relevant F-test for a significant change in prediction (R^2)

between models. A separate regression was conducted for each dimension that resulted from the Multidimensional scaling of sorting data. * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$

Table S5. Dimension Identification: Himba Free Sort Data

		Model 1 (Behavior)			Model 2 (Behavior + Discrete)		
Dimension	Variable	*B*	*SE B*	β	*B*	*SE B*	β
1	Crying	-0.677	0.283	-0.369*	-0.373	0.444	-0.203
	Frowning	-1.366	0.473	-1.081**	-0.743	0.668	-0.588
	Laughing	-0.672	0.933	-0.47	-1.205	0.931	-0.843
	Looking	-0.627	0.273	-0.238*	-0.116	0.545	-0.044
	Pouting	1.306	0.333	0.946***	1.167	0.553	0.846*
	Scowling	0.066	0.371	0.051	0.05	0.603	0.038
	Smelling	-0.553	0.3	-0.23	-0.662	0.618	-0.276
	Smiling	0.831	0.688	0.81	0.144	0.892	0.14
	Anger				0.264	0.509	0.19
	Disgust				0.476	0.623	0.367
	Fear				0.344	0.375	0.287
	Sad				0.095	0.442	0.069
	Neutral				0.729	0.327	0.635*
	Happiness				2.209	1.183	1.996
	R^2		.826			.890	
	F for change in R^2		16.050***			2.051	
2	Crying	-0.332	0.275	-0.231	-0.592	0.479	-0.412
	Frowning	0.124	0.46	0.126	0.591	0.72	0.598
	Laughing	-0.67	0.907	-0.6	-0.718	1.003	-0.642
	Looking	-1.081	0.266	-0.525***	-0.439	0.587	-0.213

	Pouting	0.016	0.324	0.015	-0.714	0.595	-0.661
	Scowling	0.34	0.36	0.335	0.415	0.649	0.409
	Smelling	-0.652	0.291	-0.347*	0.294	0.666	0.156
	Smiling	0.261	0.668	0.325	0.856	0.962	1.067
	Anger				-0.445	0.549	-0.41
	Disgust				-0.928	0.671	-0.915
	Fear				-0.684	0.404	-0.73
	Sad				0.47	0.477	0.434
	Neutral				-0.245	0.352	-0.273
	Happiness				-1.567	1.275	-1.811
	R^2		.732			.792	
	F for change in R^2		9.215***			1.012	
3	Crying	-0.37	0.299	-0.291	-0.01	0.493	-0.008
	Frowning	-1.3	0.5	-1.484*	-0.447	0.742	-0.511
	Laughing	-0.886	0.985	-0.895	-0.643	1.033	-0.649
	Looking	0.733	0.289	0.402*	0.519	0.605	0.284
	Pouting	1.38	0.352	1.443***	1.055	0.613	1.103
	Scowling	0.237	0.392	0.263	0.556	0.669	0.617
	Smelling	0.1	0.317	0.06	0.846	0.686	0.508
	Smiling	0.264	0.726	0.372	0.545	0.99	0.766
	Anger				0.579	0.565	0.603
	Disgust				-0.765	0.691	-0.851
	Fear				0.924	0.416	1.114*

Sad		-0.324	0.491	-0.338
Neutral		0.774	0.363	0.971*
Happiness		0.51	1.313	0.665
R^2	.597		.719	
F for change in R^2	4.993***		1.525	

Note. Hierarchical multiple regressions were conducted to identify the derived dimensions from the MDS analyses. Predictors were entered in blocks: Model 1 contained only behavior vectors as predictors; Model 2 contained both behavior and emotion vectors. Coefficients for each predictor within each model are presented as well as the relevant F-test for a significant change in prediction (R^2) between models. A separate regression was conducted for each dimension that resulted from the Multidimensional scaling of sorting data. * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$

Table S6. Dimension Identification: US Conceptually Anchored Sort Data

		Model 1 (Behavior)			Model 2 (Behavior + Discrete)		
Dimension	Variable	*B*	*SE B*	β	*B*	*SE B*	β
1	Crying	-0.445	0.252	-0.25	0.364	0.287	0.204
	Frowning	-0.742	0.421	-0.605	0.22	0.432	0.179
	Laughing	0.587	0.83	0.424	0.058	0.601	0.042
	Looking	-1.08	0.243	-0.423***	0.15	0.352	0.059
	Pouting	0.366	0.296	0.273	0.121	0.357	0.091
	Scowling	0.412	0.33	0.327	-0.06	0.389	-0.048
	Smelling	-0.596	0.267	-0.256*	-0.526	0.399	-0.226
	Smiling	0.185	0.612	0.186	-0.569	0.576	-0.572
	Anger				-0.203	0.329	-0.151
	Disgust				0.337	0.402	0.268
	Fear				-0.43	0.242	-0.37
	Sad				-0.425	0.286	-0.317
	Neutral				0.485	0.211	0.436*
	Happiness				1.235	0.764	1.151
	R^2		.854			.951	
	F for change in R^2		19.702***			7.034***	
2	Crying	0.114	0.178	0.064	-0.012	0.264	-0.007
	Frowning	0.489	0.298	0.403	0.061	0.397	0.051
	Laughing	-0.641	0.587	-0.467	-0.742	0.553	-0.54
	Looking	0.835	0.172	0.33***	0.136	0.324	0.054

	Pouting	-0.506	0.209	-0.382*	-0.297	0.329	-0.224
	Scowling	-0.374	0.233	-0.299	0.098	0.358	0.078
	Smelling	1.295	0.189	0.561***	0.825	0.368	0.358*
	Smiling	1.017	0.432	1.032*	0.476	0.531	0.483
	Anger				-0.303	0.303	-0.227
	Disgust				0.036	0.371	0.029
	Fear				0.138	0.223	0.12
	Sad				-0.115	0.263	-0.087
	Neutral				-0.499	0.194	-0.452*
	Happiness				0.567	0.704	0.533
	R^2		.926			.958	
	F for change in R^2		41.931***			2.706*	
3	Crying	0.388	0.215	0.309	-0.599	0.255	-0.476*
	Frowning	0.16	0.36	0.184	0.146	0.384	0.168
	Laughing	0.204	0.709	0.208	-0.147	0.535	-0.15
	Looking	-0.538	0.208	-0.298*	-0.35	0.313	-0.194
	Pouting	0.275	0.253	0.291	-0.35	0.318	-0.37
	Scowling	-1.159	0.282	-1.302***	-0.875	0.346	-0.983*
	Smelling	0.563	0.228	0.343*	1.107	0.356	0.673**
	Smiling	-0.29	0.523	-0.413	0.158	0.513	0.225
	Anger				-0.018	0.293	-0.019
	Disgust				-0.371	0.358	-0.417
	Fear				-0.242	0.216	-0.294

	Sad				1.37	0.254	1.445***
	Neutral				-0.102	0.188	-0.13
	Happiness				-0.226	0.68	-0.298
	R^2		.786			.923	
	F for change in R^2		12.403***			6.202***	
4	Crying	-0.146	0.178	-0.117	-0.12	0.239	-0.096
	Frowning	-0.337	0.297	-0.394	-0.233	0.359	-0.273
	Laughing	-0.781	0.586	-0.806	-0.312	0.5	-0.322
	Looking	0.956	0.172	0.536***	0.754	0.293	0.422*
	Pouting	0.051	0.209	0.055	-0.15	0.297	-0.161
	Scowling	-0.654	0.233	-0.744**	0.299	0.324	0.34
	Smelling	-0.18	0.188	-0.111	0.36	0.333	0.222
	Smiling	0.003	0.432	0.004	0.738	0.48	1.062
	Anger				-0.217	0.274	-0.232
	Disgust				-0.819	0.335	-0.932*
	Fear				0.581	0.202	0.716**
	Sad				-0.07	0.238	-0.075
	Neutral				0.3	0.176	0.386
	Happiness				-0.87	0.636	-1.161
	R^2		.851			.931	
	F for change in R^2		19.228***			4.069**	

Note. Hierarchical multiple regressions were conducted to identify the derived dimensions from the MDS analyses. Predictors were entered in blocks: Model 1 contained only behavior vectors as predictors; Model 2 contained both behavior and emotion vectors. Coefficients for each predictor within each model are presented as well as the relevant F-test for a significant change in prediction (R^2)

between models. A separate regression was conducted for each dimension that resulted from the Multidimensional scaling of sorting data. * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$

Table S7. Dimension Identification: Himba Conceptually Anchored Sort Data

		Model 1 (Behavior)			Model 2 (Behavior + Discrete)		
Dimension	Variable	*B*	*SE B*	β	*B*	*SE B*	β
1	Crying	-0.821	0.257	-0.434**	-0.32	0.391	-0.169
	Frowning	-1.137	0.43	-0.873*	-0.267	0.589	-0.205
	Laughing	-0.319	0.847	-0.217	-0.523	0.82	-0.355
	Looking	-0.962	0.248	-0.354***	-0.161	0.48	-0.059
	Pouting	1.21	0.302	0.851***	0.934	0.487	0.657
	Scowling	0.189	0.337	0.141	-0.148	0.531	-0.11
	Smelling	-0.871	0.272	-0.352**	-0.496	0.545	-0.2
	Smiling	0.676	0.624	0.639	0.552	0.786	0.522
	Anger				0.28	0.449	0.196
	Disgust				0.045	0.549	0.033
	Fear				0.08	0.33	0.065
	Sad				-0.167	0.39	-0.117
	Neutral				0.738	0.288	0.623*
	Happiness				0.811	1.042	0.711
	R^2		.865			.920	
	F for change in R^2		21.671***			2.393	
2	Crying	-0.445	0.239	-0.305	-0.062	0.333	-0.043
	Frowning	-0.925	0.4	-0.921*	0.41	0.5	0.408
	Laughing	-0.655	0.789	-0.577	-1.04	0.697	-0.915

