Emotion

Context Facilitates Performance on a Classic Cross-Cultural Emotion Perception Task

Katie Hoemann, Alyssa N. Crittenden, Shani Msafiri, Qiang Liu, Chaojie Li, Debi Roberson, Gregory A. Ruark, Maria Gendron, and Lisa Feldman Barrett

Online First Publication, November 26, 2018. http://dx.doi.org/10.1037/emo0000501

American Psychological Association
Emotion

1528-3542/18/$12.00 http://dx.doi.org/10.1037/emo0000501

Context Facilitates Performance on a Classic Cross-Cultural Emotion Perception Task

Katie Hoemann
Northeastern University

Alyssa N. Crittenden
University of Nevada–Las Vegas

Shani Msafiri
Hadza cultural consultant, Arusha, Tanzania

Qiang Liu and Chaojie Li
Liaoning Normal University

Debi Roberson
University of Essex

Gregory A. Ruark
U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Belvoir, Virginia

Maria Gendron
Northeastern University

Lisa Feldman Barrett
Northeastern University; Massachusetts General Hospital, Boston, Massachusetts; and Martinos Center for Biomedical Imaging, Charlestown, Massachusetts

The majority of studies designed to assess cross-cultural emotion perception use a choice-from-array task in which participants are presented with brief emotion stories and asked to choose between target and foil cues. This task has been widely criticized, evoking a lively and prolonged debate about whether it inadvertently helps participants to perform better than they otherwise would, resulting in the appearance of universality. In 3 studies, we provide a strong test of the hypothesis that the classic choice-from-array task constitutes a potent source of context that shapes performance. Participants from a remote small-scale (the Hadza hunter-gatherers of Tanzania) and 2 urban industrialized (China and the United States) cultural samples selected target vocalizations that were contrived for 6 non-English, nonuniversal emotion categories at levels significantly above chance. In studies of *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*, above chance performance is interpreted as evidence of universality. These studies support the hypothesis that choice-from-array tasks encourage evidence for cross-cultural emotion perception. We discuss these findings with reference to the history of cross-cultural emotion perception studies, and suggest several processes that may, together, give rise to the appearance of universal emotions.

Keywords: emotion, perception, vocalizations, universality, culture

Supplemental materials: http://dx.doi.org/10.1037/emo0000501.supp

Katie Hoemann, Department of Psychology, Northeastern University. Alyssa N. Crittenden, Department of Anthropology, University of Nevada–Las Vegas. Shani Msafiri, Hadza cultural consultant, Arusha, Tanzania. Qiang Liu and Chaojie Li, Research Center of Brain and Cognitive Neuroscience, Liaoning Normal University. Debi Roberson, Department of Psychology, University of Essex. Gregory A. Ruark, Foundational Science Research Unit, U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Belvoir, Virginia. Maria Gendron, Department of Psychology, Northeastern University. Lisa Feldman Barrett, Department of Psychology, Northeastern University; Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts; and Martinos Center for Biomedical Imaging, Charlestown, Massachusetts.

This research was supported by a U.S. Army Research Institute for the Behavioral and Social Sciences grant (W911NF-15-1-0647) awarded to Maria Gendron and Lisa Feldman Barrett, as well as a NIH NRSA fellowship (5F32MH105052) awarded to Maria Gendron. This research was further supported by the National Natural Science Foundation of China (NSFC 31170982) and Royal Society (IE121122) International Cooperation and Exchanges NSFC (31311130123) awarded to Qiang Liu and Debi Roberson, and extended to Maria Gendron. The views, opinions, and/or findings contained in this article are those of the authors and shall not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documents. Katie Hoemann, Maria Gendron, and Lisa Feldman Barrett developed the study concept. Maria Gendron and Lisa Feldman Barrett shared senior authorship. Katie Hoemann and Maria Gendron developed the stimuli, and Alyssa N. Crittenden, Shani Msafiri, and Chaojie Li provided cultural review. Katie Hoemann, Maria Gendron, Shani Msafiri, and Chaojie Li collected the experimental data. Katie Hoemann, Maria Gendron, and Lisa Feldman Barrett analyzed the data. All authors contributed to the writing of the manuscript. Study 1 data were collected under the approval of the Tanzanian COSTECH. The authors are grateful to A. Mabulla, M. Peterson, L. Wang, and J. Zhang for research support, and the Hadza communities and Liaoning Normal University where this research was conducted.