	Looking	-0.127	0.231	-0.061	0.697	0.408	0.333
	Pouting	1.091	0.282	0.995***	0.325	0.414	0.296
	Scowling	0.621	0.314	0.602	0.536	0.451	0.519
	Smelling	-0.125	0.254	-0.065	0.792	0.463	0.415
	Smiling	0.165	0.582	0.203	-0.003	0.668	-0.004
	Anger				0.162	0.381	0.147
	Disgust				-0.624	0.466	-0.605
	Fear				0.023	0.281	0.025
	Sad				0.043	0.331	0.039
	Neutral				0.641	0.245	0.702*
	Happiness				0.838	0.886	0.953
	R^2		.803			.903	
	F for change in R^2		13.776***			3.586*	
3	Crying	-0.363	0.233	-0.313	0.077	0.383	0.067
	Frowning	-0.602	0.39	-0.756	-0.026	0.577	-0.032
	Laughing	-0.19	0.768	-0.211	-0.134	0.803	-0.149
	Looking	1.2	0.225	0.722***	0.854	0.47	0.514
	Pouting	0.549	0.274	0.631	0.502	0.477	0.577
	Scowling	-0.017	0.305	-0.021	-0.06	0.52	-0.074
	Smelling	0.21	0.247	0.139	0.427	0.534	0.282
	Smiling	-0.004	0.566	-0.006	-0.305	0.77	-0.471
	Anger				0.456	0.44	0.522
	Disgust				-0.311	0.538	-0.38

Fear		0.574	0.324	0.76
Sad		-0.553	0.382	-0.634
Neutral		0.338	0.282	0.467
Happiness		0.873	1.021	1.25
R^2	.704		.795	
F for change in R^2	8.029***		1.549	

Note. Hierarchical multiple regressions were conducted to identify the derived dimensions from the MDS analyses. Predictors were entered in blocks: Model 1 contained only behavior vectors as predictors; Model 2 contained both behavior and emotion vectors. Coefficients for each predictor within each model are presented as well as the relevant F-test for a significant change in prediction (R^2) between models. A separate regression was conducted for each dimension that resulted from the Multidimensional scaling of sorting data. * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$

Table S8. Descriptive statistics for accuracy scores.

Emotion Portrayed	Cultural Group	Free Sorting Mean % Accuracy	Free Sorting (SD)	Anchored Sorting Mean % Accuracy	Anchored Sorting (SD)
Anger					
	US	0.578	0.046	0.523	0.043
	Himba	0.295	0.051	0.167	0.049
Fear					
	US	0.818	0.045	0.792	0.043
	Himba	0.506	0.05	0.512	0.048
Sadness					
	US	0.344	0.049	0.62	0.046
	Himba	0.103	0.055	0.25	0.053
Disgust					
	US	0.469	0.047	0.644	0.044
	Himba	0.314	0.052	0.339	0.05
Happiness					
	US	0.99	0.033	0.991	0.031
	Himba	0.75	0.036	0.732	0.035
Neutral					
	US	0.828	0.051	0.912	0.048
	Himba	0.513	0.057	0.548	0.055

Note. Means and standard deviations are presented.

Table S9. Descriptive Statistics for Discrimination Scores.

		Free Sorting		Anchored Sorting	
Emotion Portrayed	Cultural Group	Mean Proportion Discriminated	(SD)	Mean Proportion Discriminated	(SD)
Anger					
	US	0.612	0.037	0.659	0.035
	Himba	0.455	0.042	0.427	0.04
Fear					
	US	0.868	0.04	0.918	0.038
	Himba	0.592	0.045	0.745	0.043
Sadness					
	US	0.55	0.033	0.651	0.031
	Himba	0.317	0.037	0.41	0.035
Disgust					
	US	0.74	0.045	0.776	0.042
	Himba	0.56	0.05	0.527	0.048
Happiness					
	US	0.952	0.031	0.918	0.029
	Himba	0.826	0.034	0.775	0.033
Neutral					
	US	0.718	0.031	0.715	0.03
	Himba	0.469	0.035	0.509	0.034

Note. Means and standard deviations are presented.

Chapter IV

Within Cultural Stability in Mental State Inference

We examined whether the tendency to infer mental states from facial portrayals of emotion was dependent on instructions to focus on the mind. We manipulated the focus of participants from a US cultural context on mental versus behavioral states and examined the subsequent impact on emotion perception.

Mental State Attribution in Western Culture

One of the central assumptions about social cognition in Western psychology is that people routinely and automatically understand human actions in terms of the mental states that supposedly caused those actions. Spontaneously attending to the minds of others is considered central to normal functioning in Western culture (Bruner & Taguiri, 1954; Mitchell, 2009). Likewise, failure to attend to mental states is considered a central symptom in some disorders, such as the autism spectrum, leading to disruptions in social functioning (Baron-Cohen, 1995). Distortions in mind perception have recently been identified in various forms of Western psychopathology such as schizotypy and psychopathy (Gray, Jenkins, Heberlein, & Wegner, 2011), lending further support to the notion that inferring mental states in others is central to one's own mental health.

In the Western conception of "emotion perception," mental state inference is assumed to be the natural endpoint to the process. Emotion perception is often described as a cascade of "feed-forward" processing, driven by a clear external "signal" emitted non-verbally by another individual (Haxby et al., 2001). The perceiver's nervous system has been described as evolved to "decode" these signals (Schyns, Petro, & Smith, 2009)[i], such that the emotion perception is simply a recognition process, rather than an inference on the part of the perceiver. While mental state inference is assumed to occur in most models, and it is assumed that the ability to apply a label to a face is indicative of that inference, it is rarely explicitly discussed or measured.

Mental state inference may not be as routine as models of emotion have often implied, however. Failure to engage in mind perception also occurs in the normal population, and is thus not restricted to instances of pathology. That is, people vary in the extent to which they infer mental states in others as the cause of observable actions (Kozak, Marsh, & Wegner, 2006). This individual difference has been termed "action identification". Individuals low in action identification tend to view actions in simple behavioral terms (e.g., eating might be described as *chewing and swallowing*), whereas individuals who are high in action identification infer the mental states and intentions behind actions (e.g., eating might be described as *enjoying the taste of food*). Further, the tendency to engage in mental state attribution is also based on the situational constraints—when behavior is more ambiguous in its causes this increases the likelihood that a mind will be inferred behind that behavior (Waytz, Gray, Epley, & Wegner, 2010).

The distinction between action perception and mental state inference has received relatively little attention in the behavioral literature on emotion perception. But recent models of emotion, motivated by the underlying neural networks involved, yield a far different picture from the "cascade" of processing implicit in some models of emotion. Instead, emotion perception is described as a product of the mutual interaction of more basic systems in the brain (Spunt & Lieberman, 2012; Zaki, Hennigan, Weber, & Ochsner, 2010), one involved broadly in action perception (often referred to as a mirroring network involving regions such as posterior inferior frontal gyrus, dorsal premotor cortex, and the inferior parietal lobule; although this network itself may be subdivided further), and another involved broadly in mental state inference (including regions such as the dorsomedial and ventromedial prefrontal cortices, posterior cingulate cortex/precuneus, temporoparietal parietal junction, the posterior superior temporal sulcus, and the anterior temporal cortex; although this network is not necessarily specific for this function). Recent data suggests that these networks have increased functional connectivity to one another during emotion perception (Spunt & Lieberman, 2012). These two networks work in tandem to accomplish the "emotion perception" task—producing a representation of action, as well as the mental states that produced those actions. Interestingly, activity in one network or the other tends to "win" out when contextual information informing mental state inference is at odds with the nonverbal behavior observed (Zaki et al., 2010). But recruitment of the mentalizing network appears to be fairly routine—it is even engaged when perceivers view non-affective social scenes with no instructions or need to engage in mentalizing (Wagner, Kelley, & Heatherton, 2011). The extent to which the "mentalizing" network is engaged is related to trait level empathy, or the tendency to both apprehend and share in another's affective state. Taken together, these findings converge with the action identification framework described above, where a behavioral understanding of action is primary and is not necessarily coupled with mental state inference, but often is.

While variation in the tendency to perceive mental states behind actions has been characterized within individual difference and situational frameworks, there has been relatively little discussion of the possibility that cultures may also vary in the extent to which they routinely engage in mental state inference. The cross-cultural research presented in the first two chapters of this dissertation suggests that there may be cultural differences in the tendency to perceive actions versus mental states in response to the emotional facial actions or vocalizations of others. Individuals from the Himba culture, from an isolated ethnic group in Northwest Namibia, tended to describe facial actions and vocalizations with behavioral descriptors, such as "crying", rather than with mental state descriptors, such as “sadness”. Further, data driven analysis of how Himba individuals perceived the similarity between facial actions supported the conclusion that they attended to behavioral dimensions, rather than the Western mental state categories we tested. Participants from the United States, on the other hand, tended to infer mental states and made less purely behavioral attributions. These findings suggest the intriguing possibility that there is a relatively stable cultural difference in the tendency to infer mental versus behavioral content in non-verbal emotional cues.