Correspondence concerning this article should be addressed to Katie Hoemann, Department of Psychology, Northeastern University, 360 Huntington Avenue, Boston, MA 02115. E-mail: khoemann@gmail.com

There has been a lively and prolonged debate about whether or not certain emotion categories are universally expressed and recognized. According to the strongest version of the universality hypothesis, all humans (barring illness) innately produce and perceive *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise* in nonverbal behaviors, independent of cultural background and learning.[1] The original tests of this hypothesis were conducted in a handful of studies between 1969 and 1975 in two remote, small-scale societies, the Fore and Dani of Papua New Guinea (Ekman, 1972; Ekman & Friesen, 1971; Ekman, Sorenson, & Friesen, 1969; Sorenson, 1975; for a review, see Gendron, Crivelli, & Barrett, 2018). The studies used several versions of a choice-from-array task pioneered by Dashiell (1927): participants heard a brief story about an emotional episode with an emotion word embedded (e.g., "Her child has died and she feels very sad.") and were then asked to select the target stimulus (e.g., a photograph of a posed facial configuration of a frown) presented alongside one or two foils.[2] Participants chose the target stimulus more frequently than chance, leading to the conclusion that *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise* were universal emotions (on the assumption that people infer emotion in certain facial configurations only if those configurations express emotion with a certain degree of fidelity; for discussion, see Jack, Sun, Delis, Garrod, & Schyns, 2016). Since then, hundreds of studies have been published using a choice-from-array task with participants from various urban cultural contexts with some degree of exposure to western cultural practices and norms, such as Brazil, China, Estonia, Ethiopia, France, Greece, India, Indonesia, Israel, Japan, Kenya, Malaysia, Mexico, Singapore, Turkey, and Zambia (Elfenbein & Ambady, 2002; Laukka et al., 2013), providing further support for the universality hypothesis (for discussion, see Ekman, 2017). The strongest evidence continues to come from the original studies of emotion perception in the Fore and Dani of Papua New Guinea (Ekman, 1972; Ekman & Friesen, 1971; Ekman, Heider, Friesen, & Heider, 1972), as well as from choice-from-array tasks used with the Himba of Namibia (Sauter, Eisner, Ekman, & Scott, 2010) and Bhutanese villagers (Cordaro, Keltner, Tshering, Wangchuk, & Flynn, 2016), because participants in these studies had limited exposure to western cultural practices and norms, including media, minimizing alternative explanations for any cross-cultural consistencies that were observed (Norenzayan & Heine, 2005). In the present article, however, we present three studies to support the hypothesis that the choice-from-array task itself *creates* evidence for universal emotions. This evidence, we suggest, is not simply the result of demand characteristics or confirmatory bias. The task does not prime innate emotion knowledge. Instead, we suggest that choice-from-array tasks contain psychologically potent features that guide participants as they make emotional meaning of novel stimuli.

Choice-From-Array Tasks

There is now evidence from a variety of experimental studies that choice-from-array tasks provide a potent context that helps participants choose the stimuli that experimenters expect, creating stronger evidence for universal emotions than might occur otherwise (L. F. Barrett, Adolphs, Marsella, Martinez, & Pollak, in press; Russell, 1994; Russell, Bachorowski, & Fernandez-Dols, 2003). Numerous lab-based studies with both children and adults (L. F. Barrett et al., in press), as well as almost a dozen cross-cultural studies published since 2008, which sample a broader range of remote, small-scale populations (see Table 1 in the online supplemental material; also Gendron, Crivelli, et al., 2018) and ask people to infer emotional meaning in both facial poses and vocalizations, provide evidence that choice-from-array tasks are not psychologically inert. Instead, as reviewed next, these tasks can contain elements that are known to shape participants' responses (for a broader discussion of contextual factors that influence emotion perception, see Aviezer & Hassin, 2017; L. F. Barrett, Mesquita, & Gendron, 2011; de Gelder, 2016; Gendron, Mesquita, & Barrett, 2013; Hess & Hareli, 2017; Wieser & Brosch, 2012).[3]

Emotion Words

The emotion words used in choice-from-array tasks may encourage participants to assign emotional meaning to facial configurations and vocalizations in ways they would otherwise not. When emotion words are familiar (i.e., associated with known emotion concepts), they directly shape participants' perceptions, such that facial configurations like scowls and pouts are seen differently when words like "angry" and "sad" are present during the experiment versus when they are absent (Gendron, Lindquist, Barsalou, & Barrett, 2012). In such contexts, emotion words influence which facial configurations are predicted, encoded, and remembered (Chanes, Wormwood, Betz, & Barrett, 2018; Doyle & Lindquist, 2018; Fugate, Gendron, Nakashima, & Barrett, 2018; Fugate, Gouzoules, & Barrett, 2010). These findings suggest the interpretation that emotion words are aids that help perceivers recognize universal emotions more easily. Emotion words may assist perceivers in choosing the target stimulus from an array by strengthening their predictions or enhancing their sensitivity for the target or its perceptual features (Chanes et al., 2018; Mohanty & Sussman, 2013; Vogt, De Houwer, Moors, Van Damme, & Crombez, 2010). Other evidence suggests that the effect of emotion words is more potent than that, potentially *creating* emotion perceptions that would otherwise not occur. For example, participants label scowling faces as "determined" or "puzzled," wide-eyed faces as "hopeful" and gasping faces as "pained" when they are provided with stories about those emotions rather than with stories of *anger*, *surprise*, and *fear* (Carroll & Russell, 1996, Study 2). Scowling faces are also more likely to be perceived as fearful when paired with the description of danger (Carroll & Russell, 1996, Study 1).