While a stable cross-cultural difference between Himba and US cultures is a possibility, it also may be the case that the observed differences in mental state inference between cultures were driven by the constraints of our experimental paradigm. For example, one possibility is that during the translation process, the translator emphasized focusing on behaviors, unbeknownst to the researchers. This would leave open the possibility that the cross-cultural pattern observed was simply an issue with the interpretation and understanding of instructions, rather than a real cultural difference between Himba and US participants. If it is the case that no stable cultural difference exists, and instead our effects were driven by a simple issue with instructions, we should be able to easily induce the “Himba” pattern of sorting in US participants simply by asking them to sort by behaviors (rather than emotions).

The Present Study

To address the possibility that cross-cultural differences observed in Chapter 2 were driven by differences in instructions, as opposed to a more stable cultural difference in the tendency to infer mental states (perhaps based on variation in the concept of emotion), we decided to replicate our original experiment within a US culture context, but to manipulate the manner in which we instruct participants. We varied the extent to which our instructions emphasized mental versus behavioral content. Specifically, participants were instructed to sort a set of faces portraying emotional actions based on: (1) behavior, (2) emotion, or (3) mental states. If the Himba participants sorted items in a distinct manner from US participants simply because the translation process led to an emphasis on behavior, participants in a US cultural context instructed to emphasize behavior should produce a similar sorting pattern to the Himba participants. If, on the other hand, a more stable cultural difference is at hand, asking US participants to sort portrayals of emotional facial actions more behaviorally versus mentally shouldn't dramatically shape sorting, because US participants should engage in mental state inference regardless of condition. This second pattern of findings would suggest that (a) simple instruction differences are unlikely to account for cross-cultural differences observed and (b) that the tendency to engage in mental state inference (to emotional cues) is relatively routine and inflexible in a US cultural sample exposed to emotional stimuli.

Methods

Participants. Participants were 76 Northeastern University undergraduate students (32 male, 44 female; mean age= 18.85, sd= 1.12). All participants were Native English speakers and over 18 years of age. Participants were recruited via flyers around campus or from the Psychology Department participant pool.

Stimuli. Stimuli were 4x6 cards containing photographs of facial expressions of emotion posed by African American individuals. Stimuli were the same as those used in Study 2.

Procedure. All participants were consented prior to the start of the experimental session. As in Study 2, all participants in this experiment were asked to complete a sorting task as well as answer questions about their sort. Participants were randomly assigned to one of three possible "focus" conditions: behavioral (N=25; 11 male), emotion (N=25; 10 male), or mental (N=26; 11 male). The stimuli were identical in all three conditions, but participants were instructed to sort them in a distinct manner depending on condition.

Participants in the *behavior* condition were instructed*: "I am going to give you a pile of pictures. The people in this pile are all doing something. Some of them are behaving in same way and some of the people are behaving differently. What I want you to do now is to sort the faces in the pile based on how the people are behaving or what it is that they are doing. You should create piles where each person is doing exactly the same thing. At the end, each pile you've made should have pictures of only people who are behaving in the same way. Are these instructions clear? You can create as many piles as you need to. This is not timed, so feel free to take as long as you need. You can also change the piles whenever you wish, while you are sorting or at the end. Do you have any questions?"*

Participants in the *mental state* condition were instructed: *"I am going to give you a pile of pictures. The people in this pile all have an experience going on in their minds. Some of them have the same sort of mental experience, whereas others have a different type of experience in their minds. What I want you to do now is to sort the faces in the pile based on what is happening inside the minds of the people in the pictures. You should create piles where each person has exactly the same type of mental state. At the* *end, each pile you've made should have pictures of only people who experiencing the same thing in their minds. Are these instructions clear? You can create as many piles as you need to. This is not timed, so feel free to take as long as you need. You can also change the piles whenever you wish, while you are sorting or at the end. Do you have any questions?"*

Participants in the *emotion* condition were instructed: *"I am going to give you a pile of pictures. The people in this pile are all feeling something. Some of them feel the same emotion and some of the people feel differently. What I want you to do now is to sort the faces in the pile based on how the people feel. You should create piles where each person feels exactly the same way. At the end, each pile you've made should have pictures of only people who feel the same emotion. Are these instructions clear? You can create as many piles as you need to. This is not timed, so feel free to take as long as you need. You can also change the piles whenever you wish, while you are sorting or at the end. Do you have any questions?"*

Participants were then allowed to sort the images into piles based on the instructions they were provided with. Once the participant sorted the images into distinct piles, the participant was then asked provide a word to describe the content of each pile. This was initially asked about each pile on the table in an open-ended manner (i.e., "What is in this pile?"). The experimenter then filled in a record sheet with any facial descriptors or emotions/behaviors/mental states that the participant produced. These responses were marked as "unprompted" because the experimenter did not question specifically about other content. This line of questioning allows for assessment of how the piles are spontaneously anchored by the participant. Next, the experimenter asked about facial behaviors (i.e., “What is on the face of these people?” or “What is the facial expression?” or “What is this person doing with their face that tells you about their emotion?”). The second and third phrasing of this prompt was only given if participants did not provide a response to the initial prompting. Finally, the experimenter asked about behavior/mental state or emotion (i.e., “*What (emotion/behavior/mental state) are these people (feeling/doing/experiencing)?”).* Any relevant response at this prompt would be marked as “prompted”.

Results

Data Coding. Data were coded in the same manner as Study 2.

Cluster Analysis. We performed a cluster analysis on the data from each instruction group separately to examine whether the clustering of items by emotions varied based on the instruction condition that participants were placed in. Data were summarized within a group as a co-occurrence matrix and subsequently converted into a dissimilarity matrix. An agglomerative average-linkage hierarchical cluster analysis (Sokal & Michener, 1958) was then conducted. The results (Figure 1) of the six cluster solutions are very consistent across the three instruction conditions (behavior, emotion, and mental state). Participants clustered facial expressions into the 6 discrete emotion categories in a manner consistent with the Western cultural model for emotion, regardless of condition. A six-cluster solution was reported on based on the increases in within cluster average item distance when the solution contained less than six clusters, (i.e., as items were grouped into larger, increasingly inclusive clusters, the clusters became less coherent; Figures 2-4), for theoretical reasons (since 6 discrete expression portrayals were included in the set), as well as to make the analyses maximally comparable to those conducted in Study 2.

Multidimensional Scaling. Consistent with the approach laid out in Study 2, in order to best represent the data without further imposing emotion categories onto participants sorting behavior as a means to quantify it, we employed multidimensional scaling (MDS). MDS allowed us to examine the underlying dimensions that structure sorting depending on instruction condition. This was accomplished by computing a co-occurrence matrix (Coxon, 1982) for each of the 3 groups (behavior, emotion, and mental state focused) and subjecting those matrices to MDS in ALSCAL. MDS provides the structure of a set of objects from data that approximate the distances between pairs of the objects. The points are arranged in space, defined by N dimensions, so that the distances between pairs of points in that space have the strongest possible relation to the similarities among the pairs of objects. A 4 dimensional solution was selected as the best fit across groups based on the elbow on the stress-by-dimensionality plot (Figure 5), although the 2 dimensional solutions (visualized in Figure 6) had sufficiently low stress to adequately capture the sorting data well. The scaling solutions resulting from these analyses can be treated in several different ways. In a first approach, we looked at the distances in multidimensional space between portrayals of the same emotion category (Figure 6). What we can see fairly clearly is very few systematic differences across groups, either based on visual inspection of the MDS solutions rendered in 2D space (Figure 6) or based on clustering of emotions in multidimensional space (Figure 7).

Additionally, we performed a set of hierarchical multiple regressions on the scaling solutions in order to examine the underlying dimensions themselves. Because interpretation of the dimensions yielded by multidimensional scaling can be difficult to accomplish completely subjectively (particularly for higher dimensionality solutions), we examined the dimensions for each solution in a data driven manner. As in Study 2, we used a set of normative ratings on each of the 36 images in the stimulus set. The set of 36 images were rated on a given attribute (14 total) by separate sets of 40 observers each on Amazon's Mechanical Turk (for a total of 560 participants). Attributes were behavioral descriptors: crying, frowning, laughing, looking, pouting, scowling, smelling, smiling and emotional descriptors: anger, fear, disgust, sadness, neutral, happiness. We then performed hierarchical multiple regressions on the dimension coordinates of each face stimulus (for each dimension). By using hierarchical MDS, we were able to enter the different types of stimulus attribute variables (behaviors, discrete emotions) as blocks and examined the extent to which a given "model" is redundant (i.e., did not add much predictive power for accounting for stimulus coordinates on a given dimension). We examined whether behavior and discrete emotions are redundant predictors of the solution dimensions or whether emotions predict unique variance over and above behavior. The results of the hierarchical multiple regressions are summarized in Table 1 and *F*-tests are reported in Tables 2-4, for each of the groups, behavior, emotion and mental state, respectively. Converging with the other analysis approaches, participants in the behavioral condition appeared to use discrete emotion categories over and above the behavior to structure the content of their sort. Across all three conditions, the first two dimensions were best explained by a discrete emotion model (over and above a behavioral model). Contrary to predictions, the emotion condition tended to engage discrete emotion sorting to a lesser extent than the behavioral and mental state conditions. That is, the third dimension of the solution for participants in the emotion condition was explained by behavior, whereas it was explained by discrete emotion for participants in the other two conditions.