Words are particularly potent when people are presented with cues, such as facial configurations, that are unfamiliar and have no prior emotional meaning (Fugate et al., 2010), consistent with the

[1] For the purposes of illustrating the theoretical origins of tests of cross-cultural emotion perception, we refer to the strongest, traditional version of the universality hypothesis, and the debate as one between total universality versus cultural relativism. However, more recent accounts have described universality as a graded phenomenon (e.g., Cordaro et al., 2018; Keltner & Haidt, 1999). This debate can alternatively be framed in terms of sources of cross-cultural (and intracultural) consistency versus diversity (Crivelli & Gendron, 2017a; Gendron, Crivelli, et al., 2018; Russell, 1995).

[2] An alternative version is to show photos of facial configurations, one at a time, along with a small set of emotion words. Participants are tasked with selecting the best matching emotion word for each photo.

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

broader finding that words support perception for unfamiliar objects (Lupyan, Rakison, & McClelland, 2007). Furthermore, experimental tasks that place fewer constraints on how participants respond, such as asking them to freely label facial configurations or vocalizations, rarely provide evidence for the universality hypothesis (see Table 1 in the online supplemental material; L. F. Barrett et al., in press; Gendron, 2017; Gendron, Crivelli, et al., 2018). Finally, children who do not explicitly possess conceptual knowledge about emotions such as *anger*, *sadness*, and *fear*, and who do not understand the meaning of emotion words beyond their affective content, as well as adults who have lost that semantic knowledge from neurodegeneration, are experientially blind to the emotional meaning of facial configurations and vocalizations: they only perceive affect (Lindquist, Gendron, Barrett, & Dickerson, 2014; Widen, 2016).

Elaborate Manipulation Checks

Choice-from-array tasks have been known to use an elaborate manipulation check procedure that has the potential to teach novel emotion concepts to naïve participants. For example, Sauter, Eisner, Calder, and Scott (2010) used what they referred to as a "rigorous" manipulation check that did more than ask for verbal confirmation of whether individuals from the Himba culture of northwestern Namibia understood *anger*, *sadness*, *fear*, and the other concepts represented in brief stories in the experiment. As stated in Sauter et al. (2010), after hearing each emotion story, participants were "asked how the person was *feeling*" (p. 2411, italics added) to confirm that "they had understood the intended emotion of the story" (p. 2408). In their later commentary, Sauter, Eisner, Ekman, and Scott (2015) elaborated that:

> Participants [were allowed] to listen several times to a given recorded story (if needed), until they could explain the intended emotion in their own words, before they proceeded to the experimental trials for that story. The inclusion of a rigorous manipulation check with experimenter verification, rather than reliance on participants' reports, was crucial. (p. 355)

Individuals in the Himba cultural group display opacity of mind, however; they do not make mental state inferences as frequently as more westernized people (H. C. Barrett et al., 2016), focusing on emotion as situated action rather than internal feeling.[4] As a consequence, when Sauter and colleagues asked participants to verbally describe the English emotion concept that was portrayed in the story, they perhaps (unwittingly) encouraged concept learning in their experiment (as hypothesized by Gendron, Roberson, & Barrett, 2015). They did not allow Himba participants to proceed to the experimental trials for a given story until they could conceptualize the emotion stories in a manner consistent with western cultural expectations.

By contrast, a subsequent study using a choice-from-array task with the Himba, but without the elaborate manipulation check, was unable to replicate the Sauter et al. (2010) findings (Gendron, Roberson, van der Vyver, & Barrett, 2014a). Our study used a manipulation check only to verify that participants understood how to perform the task. Participants were given the opportunity to indicate that they understood the emotion stories from their own cultural perspective. We took great pains to ensure participants were attending to the experimental task, understood the action described in the brief stories, and understood what was required of them: we conducted an attention check before every trial, verbally confirmed understanding (yes or no response), and allowed participants to replay the story for any reason. The concern that participants may have missed some of the stories, because of an attentional lapse or a failure of memory, is assuaged by that fact that the stories were repeated many times over the course of the experiment. Analyses of the different foil conditions revealed that Himba participants performed better than chance only when the target and foil differed in valence, such that pleasantness or unpleasantness could be used to distinguish between the two vocalizations. The general pattern of these findings was replicated in a free-labeling task (Gendron et al., 2014a, Study 1).

Repeated Trials

Choice-from-array tasks present the same stimuli over and over. The fact that participants are repeatedly exposed to the same facial configurations and emotion words creates a context for them to learn the intended pairings, even if they do not know them to begin with (Nelson & Russell, 2016a). For example, children learn to label an artificially constructed facial expression (e.g., a blowfish expression) with the word "pax" in a choice-from-array task (Nelson & Russell, 2016b). Additionally, participants were shown to use a process-of-elimination strategy to complete choice-from-array tasks, boosting agreement levels (DiGirolamo & Russell, 2017). In fact, if the target stimulus is presented against only one foil, all that is required in this task arrangement is to figure out which of two vocalizations or two faces is least expected in the given context. For example, after hearing a story about anger, a participant hears a shout and a laugh, and can choose the shout merely by realizing the laugh is not correct (on the basis of valence). This is similar to selecting the correct answer on a multiple-choice test by eliminating the incorrect alternatives.