Pile Label Analyses. We examined the words that participants freely offered to name their piles in order to examine whether participants differentially anchored their piles with behavioral versus mental state content, depending on the focus that their instructions emphasized. We found both a main effect of instruction condition, F (1, 73) = 84.975, $p < .001$, $\eta_p^2 = .700$, label type, F (2, 146) = 82.912, $p < .001$, $\eta_p^2 = .532$, and an interaction between the two factors, F (4, 146) = 25.040, $p < .001$, $\eta_p^2 = .407$ (Figure 8). Follow-up single factor ANOVAs on each label type separately revealed that the production of behavioral labels did not differ by instruction group, F (2, 73) = 1.286, p = .282. Both discrete emotion labels, F (2, 73) = 110.704, $p < .001$, and other affective/emotion terms, F (2, 73) = 22.566, $p < .001$, differed by instruction condition, however. Follow-up bonferroni corrected comparisons revealed that participants in the behavioral instruction condition generated less discrete emotion labels and other affective/emotion terms compared to participants in the emotion and mental state instruction condition (all p's < .001). Participants in the mental state and discrete emotion condition did not differ from one another in the number of labels produced for either content type (all p's > .05). These data indicate that individuals in the behavioral condition had difficulty using *behaviors* to structure their sort. Yet the less frequent use of emotion terms in the behavioral instruction condition indicates that those participants were attempting to comply with instructions to sort by behavior (and as a byproduct ignored mental state content).

Discussion

We found that attempts to manipulate participants' emphasis on behaviors versus mental states had little impact on participants' reliance on discrete emotion categories to make meaning of facial expressions. Specifically, across a number of analysis approaches we found evidence that US participants who were instructed to focus on the behaviors that faces were engaged in still tended to infer the mental states that caused those behaviors, and used those mental state categories to make sense of the similarities between facial expressions. This was reflected in the recovery of 6 Western discrete emotion categories in the cluster analyses. Further, a data driven analysis of the underlying dimensions structuring participants' cognitive maps (derived via multidimensional scaling) revealed that mental state (i.e., discrete emotion) ratings better characterized those dimensions than did simple behavioral ratings.

The present findings help to rule out an "instruction" based explanation of the cross-cultural differences observed in Study 2 (Chapter 3). If Himba participants merely received instructions that over-emphasized behavioral aspects of the faces rather than mental state (i.e., to engage in action perception rather than focus on mental states) but do not actually differ from US participants in any stable, substantive way in their perception of emotion, we should have replicated the Himba pattern of data in our behavioral sorting condition. Because this study failed to yield the "Himba pattern" of results when US participants were instructed to sort by behavior, these findings instead suggest that the patterns observed in Study 2 (Chapter 3) reflect a cultural difference between the US and Himba samples tested.

There may be several main drivers of the cultural differences observed between the Himba and US, beyond a simple instructional difference that this study helps to rule out. One intriguing possibility is that the concept of emotion in Himba culture has less emphasis on mental/internal states, such that the differences between Himba and US perceivers were not due to a failure of the translation process, but due to a deeper cultural difference in the concept itself. Considerable cross-cultural differences in individual concepts for emotions have been noted (Russell, 1991; Wierzbicka, 1992). Further, differences appear to exist between cultures in the concept of *emotion* itself. For example, the English language definition of emotion has a heavy emphasis on mental state content. Webster's defines emotion as "a conscious mental reaction (as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body". While perhaps psychologists would not agree with this entire definition, the emphasis on the "internal" and "mental" is heavily emphasized in scientific definitions of emotion as well. In contrast, the Japanese concept of emotion emphasizes dimensions of social relatedness and is thus constituted by an interaction between people that cannot be located within an individual (Mesquita, 2007). Although relatively little is known about the concept of emotion in Himba culture from a psychological (or even folk) standpoint, it is possible that the term implies relatively less mental (and perhaps individual) content than the individualistic concept of emotion in the United States. As a result, the cultural differences observed between our two samples may be due to differences in the meaning of the term emotion itself.

Further, the finding that Himba perceivers are less likely to engage in mental state attribution compared to US perceivers is counterintuitive given evidence that the more ambiguous, unfamiliar, or unpredictable a situation is, the more likely it is that individuals will engage in mental state inference (Waytz, et al. 2010). It is likely that the experimental task and stimuli met these situational criteria for Himba perceivers—yet the Himba data indicate that they were less likely than US participants to engage in mental state inference. An interesting future direction would be to examine whether situational manipulations impact mental state inference in the same manner in Himba culture. One possibility is that, for the Himba, using behavioral categories actually satisfies the goals of mental state inference. That is, within a Western cultural context, we infer mental states in order to understand current behavior, predict what will happen next, and tune our own behavior appropriately as the situation unfolds, all of which may be in the service of fostering a social connection with another individual (Epley, Waytz, & Cacioppo, 2007). But to the extent that emotion (and other mental state) categories are not special "kinds" of categories given by biology (Barrett, 2009), it may be that other types of categories in other cultures service similar goals to mental state categories in the United States.

The present findings also weigh in on the interesting question of whether mental state inference is as automatic in US culture as psychologists have assumed. The present data align with the assumption that engaging in mental state inference does appear to be relatively relentless in US culture. Despite instructions to engage in action perception, US perceivers in our sample still tended to infer mental states behind those actions. That is, mental state inference was persistent in the sample of participants that we tested. These data are consistent with a host of neuroimaging data suggesting that the affective or emotional aspects of faces are not completely "ignored" when the task is to engage in some other judgment like gender identification—indeed, these tasks are often referred to as "implicit" emotion perception tasks.

The present study is not without limitations. While tendency to engage in mental state inference seems relatively inflexible in US culture, this may be due to a general disregard for instructions on the part of US participants compared to Himba participants. Our data cannot rule this potential explanation out entirely, but the pile labeling data we collected do suggest that the US participants were attending to the instructions. That is, individuals in the behavioral sorting condition spontaneously produced less mental state labels (although they did not produce more behavioral labels), consistent with the interpretation that they understood they should be sorting by something other than mental states. In a sense, the labeling data appear to reflect a demand characteristic, which would not have resulted is the participants simply disregarded the instructions. A second limitation is that malleability in performance in the sort task based on the phrasing of instructions may itself be culturally relative. Thus we cannot rule out the possibility that manipulating instructions for participants in Himba culture might have yielded larger between group differences.

References

Baron-Cohen, S. (1995). *Mindblindness : an essay on autism and theory of mind*. Cambridge, Mass.: MIT Press.

Barrett, L. F. (2009). The Future of Psychology: Connecting Mind to Brain. *Perspectives on Psychological Science, 4*(4), 326-339.

Bruner, J. S., & Taguiri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 634-654). Reading, MA: Addison-Wesley.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864-886. doi: Doi 10.1037/0033-295x.114.4.864

Gray, K., Jenkins, A. C., Heberlein, A. S., & Wegner, D. M. (2011). Distortions of mind perception in psychopathology. *Proceedings of the National Academy of Sciences of the United States of America, 108*(2), 477-479. doi: Doi 10.1073/Pnas.1015493108

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430. doi: 10.1126/science.1063736

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 of the United States of America, 109*(19), 7241-7244. doi: Doi 10.1073/Pnas.1200155109

Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *J Pers Soc Psychol, 90*(4), 543-555. doi: 10.1037/0022-3514.90.4.543

Mesquita, B. (2007). Emotions are culturally situated. *Social Science Information Sur Les Sciences Sociales, 46*(3), 410-415. doi: Doi 10.1177/05390184070460030107

Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1521), 1309-1316. doi: Doi 10.1098/Rstb.2008.0318

Russell, J. A. (1991). Culture and the Categorization of Emotions. *Psychol Bull, 110*(3), 426-450. doi: Doi 10.1037/0033-2909.110.3.426

Schyns, P. G., Petro, L. S., & Smith, M. L. (2009). Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: behavioral and brain evidence. *PLoS One, 4*(5), e5625. doi: 10.1371/journal.pone.0005625

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *The University of Kansas Scientific Bulletin, 38*, 1409-1438.

Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *Neuroimage, 59*(3), 3050-3059. doi: 10.1016/j.neuroimage.2011.10.005

Wagner, D. D., Kelley, W. M., & Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cereb Cortex, 21*(12), 2788-2796. doi: 10.1093/cercor/bhr074

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends Cogn Sci, 14*(8), 383-388. doi: Doi 10.1016/J.Tics.2010.05.006

Wierzbicka, A. (1992). Defining Emotion Concepts. *Cognitive Science, 16*(4), 539-581. doi: Doi 10.1207/S15516709cog1604_4

Zaki, J., Hennigan, K., Weber, J., & Ochsner, K. N. (2010). Social cognitive conflict resolution: contributions of domain-general and domain-specific neural systems. *J Neurosci, 30*(25), 8481-8488. doi: 10.1523/JNEUROSCI.0382-10.2010

Footnotes

[i] Schyns and colleagues have recently published work (Jack, Garrod, Yu, Caldara, & Schyns, 2012) indicating that facial expressions of emotion are culturally relative, suggesting that they may no longer support this claim.

Table 1. Multidimensional Scaling (MDS) Dimensions Identified using Hierarchical Multiple Regressions.