Choice-from-array tasks have also been known to block trials by emotion category, which may further encourage perceptual learning and a process-of-elimination strategy. For example, Sauter et al. (2010) blocked their trials: a Himba participant heard a story, verbalized an English emotion concept, and then heard a series of trials with portrayed vocal bursts for the target emotion category and a foil. For example, after hearing a story about sadness several times and describing the English concept for *sadness*, a participant would hear a cry (target) and a laugh (foil), then a shout (foil) and a cry (target), and so on. When Himba participants completed a similar task with randomized trials, they did not choose vocalizations for *anger*, *sadness*, *fear*, and the other categories tested in a

[4] Anecdotal evidence is consistent with the idea that people in remote, small-scale villages many not understand emotion concepts in the same way as people from urban, western cultural contexts. Describing his first visit to the Fore tribe in New Guinea, Ekman (2007) wrote "I asked them to make up a story about each facial expression [photograph]. 'Tell me what is happening now, what happened before to make the person show this expression, and what is going to happen next.' It was like pulling teeth. I am not certain whether it was the translation process, or the fact that they have no idea what I wanted to hear or why I wanted them to do this. Perhaps making up stories about strangers was just something the Fore didn't do." (p. 7) However, it is also possible that they do not understand emotion as people from the United States do, and they may not have the same emotion concepts.

way that was supportive of the universality hypothesis (Gendron et al., 2014a).

Meaning Making

It is tempting to assume that these design considerations are mere methodological footnotes, but we suggest an alternative hypothesis. Taken together, these findings suggest that the design elements in a choice-from-array task are psychologically potent: they provide a context for perceivers to make meaning of ambiguous physical cues like facial configurations and vocalizations, particularly when those cues have no inherent emotional meaning. As demonstrated by a recent review, hundreds of studies of healthy adults across cultures, newborns and young children, and people who are congenitally blind or deaf consistently find that specific facial movements, such as smiling, scowling, and frowning are context-specific in their emotional meaning: instances of the same emotion category, such as instances of anger, are expressed with more variable facial movements than generally acknowledged, and similar facial movements, such as a scowl, can communicate a variety of different emotions, or even carry nonemotional information (L. F. Barrett et al., in press). As a result, perceivers implicitly use situational context to make meaning and infer emotional information in facial configurations. The same is likely true of vocalizations (Russell et al., 2003). As a result, choice-from-array tasks may provide a context that allows participants to infer emotional meaning differently than they otherwise would, shaping their emotion perception performance. If this hypothesis is correct, then it provides an alternative explanation for the hundreds of studies that support the hypothesis that people universally recognize emotion in nonverbal behaviors in an automatic and obligatory way.

The hypothesis that choice-from-array tasks facilitate emotion perception fits with the observation that humans are active meaning-makers. Human brains are wired to transduce changes in light, air pressure, and chemicals, and go beyond the information given (Bruner, 1957) to create the sights, sounds, and smells of our surrounds. The same changes in air pressure can be experienced as a person's laugh in one setting and a sob in another (Belin, Fillion-Bilodeau, & Gosselin, 2008). The psychological literature is full of experiments that manipulate context to examine its effect on meaning. For example, visual context facilitates the recognition of scene-consistent objects (for a review, see Bar, 2004). Furthermore, it is well known that the acoustic and linguistic context influences which phonemes are heard (Massaro & Cohen, 1983). There are notable examples of how certain experimental design features evoke powerful psychological effects that are hidden contexts for other psychological processes (such as controlled processing; for a review, see Pashler, Johnston, & Ruthruff, 2001). Context effects are well-established in how people perceive emotion in faces, voices, and bodies (e.g., Aviezer et al., 2008; Aviezer, Trope, & Todorov, 2012; Calbi, Angelini, Gallese, & Umiltà, 2017; Fernández-Dols & Ruiz-Belda, 1997; Fridlund, 1991; Mobbs et al., 2006; Ruiz-Belda, Fernández-Dols, Carrera, & Barchard, 2003; Van den Stock, Righart, & de Gelder, 2007; Wallbott, 1988).

Our hypothesis here is that choice-from-array tasks contain features that encourage participants to infer particular emotional meaning in vocalizations such as shouts, sighs, and laughs, *creating* the appearance of universal emotions. Participant performance on tests of cross-cultural emotion perception may be the result of multiple processes, including identifying perceptual similarities or using process-of-elimination strategies (e.g., Nelson & Russell, 2016a), perceiving affect (e.g., Gendron et al., 2014a), and learning categories online (e.g., Ferry, Hespos, & Waxman, 2010), calling into question the validity of interpreting such performance as direct evidence for the innate universality of certain emotion categories. In addition, it is possible that participants may use *conceptual combination* to complete cross-cultural emotion perception tasks when faced with unfamiliar emotion categories and exemplars. Conceptual combination is a fundamental cognitive capacity that allows individuals to construct instances of novel categories (such as emotion categories not present in their culture) by flexibly combining previously acquired conceptual knowledge (Barsalou, 1987). Conceptual combination does not imply that all properties of the original concepts will be invoked in a novel instance (Wu & Barsalou, 2009), but only those that are relevant for a particular situation (e.g., a given emotion story). This is the way that people can perceive instances of novel categories, including emotion categories, for which they have no single vocabulary word or even prior experience (for discussion, see L. F. Barrett, 2017a). As such, conceptual combination is a plausible process by which participants leverage features of the experimental context to infer emotional meaning in novel stimuli.