	Dimension 1		Dimension 2		Dimension 3		Dimension 4	
Sort Solution	Behavior	Discrete	Behavior	Discrete	Behavior	Discrete	Behavior	Discrete
Behavior	**.869*** **	**.091*** **	**.906*** **	**.060** **	**.824*** **	**.093** **	**.745*** **	.101
Emotion	**.909*** **	**.054** **	**.896*** **	**.080*** **	**.779*** **	.053	**.708*** **	.055
Mental State	**.826*** **	**.124*** **	**.916*** **	**.039** **	**.821*** **	**.103*** **	**.821*** **	.068

Note. The first block of regressors was comprised of eight sets of ratings where the 36 face stimuli were rated for the extent to which they depicted specific behaviors (e.g., smelling, crying); R^2 values are presented in the Behavior column under each dimension. The second block of regressors was comprised of six sets of ratings where each of the 36 faces was rated for the extent to which they represented discrete emotions (e.g., disgust, sadness); the R^2 increase is presented under each dimension labeled Discrete. The change in R^2 reflects the additional variability accounted for by the discrete emotion predictors over and above behavioral predictors. Thus dimensions for which the change in R^2 is significant reflect discrete emotion information, rather than strictly behavioral information. Significant effects are bolded; * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$.

Table 2. Dimension Identification: Behavioral Focus Sorting

		Model 1 (Behavior)			Model 2 (Behavior + Emotion)		
Dimension	Variable	*B*	*SE B*	β	*B*	*SE B*	β
1	Crying	0.211	0.245	0.115	-0.302	0.268	-0.165
	Frowning	1.382	0.41	1.096**	-0.098	0.403	-0.078
	Laughing	-0.691	0.809	-0.485	-0.615	0.561	-0.432
	Looking	0.889	0.237	0.338***	-0.197	0.329	-0.075
	Pouting	-0.945	0.289	-0.686**	-0.097	0.333	-0.07
	Scowling	-0.752	0.322	-0.58*	-0.781	0.363	-0.602*
	Smelling	1.402	0.26	0.585***	0.134	0.373	0.056
	Smiling	0.973	0.596	0.951	0.554	0.538	0.541
	Anger				-0.169	0.307	-0.122
	Sad				0.038	0.267	0.028
	Disgust				0.892	0.376	0.689*
	Fear				-0.14	0.226	-0.118
	Neutral				-0.958	0.197	-0.836***
	Happiness				-0.022	0.714	-0.02
	R^2		.869			.960	
	F for change in R^2		22.241***			7.969***	
2	Crying	-0.319	0.201	-0.18	0.066	0.24	0.037
	Frowning	-0.66	0.336	-0.542	0.434	0.361	0.356
	Laughing	0.604	0.663	0.439	-0.148	0.503	-0.107
	Looking	-0.526	0.194	-0.207*	-0.15	0.295	-0.059

	Pouting	0.244	0.237	0.184	-0.223	0.299	-0.168
	Scowling	0.338	0.263	0.27	0.127	0.326	0.102
	Smelling	-0.349	0.213	-0.151	-0.118	0.334	-0.051
	Smiling	0.29	0.489	0.294	-0.804	0.482	-0.814
	Anger				0.239	0.275	0.179
	Sad				-0.071	0.239	-0.053
	Disgust				-0.052	0.337	-0.042
	Fear				0.089	0.203	0.077
	Neutral				0.365	0.177	0.329
	Happiness				2.339	0.639	2.194***
	R^2		.906			.966	
	F for change in R^2		32.349***			6.097**	
3	Crying	-0.083	0.204	-0.063	-0.459	0.276	-0.35
	Frowning	-0.057	0.341	-0.063	-0.301	0.416	-0.334
	Laughing	-0.341	0.671	-0.334	0.347	0.579	0.34
	Looking	0.969	0.197	0.515***	0.528	0.339	0.281
	Pouting	-0.481	0.24	-0.488	-0.494	0.344	-0.501
	Scowling	-0.493	0.267	-0.531	-0.027	0.375	-0.03
	Smelling	-0.075	0.216	-0.044	0.432	0.385	0.252
	Smiling	-0.443	0.495	-0.605	0.832	0.555	1.136
	Anger				0.449	0.317	0.454
	Sad				0.313	0.275	0.317
	Disgust				-0.514	0.388	-0.555

	Fear				0.743	0.233	0.869**
	Neutral				0.457	0.203	0.557*
	Happiness				-1.277	0.736	-1.616
	R^2		.771			.861	
	F for change in R^2		15.749***			3.926**	
4	Crying	-0.341	0.204	-0.311	0.172	0.314	0.157
	Frowning	-0.055	0.342	-0.073	0.41	0.472	0.543
	Laughing	-0.757	0.674	-0.887	-0.702	0.657	-0.823
	Looking	0.371	0.197	0.236	0.5	0.385	0.318
	Pouting	-0.295	0.241	-0.358	-0.472	0.39	-0.574
	Scowling	0.863	0.268	1.112**	0.836	0.425	1.078
	Smelling	-0.06	0.217	-0.042	0.24	0.437	0.168
	Smiling	0.695	0.497	1.136	0.464	0.63	0.758
	Anger				-0.375	0.36	-0.453
	Sad				-0.676	0.312	-0.818*
	Disgust				-0.564	0.44	-0.729
	Fear				-0.297	0.265	-0.416
	Neutral				-0.272	0.231	-0.396
	Happiness				-0.513	0.835	-0.777
	R^2		.745			.847	
	F for change in R^2		9.876***			2.317	

Note. Hierarchical multiple regressions were conducted to identify the derived dimensions from the MDS analyses. Predictors were entered in blocks: Model 1 contained only behavior vectors as predictors; Model 2 contained both behavior and emotion vectors. Coefficients for each predictor within each model are presented as well as the relevant F-test for a significant change in prediction (R^2)

between models. A separate regression was conducted for each dimension that resulted from the Multidimensional scaling of sorting data. * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$

Table 3. Dimension Identification: Emotion Focus Sorting

		Model 1 (Behavior)			Model 2 (Behavior + Emotion)		
Dimension	Variable	*B*	*SE B*	β	*B*	*SE B*	β
1	Crying	-0.118	0.203	-0.065	0.201	0.257	0.11
	Frowning	-0.888	0.34	-0.709*	-0.098	0.386	-0.078
	Laughing	0.302	0.669	0.213	0.598	0.538	0.422
	Looking	-0.776	0.196	-0.297***	0.14	0.315	0.054
	Pouting	0.549	0.239	0.402*	0.079	0.319	0.058
	Scowling	0.371	0.266	0.288	0.498	0.348	0.386
	Smelling	-1.053	0.215	-0.442***	-0.106	0.357	-0.045
	Smiling	-0.992	0.493	-0.976	-0.053	0.516	-0.052
	Anger				0.108	0.294	0.079
	Sad				0.029	0.256	0.021
	Disgust				-0.581	0.36	-0.452
	Fear				0.189	0.217	0.159
	Neutral				0.895	0.189	0.786***
	Happiness				-0.889	0.684	-0.81
	R^2		.909			.963	
	F for change in R^2		33.711***			5.045**	
2	Crying	0.293	0.208	0.167	-0.332	0.196	-0.19
	Frowning	0.994	0.349	0.825**	-0.275	0.295	-0.228
	Laughing	-0.374	0.688	-0.275	0.11	0.411	0.081
	Looking	0.986	0.201	0.393***	-0.138	0.241	-0.055

	Pouting	-0.304	0.245	-0.231	0.345	0.244	0.262
	Scowling	-0.457	0.273	-0.369	-0.337	0.266	-0.272
	Smelling	0.517	0.221	0.226*	-0.133	0.273	-0.058
	Smiling	-0.193	0.507	-0.198	0.262	0.395	0.268
	Anger				0.225	0.225	0.17
	Sad				0.2	0.196	0.151
	Disgust				0.377	0.275	0.305
	Fear				0.301	0.166	0.264
	Neutral				-0.49	0.144	-0.448**
	Happiness				-0.853	0.523	-0.809
	R^2		.896			.961	
	F for change in R^2		29.139***			11.916***	
3	Crying	-0.401	0.237	-0.293	-0.156	0.409	-0.114
	Frowning	0.361	0.397	0.384	-0.298	0.615	-0.317
	Laughing	-0.789	0.783	-0.742	-0.485	0.856	-0.456
	Looking	0.995	0.229	0.507***	0.808	0.501	0.412
	Pouting	-0.107	0.279	-0.104	0.509	0.508	0.495
	Scowling	-1.189	0.311	-1.23**	-0.673	0.554	-0.696
	Smelling	0.039	0.252	0.022	-0.692	0.569	-0.387
	Smiling	-0.092	0.577	-0.12	-0.072	0.821	-0.095
	Anger				-0.278	0.469	-0.27
	Sad				-0.382	0.407	-0.372
	Disgust				0.609	0.573	0.631

	Fear				0.415	0.345	0.466
	Neutral				0.187	0.301	0.219
	Happiness				0.162	1.088	0.197
	R^2		.779			.833	
	F for change in R^2		11.907***			1.116	
4	Crying	-0.374	0.215	-0.347	-0.041	0.384	-0.038
	Frowning	-0.103	0.36	-0.138	0.315	0.577	0.424
	Laughing	0.002	0.71	0.003	0.042	0.804	0.05
	Looking	0.428	0.208	0.277*	0.374	0.471	0.242
	Pouting	-0.246	0.253	-0.304	-0.404	0.477	-0.499
	Scowling	0.852	0.282	1.117**	0.81	0.521	1.062
	Smelling	-0.358	0.228	-0.254	-0.072	0.534	-0.051
	Smiling	0.098	0.524	0.163	-0.073	0.771	-0.121
	Anger				-0.068	0.44	-0.084
	Sad				-0.457	0.382	-0.562
	Disgust				-0.468	0.538	-0.614
	Fear				-0.042	0.324	-0.06
	Neutral				-0.11	0.282	-0.163
	Happiness				-0.143	1.022	-0.22
	R^2		.708			.763	
	F for change in R^2		8.182***			.810	

Note. Hierarchical multiple regressions were conducted to identify the derived dimensions from the MDS analyses. Predictors were entered in blocks: Model 1 contained only behavior vectors as predictors; Model 2 contained both behavior and emotion vectors. Coefficients for each predictor within each model are presented as well as the relevant F-test for a significant change in prediction (R^2)

between models. A separate regression was conducted for each dimension that resulted from the Multidimensional scaling of sorting data. * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$

Table S3. Dimension Identification: Mental State Focus Sorting

		Model 1 (Behavior)			Model 2 (Behavior + Emotion)		
Dimension	Variable	*B*	*SE B*	β	*B*	*SE B*	β
1	Crying	0.424	0.275	0.238	-0.354	0.292	-0.199
	Frowning	1.125	0.46	0.918*	-0.185	0.439	-0.151
	Laughing	-0.363	0.906	-0.262	-0.591	0.611	-0.426
	Looking	0.765	0.265	0.299**	-0.445	0.358	-0.174
	Pouting	-0.723	0.323	-0.54*	-0.161	0.363	-0.12
	Scowling	-0.36	0.36	-0.286	-0.492	0.396	-0.39
	Smelling	1.088	0.291	0.467**	0	0.406	0
	Smiling	0.916	0.668	0.92	0.205	0.586	0.206
	Anger				-0.168	0.334	-0.125
	Sad				0.333	0.29	0.248
	Disgust				0.638	0.409	0.507
	Fear				-0.385	0.246	-0.331
	Neutral				-1.294	0.215	-1.16***
	Happiness				0.246	0.777	0.229
	R^2		.826			.950	
	F for change in R^2		16.018***			8.644***	
2	Crying	-0.042	0.182	-0.025	0.638	0.264	0.376*
	Frowning	-1.099	0.305	-0.94***	-0.531	0.398	-0.454
	Laughing	0.172	0.601	0.13	0.107	0.554	0.081
	Looking	-0.339	0.176	-0.139	0.612	0.324	0.251

	Pouting	0.15	0.214	0.117	-0.016	0.329	-0.012
	Scowling	0.641	0.239	0.533*	0.716	0.359	0.596
	Smelling	-0.101	0.193	-0.045	0.094	0.368	0.042
	Smiling	0.467	0.443	0.492	0.359	0.531	0.378
	Anger				-0.508	0.303	-0.396
	Sad				-0.475	0.263	-0.371
	Disgust				-0.062	0.371	-0.051
	Fear				-0.287	0.223	-0.259
	Neutral				0.339	0.194	0.319
	Happiness				-0.079	0.704	-0.078
	R^2		.916			.955	
	F for change in R^2		36.765***			2.995*	
3	Crying	-0.11	0.229	-0.075	0.254	0.295	0.173
	Frowning	0.04	0.383	0.04	-0.983	0.443	-0.974*
	Laughing	-0.449	0.755	-0.393	0.434	0.617	0.38
	Looking	1.241	0.221	0.59***	0.874	0.362	0.416*
	Pouting	-0.527	0.27	-0.478	0.458	0.367	0.416
	Scowling	-0.773	0.3	-0.745*	-0.254	0.4	-0.245
	Smelling	0.578	0.243	0.301*	-0.237	0.41	-0.124
	Smiling	-0.453	0.557	-0.553	0.245	0.592	0.299
	Anger				-0.068	0.338	-0.061
	Sad				-0.677	0.293	-0.613*
	Disgust				0.716	0.413	0.691

	Fear				0.754	0.249	0.788**
	Neutral				0.442	0.217	0.482
	Happiness				-0.807	0.785	-0.914
	R^2		.821			.924	
	F for change in R^2		15.520***			4.762**	
4	Crying	-0.051	0.171	-0.047	0.422	0.266	0.386
	Frowning	-0.555	0.286	-0.738	-0.261	0.4	-0.348
	Laughing	0.042	0.563	0.05	0.036	0.557	0.043
	Looking	0.312	0.165	0.199	0.255	0.326	0.163
	Pouting	-0.203	0.201	-0.247	-0.124	0.331	-0.152
	Scowling	1.242	0.224	1.607***	0.928	0.361	1.201*
	Smelling	-0.02	0.181	-0.014	-0.07	0.37	-0.049
	Smiling	0.078	0.415	0.128	-0.347	0.534	-0.569
	Anger				0.019	0.305	0.024
	Sad				-0.645	0.265	-0.784*
	Disgust				-0.063	0.373	-0.082
	Fear				-0.087	0.225	-0.122
	Neutral				-0.12	0.196	-0.175
	Happiness				0.213	0.709	0.324
	R^2		.821			.889	
	F for change in R^2		15.475***			2.139	

Note. Hierarchical multiple regressions were conducted to identify the derived dimensions from the MDS analyses. Predictors were entered in blocks: Model 1 contained only behavior vectors as predictors; Model 2 contained both behavior and emotion vectors. Coefficients for each predictor within each model are presented as well as the relevant F-test for a significant change in prediction (R^2)

between models. A separate regression was conducted for each dimension that resulted from the Multidimensional scaling of sorting data. * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$

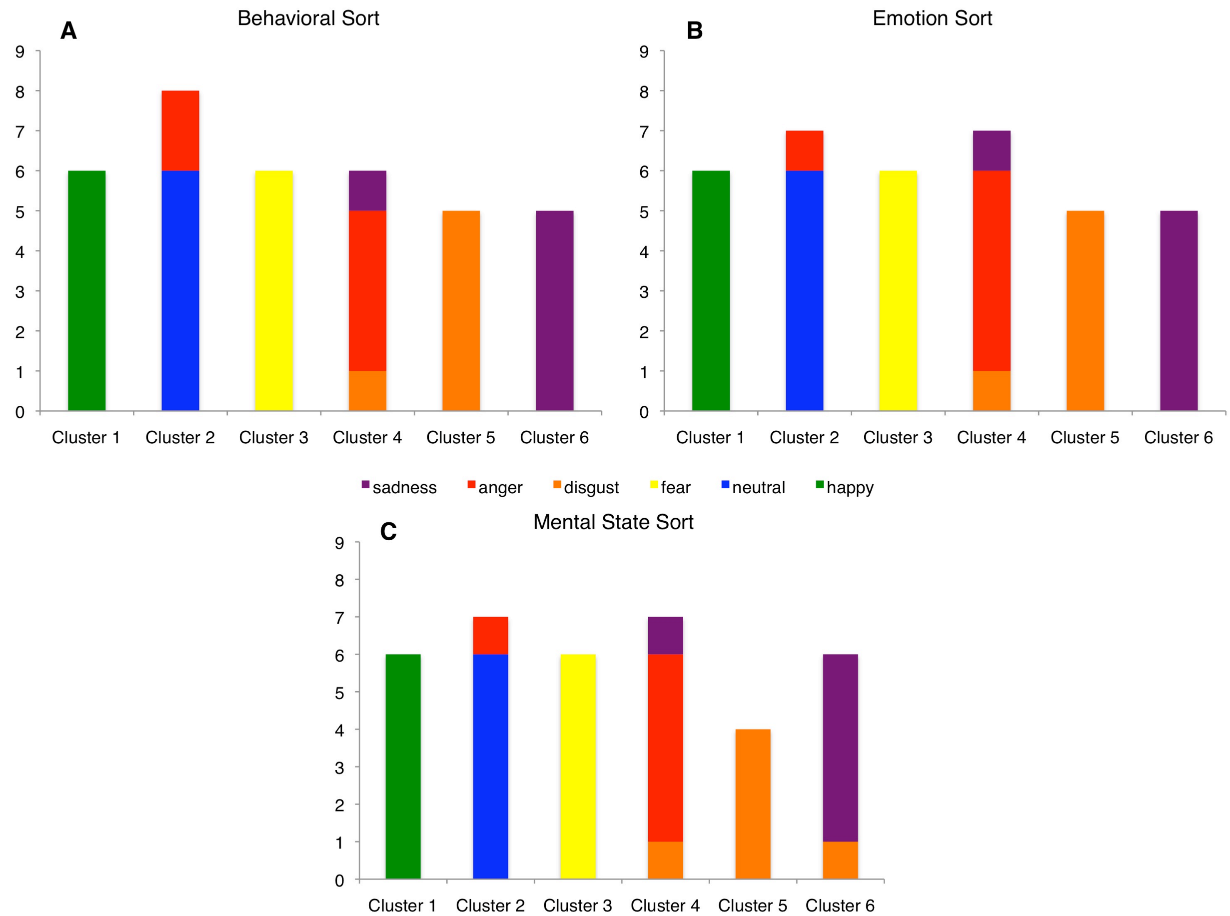


Fig 1. The results of the six-cluster solutions from the hierarchical cluster analyses are plotted in A-C with cluster on the x-axis. The y-axis represents the number of items grouped into a given cluster, with contents stacked by the emotion portrayed in each posed facial expression. Stacked bars containing several different colors indicate that faces portraying different discrete emotions were clustered together. Bars with a single color (or predominance of a given color) indicate relatively clean clustering of faces depicting one emotion category. The behavioral sort (**A**) cluster solution contains very consistent clustering with the solutions obtained for emotion (**B**) and mental state sorts (**C**).