The Present Studies

In the studies that follow, we tested the potency of the classic choice-from-array task to create a context that allows participants to infer emotional meaning for novel vocalizations that have no inherent emotional meaning for them. We present data from one sample of U.S. participants (Study 3) and from samples in two other cultural contexts. In Study 1, we test individuals from a remote, small-scale society with relatively little exposure to western cultural norms, practices, and values (Henrich, Heine, & Norenzayan, 2010; Norenzayan & Heine, 2005)—hunter-gatherers from northern central Tanzania—who may not, on their own, make emotional meaning of vocalizations such as laughs, sighs, and shouts in the same way as U.S. participants. Our recent study of emotion perception in the Hadza indicates that they do not freely label scowls, smiles, frowns, and other facial poses in the presumed universal way (Gendron, Hoemann, et al., 2018). Therefore, Study 1 provides a strong test of our hypothesis. Study 2 tests Chinese participants living in China who have access to western cultural norms, practices, and values, but who also have their own, enhancing the variable emotional meanings that are available for physical cues such as facial configurations and vocalizations. Recent research with dynamic facial movements comparing United States and Chinese participants shows that instances of the same emotion are expressed with multiple sets of facial movements, and similar facial movements express different emotion categories (Jack et al., 2016). As a consequence, Study 2 provides a test of the choice-from-array task's potency for creating a context for emotional meaning making under conditions of enhanced ambiguity. Across three studies, we demonstrate that the choice-from-array task typically used in studies of cross-cultural emotion perception actively (although unintentionally) facilitates task performance, resulting in findings that make nonuniversal emotion categories appear universal.

In all three studies, the goal was not to separately test the potency of individual elements in the choice-from-array task, such as emotion words, repeated and blocked trials, or elaborate manipulation checks. Rather, we sought to demonstrate that, together, these elements create a highly structured version of the task that encourages emotional meaning making and, therefore, task performance, providing evidence for universal emotions that may not otherwise emerge. A more mechanistic approach is not optimal at this stage of our research for two reasons. First, while it may be possible to conduct systematic control conditions in laboratory experiments, it is often not feasible in field studies with remote populations limited in number (such as the Hadza), where researchers must operate under site constraints. Second, it is common practice to establish the replicability and robustness of an effect before investing the resources to undertake the long series of studies required to isolate each feature and test its unique contribution, or manipulate features to model their synergistic impact.

There are several notable aspects to the studies we report in this article. First, to provide a stringent test of our hypothesis, we traveled to northern central Tanzania to test emotion perception in members of a remote, small-scale hunter-gatherer culture, the Hadza (Study 1). This article reports the first study of emotion perception ever conducted with the Hadza (the second being our study of emotion perception in faces, reported in Gendron, Hoemann, et al., 2018). These studies are particularly important because, according to ideas from evolutionary psychology, universal and innate emotional expressions evolved to solve to the recurring fitness challenges of hunting and gathering in small groups on the African savanna (Pinker, 1997; Shariff & Tracy, 2011; Tooby & Cosmides, 2008). Therefore, the Hadza provide the strongest test of whether certain emotion categories are universal. Their cultural isolation is rapidly under assault, becoming contaminated by tourism and assimilation, and so the opportunity to study their emotional lives is rapidly disappearing. We examined the replicability of our findings by conducting the same study in an industrialized culture from the East (China, Study 2) and the West (the United States, Study 3).

Second, we chose to study the impact of the choice-from-array task on six emotion categories that would be novel to our participants: *gigil*, *greng jai*, *glückschmerz*, *itoshii*, *lajja*, and *liget* (see Table 1). These categories are not translatable by a single word in English, meaning that participants are less likely to invoke these specific category boundaries in their daily lives. (It does not mean that participants would be unable to understand or relate in some way to the experienced described.) Nor do these emotion categories meet the usual criteria for universality: they do not appear in Darwin's *The Expression of the Emotions in Man and Animals* (Darwin, 1872/2005), they do not appear to be evidenced in the behaviors of nonhuman animals (see Tracy & Randles, 2011), and they are not thought to solve a recurring evolutionary challenge for our Pleistocene ancestors (see Shariff & Tracy, 2011). We provided participants with short descriptive scenarios and stipulated (i.e., made-up) vocalizations for each emotion category (Table 1, see online supplemental material for details). We verified that the six novel emotion categories were unfamiliar in the cultural samples being tested, in that participants did not consistently and specifically associate the scenarios or vocalizations with preexisting emotion categories in their native language. This verification process is reported in the method section of each study.