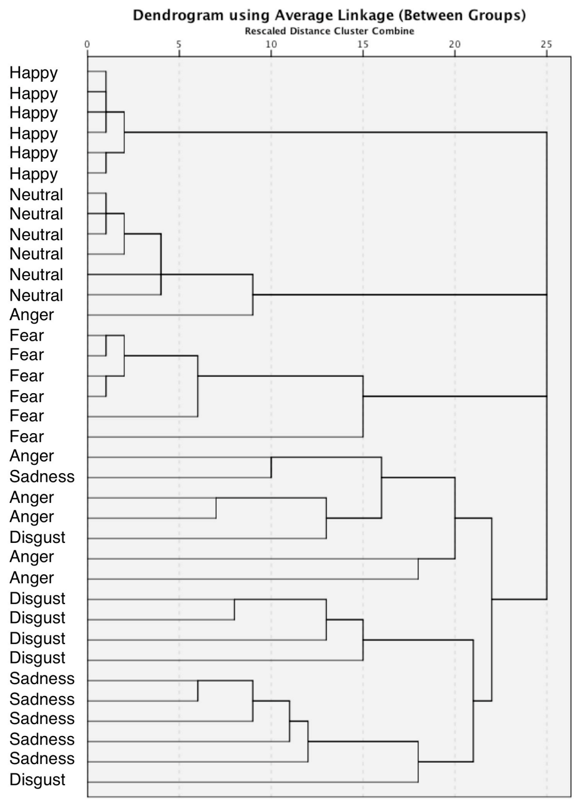


Figure 2. Cluster analysis dendrogram for participants in the behavioral focus condition. As an increasing number of items are included in a cluster, the more inclusive cluster also increases the average distance within the cluster. As each cluster (which starts out as a single item) is united with another cluster (i.e., agglomerated), this is indicated in the dendrogram by a bracket. The longer the bracket between two clusters (i.e., the bigger the distance before two clusters are joined), the more spread out the newly formed cluster is (i.e., the bigger the difference between the items joining the new cluster).

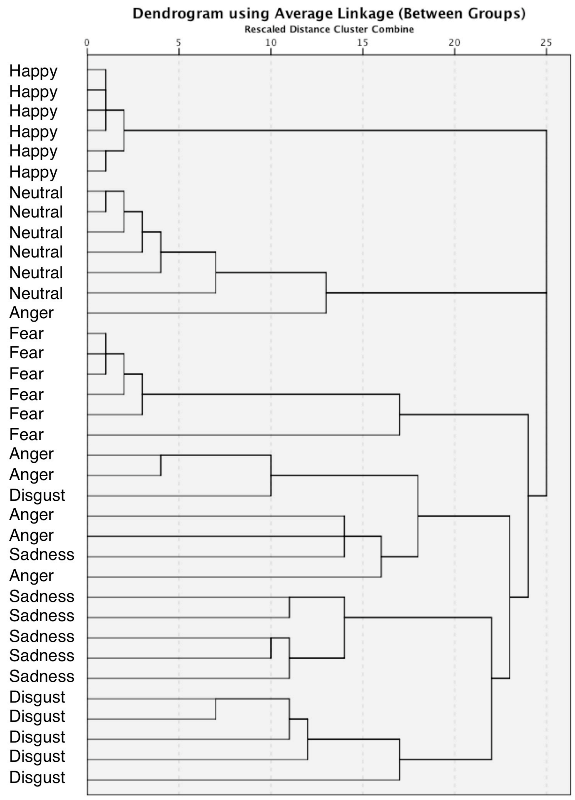


Figure 3. Cluster analysis dendrogram for participants in the emotion focus condition. As an increasing number of items are included in a cluster, the more inclusive cluster also increases the average distance within the cluster. As each cluster (which starts out as a single item) is united with another cluster (i.e., agglomerated), this is indicated in the dendrogram by a bracket. The longer the bracket between two clusters (i.e., the bigger the distance before two clusters are joined), the more spread out the newly formed cluster is (i.e., the bigger the difference between the items joining the new cluster).

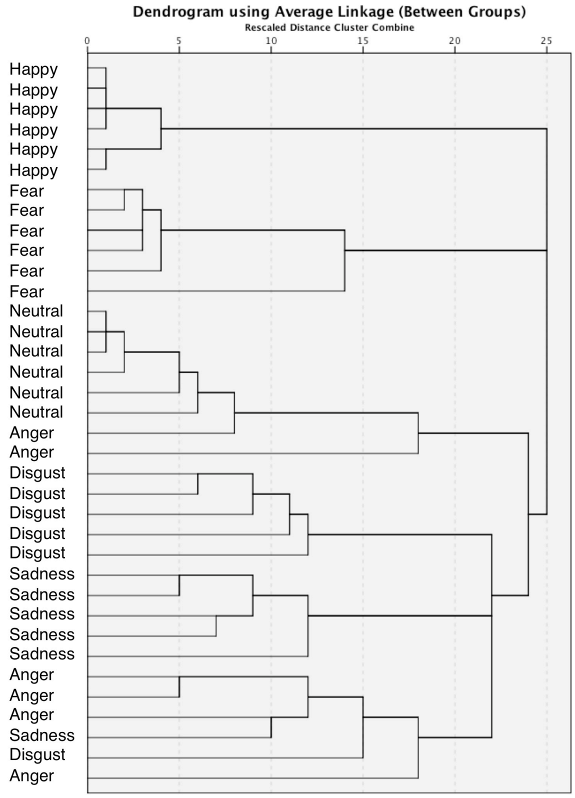


Figure 4. Cluster analysis dendrogram for participants in the mental state focus condition. As an increasing number of items are included in a cluster, the more inclusive cluster also increases the average distance within the cluster. As each cluster (which starts out as a single item) is united with another cluster (i.e., agglomerated), this is indicated in the dendrogram by a bracket. The longer the bracket between two clusters (i.e., the bigger the distance before two clusters are joined), the more spread out the newly formed cluster is (i.e., the bigger the difference between the items joining the new cluster).

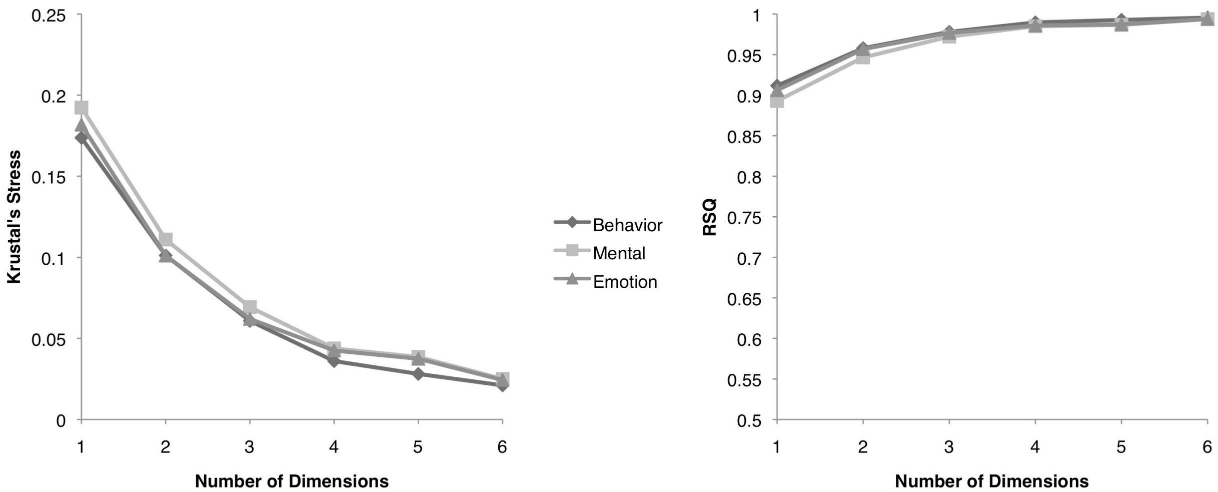


Fig. 5. Stress-by-dimensionality (left) and RSQ-by-dimensionality (right) plots for multidimensional scaling solutions of 1-6 dimensionality for three separate instruction groups: behavior, mental state, and emotion. The number of dimensions in a given solution is plotted on the x-axis. A sharp elbow can be observed in all three conditions on the stress-by-dimensionality plot at 4 dimensions, indicating that additional dimensions beyond 4 do not aid in representing the structure of the sorting data.

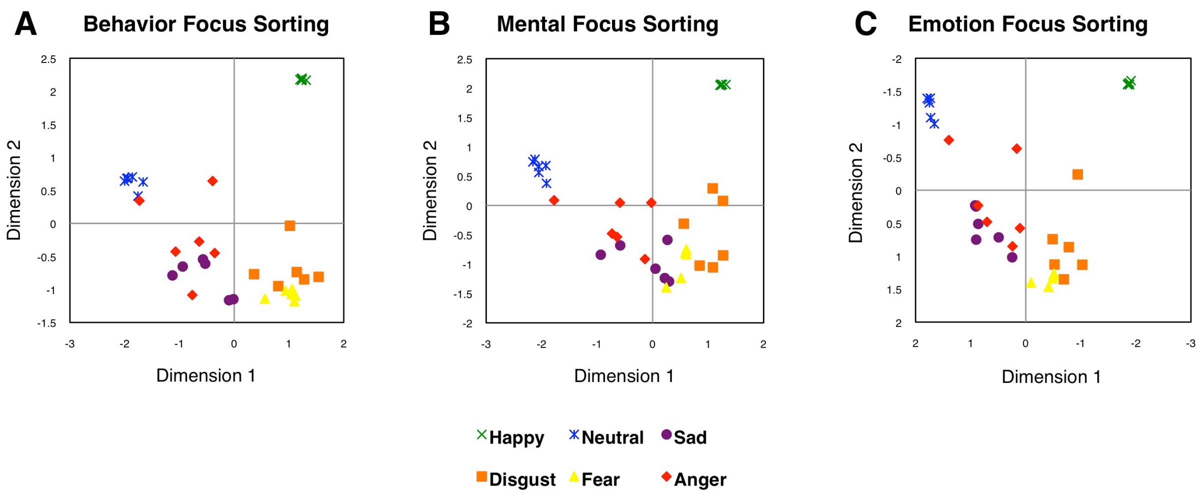


Fig. 6. MDS derived configurations plotted in two-dimensional space for US participants instructed to sort based on behaviors (**A**) mental states (**B**) or emotions (**C**). Items are plotted by discrete emotion portrayal type. Faces that were typically sorted together are closer together in space. Clustering by discrete emotion is clearly evident and few differences emerge across group solutions. One notable shift, however, is the shift in neutral portrayals toward happy portrayals on dimension 2, specifically for the emotion focus condition.

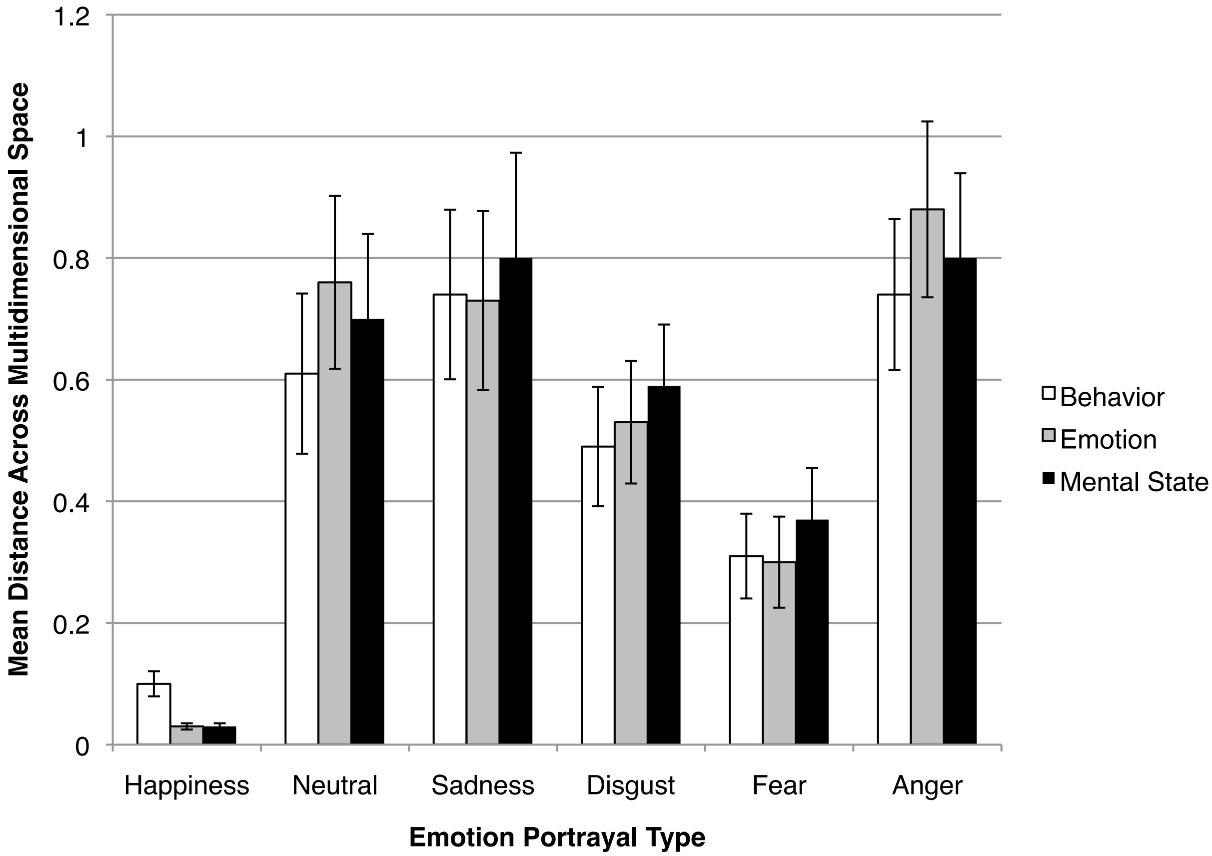


Fig. 7. Mean distance between faces portraying a given emotion across multidimensional space. We quantified the clustering of items within a category across all dimensions and plotted these mean distances (+/- standard error) for each instruction group.

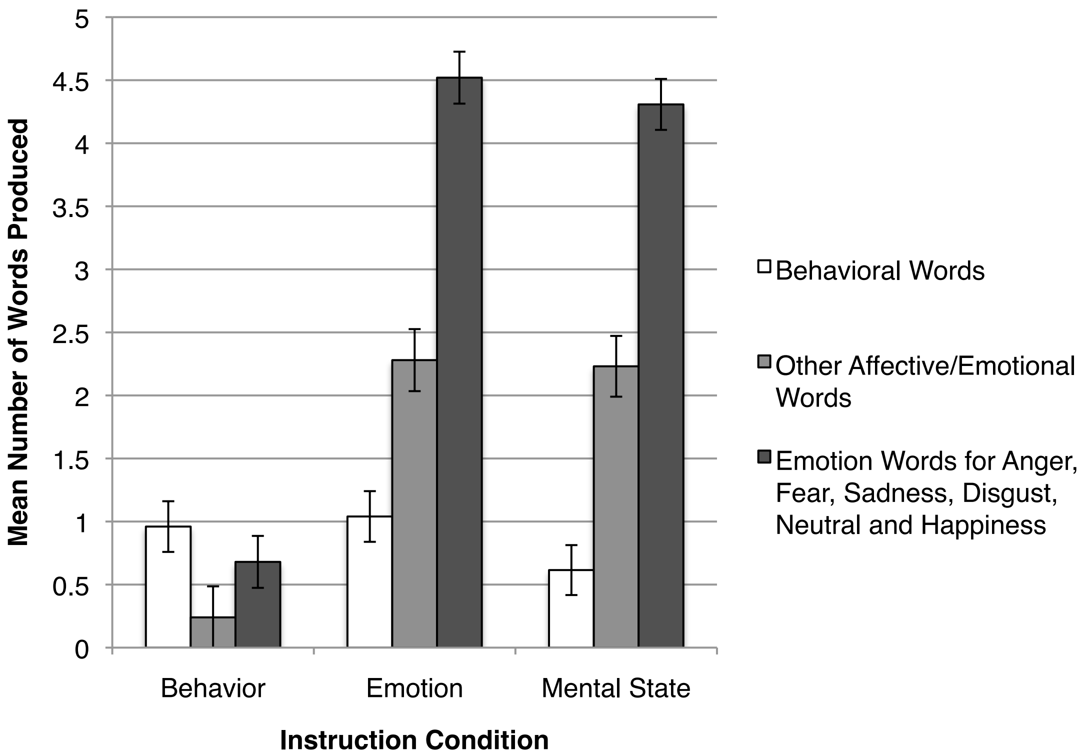


Fig. 8. Pile labels used by instruction condition: Behavior, Emotion or Mental State. Mean number of words produced by each group (± standard error) is plotted on the y-axis broken down by word type.

Acknowledgements

I would like to thank the community of individuals who have inspired, taught, and supported me throughout the dissertation process. First, I would like to thank Lisa Feldman Barrett for the many years of guidance, encouragement, and a lively intellectual exchange. I would like to thank Lisa for providing me with the opportunity to travel to Namibia to conduct this research. The research trip I took to collect these data involved many unknowns, not unlike the dissertation process itself. It took a mutual leap of faith for us to embark on this research project. The trip would not have been possible without Debi Roberson, our collaborator on this work, who was the logistical mastermind behind the trip's success. We benefitted greatly from Debi's sage advice and years of experience in cross-cultural research, not to mention the wonderful company that Debi provided out in the field. Further, I would like to thank Marietta van der Vyver, our collaborator in Namibia who greatly facilitated our research program there. The final member of our cross-cultural research team, our translator, Kemmu Jakara, recently passed away. His work as guide on our trip was deeply valued and we owe a debt of gratitude to him for his translation work.

I would also like to thank my dissertation committee for their valuable guidance during this dissertation process, but also throughout my graduate years. It was in Jim Russell's course on theories of emotion during my first year of graduate school that I was first exposed to many of the ideas tested in this dissertation. Jim has also provided many key insights through his own publications and through feedback supplied on this dissertation. I would like to thank Hiram Brownell for his meticulous feedback and insightful suggestions regarding the statistical analyses in this dissertation. Finally I would like to thank Bajta Mesquita for serving as the external reviewer on this dissertation. Bajta's work on cultural variability in emotions is key to our understanding of emotions as culturally constrained phenomena, a lineage of work that this dissertation aims to contribute to.

I would like to thank my research assistants who spent countless, and no doubt painful, hours helping me with the entry, checking and coding of the data contained in this dissertation. I would also like to acknowledge the amazing intellectual community that I find myself a part of in the Interdisciplinary Affective Science Lab (members past and present). I am grateful to my lab-mates not only for their constant intellectual stimulation and inspiration, but also for their humor and generosity.