On each trial in our studies, participants heard scenarios describing an emotional experience for which they did not have a preexisting category or word, and then heard two contrived vocalizations, one that was invented as the target for the novel emotion category and the other invented for different novel emotion cate-

Table 1
Novel Emotion Categories

Category	Pronunciation	Origin	Description	Scenario
Gigil	GHEE-ghil	Philippines	The overwhelming urge to squeeze or pinch something that is very cute	Someone sees a small, chubby, lovely baby and wants to squeeze it tightly. They feel gigil.
Glückschmerz	GLOOK-shmairts	Germany	Displeasure derived from another's pleasure	Someone hears that a bad person had some good fortune, and feels upset about it. They feel glückschmerz.
Greng jai	kreng-JAI	Thailand	The feeling you get when you do not want someone to do something for you because it would be a pain for them	Someone is offered help from others, but does not want it, because it is too much trouble for the others. The person feels greng jai.
Itoshii	ee-toe-SHEE	Japan	Bittersweet longing for an absent loved one	Someone thinks pleasant things about their loved one who has moved away (to another camp).[a] They feel itoshii.
Lajja	lah-ZHAH	India	Respectful restraint or playful shame; pleasant adherence to social norms	Someone makes a small mistake that others will notice and feels bad, but also acts playful. They feel lajja.
Liget	LI-gut	Ilongot (Philippines)	Intense focus, passion, and energy associated with actively pursuing a challenge	Someone works very hard toward a goal, and feels a rush of energy and intense focus. They feel liget.

Note. Additional testing verified that these categories were unknown to the three cultural samples. Participants were asked to freely label vocalizations and scenarios developed for each category. Examination of the labels confirmed that participants produced neither consistent nor specific labels. See study-specific methods and results sections, as well as online supplemental material for details.
[a] The content in parentheses was included in Study 1 for clarification, but not Studies 2 and 3.

gory (the foil). We hypothesized that the choice-from-array task would aid participants in selecting stipulated target vocalizations for six novel emotion categories from around the world with a level of agreement normally interpreted as evidence for universality. That is, the choice-from-array task would make six nonuniversal categories look universal in three different samples from around the world.

If certain emotion categories are indeed universal, and the choice-from-array task does not encourage participants to choose the expected answers, we would expect to find support for the null hypothesis: participants would not choose the stipulated vocalizations for these novel emotion categories at levels greater than chance. In contrast, support for our hypothesis would be found if participants perform at above-chance levels, choosing the vocalizations we invented for the emotion categories that are not traditionally part of United States, Chinese, or Hadza culture. Specifically, these results would be consistent with (even if they do not directly test) an account of novel emotion meaning-making based on conceptual combination, in addition to other general processes like affect perception and category learning.

Use of a Bayesian Analytic Approach

We used Bayesian hypothesis testing to quantify the evidence for the universality hypothesis (corresponding to the null hypothesis) that participants would select target vocalizations for the novel emotion categories at chance levels, as well as our alternative (task-as-context, henceforth "context") hypothesis that they would select the vocalizations at levels greater than chance. The ability to quantify evidence in favor of both null and alternative hypotheses is one of the main advantages of Bayesian hypothesis testing over frequentist approaches such as *t* tests and analysis of variances (ANOVAs; e.g., Edwards, Lindman, & Savage, 1963; Rouder, Morey, Speckman, & Province, 2012; Wagenmakers et al., 2015) because *p* values cannot provide support for the null (Wetzels et al., 2011). Bayesian hypothesis testing allowed us to assess the ratio between the probability of our context hypothesis given the data, as well as the probability of the null (universal) hypothesis given the data.

Another key advantage to Bayesian hypothesis testing is that it yields a de facto power analysis and replaced the need for a separate power analysis (Berger & Mortera, 1999; Berger & Wolpert, 1988; Rouder, 2014; Wagenmakers et al., 2015). The main statistic of interest is the Bayes factor, which expresses the relative evidence for null and alternative hypotheses and consistently trends toward truth as data accumulate, unlike a *p* value (Rouder, Speckman, Sun, Morey, & Iverson, 2009). Using this approach, we were able to determine that we had adequate sample sizes to test both hypotheses. We supplemented our Bayesian analyses with hierarchical generalized linear modeling analyses (HGLM; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004), as discussed in the results sections of each study.

Study 1: Remote Small-Scale Sample—The Hadza of Tanzania

The Hadza are a seminomadic, hunter-gatherer population who live in small bands in Tanzania and remain relatively isolated from other cultural practices and norms. They are egalitarian foragers who bring food back to camp for distribution (Berbesque, Wood, Crittenden, Mabulla, & Marlowe, 2016). They consume a diet that is primarily based on the collection of wild foods (Apicella & Crittenden, 2015); the women gather and the men hunt and collect honey. The Hadza are not a Paleolithic population, but have been continuously hunting and gathering for thousands of years in East Africa, and they are one of the last groups to live in a social and physical context closer to that of our Pleistocene ancestors than our own (Marlowe, 2010). An estimated 150–200 Hadza individuals currently live in remote camps (Jones, 2016). In Study 1, we recruited and tested 55 of these individuals. Before our visit to the Hadza in 2016, they had not participated in any studies of emotion perception, although some had previously participated in social cognition research (H. C. Barrett et al., 2016; Bryant et al., 2016). During this visit, we conducted two studies (Study 1 reported here, and a second study with posed emotional faces reported in Gendron, Hoemann, et al., 2018).

Method

Participants. Participants were 55 native Hadzane speakers (21 female) from the Hadza ethnic group, recruited from the area surrounding three camps located southeast of Lake Eyasi in northern central Tanzania: Camp 1 ($n = 20$), Camp 2 ($n = 27$), and Camp 3 ($n = 8$; see Figure 7 in the online supplemental material). All three camps were far from the regional towns, and there was no evidence of recent tourism to these areas. During the time of data collection, the majority of the Hadza individuals we tested did not travel except to other Hadza camps, although regional towns and small cities were visited occasionally for access to supplies and medical care. Only three individuals we tested reported venturing further afield to the larger cities of Arusha, Dar es Salaam, or parts of Kenya. Participants ranged in age from 18 to mid-70s, with a median decade of 30–40; Hadza individuals do not keep track of their biological age, so estimates were based on maternal or paternal history and personal knowledge of our translator (the third author, S.M.), who is ethnically Hadza as well as an experienced research assistant.

Participants were tested in their native language, Hadzane, which is a click-based language considered to be a linguistic isolate (Sands, 1998). Almost all Hadza speak some Swahili, however, as they use this language to communicate with neighboring ethnic groups. Twenty adults in our sample reported fair to excellent knowledge of Swahili, whereas the other 35 mostly reported little knowledge of Swahili. In addition, five individuals reported modest knowledge of another neighboring language, such as Datoga, Iraqw, or Isanzu. Only one participant claimed rudimentary knowledge of English, but did not use it with the experimenters. Sixteen participants in the sample had attended local primary school for between 2 and 7 years. One participant attended the regional secondary school for 2 years.

Data collection was approved by the Office of Human Subject Research Protection Institutional Review Board at Northeastern University as well as the Tanzanian Commission for Science and Technology (COSTECH). Hadza participants were verbally consented before participation and were remunerated with gifts (clothing, cookware, etc.).

Stimuli. Stimuli were 6 short scenarios (1 per emotion category) and 60 nonword vocalizations (10 per emotion category).

For comparison, Sauter et al. (2010) used short scenarios from nine emotion categories (*achievement*, *amusement*, *anger*, *disgust*, *fear*, *sensual pleasure*, *relief*, *sadness*, and *surprise*), each with 10 nonword vocalizations.

Emotion concept selection. We selected six emotion concepts from the Interdisciplinary Affective Science Laboratory's database of emotion concepts that have been deemed "untranslatable" into English. We compiled this database (available upon request) using published psychological, anthropological, and linguistic literatures (e.g., Lomas, 2016; Rosaldo, 1980; Russell, 1991), as well as Web sites and publications in the popular media (e.g., BetterThanEnglish, 2012; De Boinod, 2007; Lin, 2013). We selected six emotion concepts that were easy to portray with static facial configurations and vocalizations. None of these concepts are labeled with single words in English. None are frequently described in the English media. Therefore, none have been studied in a scientific context or stipulated as universal. We refer to these as "novel" emotion concepts because they are unfamiliar to English and Hadzane speakers alike (i.e., they are not represented as existing emotion categories in either English or Hadzane). Fluent English speakers from the United States on Amazon's Mechanical Turk (MTurk) rated the six emotion concepts on valence and arousal (n = approx. 15 per concept; 43.5% female; mean age = 32.2, SD = 8.79; see Figure 1 in the online supplemental material).

Emotion scenarios. We generated six short scenarios, one for each novel emotion concept. The scenarios (see Table 1) were developed based on concept descriptions by native speakers of the source languages, and were adapted for cultural fit by the second author (A.N.C.), a nutritional and ecological anthropologist with extensive field experience working among the Hadza. Our translator, the third author (S.M.), a native speaker of Hadzane and a fluent speaker of English, also reviewed the scenarios, and recorded them in Hadzane. He is one of the few Hadza to have attended university, and has previously been involved in anthropological research (e.g., Raichlen et al., 2017). Each scenario was one to two sentences long, and did not differ substantially in structure from those used in Sauter et al. (2010), which were themselves adapted from the scenarios used by Ekman et al. (1969). Each scenario briefly described an emotional situation in concrete terms and concluded with a description of how the protagonist felt using the target emotion word in its original language. For example, "Someone sees a small, chubby, lovely baby and wants to squeeze it tightly. They feel *gigil*."

Vocalizations. We (the authors) *invented* the short vocal bursts that served as posed vocal expressions of each emotion category (i.e., we stipulated their pitch, tone, and duration) based on our folk understanding of the novel emotion concepts in Table 1. We then directed five posers from the United States (3 women, 2 men) to generate these short vocal bursts. Imitating emotion poses is a well-established method for stimulus generation in studies of emotion perception (e.g., Ekman, Levenson, & Friesen, 1983; Schröder, 2003). Posers were first provided with descriptions of the desired vocalizations for each emotion category (Table 3 in the online supplemental material), which were also modeled by the first author. Posers then generated the vocalizations while imagining the emotional situations described in the scenarios. All vocalizations were cleaned for ambient noise and adjusted for mean peak amplitude using Audacity (Audacity Team, 2015).

Next, we examined whether the vocalizations portraying the same emotion category clustered together in multidimensional space. We opted for a multidimensional scaling (MDS) approach (Hout, Papesh, & Goldinger, 2013; Jaworska & Chupetlovska-Anastasova, 2009) because it allowed us to discover, rather than confirm, whether the vocalization associated with the same emotion category are similarly positioned in perceptual space. Prior MDS analyses confirm that vocal stimuli sharing similar acoustic features do cluster together (Gygi, Kidd, & Watson, 2007). Pilot participants (N = 7) heard pairs of vocalizations, one member of the pair played after the other, and then rated their similarity from 1 (*very similar*) to 9 (*very dissimilar*). All possible pairwise comparisons were presented in randomized order, with vocalizations randomly assigned to play first or second. A single matrix of mean similarity ratings was computed for all vocalization pairs and subjected to an alternating least squares approach to scaling (ALSCAL) procedure (Young & Lewyckyj, 1979). Results indicated that vocalizations posed for the same novel emotion category clustered together in multidimensional space (Figure 4 and 5 in the online supplemental material). A stress-by-dimensionality plot (Figure 3 in the online supplemental material) indicated that the three-dimensional solution provided the best fit for the data. The dimension loadings for the vocalizations suggested that the first two dimensions are valence and arousal, replicating prior findings from MDS analyses of affective stimuli (e.g., Russell, 1980; Sauter et al., 2010 also report similar findings using principal component analysis [PCA]). Cluster locations were similar to explicit valence and arousal ratings for the emotion scenarios (Figure 1 in the online supplemental material). A separate sample of MTurk raters (N = 25; 8 female; mean age = 31.56, SD = 7.04) also verified that the posed vocalizations for the same novel emotion category clustered together in multidimensional space (Figure 6 in the online supplemental material). We elicited a second set of vocalizations from five Namibian posers (3 women, 2 men) as part of a separate study, based on these normed vocalizations, for a total of 10 per emotion category.

Procedure.

Verifying vocalization novelty. Before completing the choice-from-array task, 19 participants freely labeled six of the made-up vocalizations, one for each emotion category under investigation. Participants were asked to describe how the target person making each sound felt using a word or short phrase. Participants who provided a description of a situation or behavior were prompted to describe how the target person was feeling, to replicate the method described by Sauter et al. (2010, 2015). Participant-provided labels were coded according to whether they represented a known emotion category in English, and whether or not this category was subordinate to or synonymous with a purportedly universal emotion category (i.e., *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*; see online supplemental material for details on the coding procedure). We expected, and found, that the consistency and specificity of labeling each vocalization was low. Participants did not label the vocalizations with the words for the novel emotion concepts, nor did they provide labels such as angry, fearful, sad, or their synonyms. In fact, most participants provided labels for the vocalizations that did not represent a known emotion category in English (for *gigil*, 79% of responses did not refer to a discrete emotion category; *glückschmerz*, 37%; *greng jai*, 81%; *itoshii*,78%; *lajja*, 33%; *liget*, 53%; panel A of Figure 8 in the online supplemental material). Furthermore, participants labeled each vocalization with words that ranged in their affective features, using both positively or negatively valenced labels, as well as labels that were high or low in arousal, indicating that there was little consistency and specificity in the affective features associated with the vocalizations (panel A of Figure 9 and 10 in the online supplemental material).

Choice-from-array task. Each participant completed the classic choice-from-array task, following the methods used in Sauter et al. (2010; for additional details, see Sauter et al., 2015). All participants were tested individually using a Dell ATG laptop and headphones. Participants listened to a prerecorded scenario in their native language (e.g., in English, "Someone sees a small, chubby, lovely baby and wants to squeeze it tightly. They feel *gigil*."). Participants then completed an extensive manipulation check (panel A of Figure 1) as described in Sauter et al. (2015). They were asked to describe how the scenario protagonist feels, and a correct answer would contain features other than simply repeating the target emotion label used in the scenario. For example, a response (in Hadzane or Swahili) of "he feels *gigil*" was not accepted, whereas a response of "he feels good," "he feels loving," or "he feels angry" was acceptable. Participants were allowed to listen to the recorded scenario multiple times, and received verbal feedback in their native language on their responses. Because the Hadza are a preliterate society, we communicated verbal instructions and feedback through our Hadzane translator (S.M.).

Once participants had passed the manipulation check for a given emotion category, they then completed a block of trials for that category (panel B of Figure 1). For example, on every trial within a *gigil* block, participants listened to the audio recording of the scenario followed by target (an invented *gigil* sound) and a foil vocalization (e.g., an invented sound for the *lajja* category). Each trial drew from a list of 10 possible vocalizations for each novel emotion category. Foils were drawn randomly from one of the other (nontarget) categories and were matched for sex of poser. As the first vocalization played, an icon appeared on the left side of the computer screen, and then disappeared. While the second vocalization played, the mirror version of the icon appeared on the right side of the screen, and then disappeared. Both icons then appeared simultaneously, and participants pressed the left or right icon to indicate which vocalization matched the emotion portrayed in the scenario. Target position (left or right) was randomized for target and foil. On a given trial, scenarios and vocalizations were repeated for participants who wished to hear them again. Upon completing the first trial, participants heard the *gigil* scenario

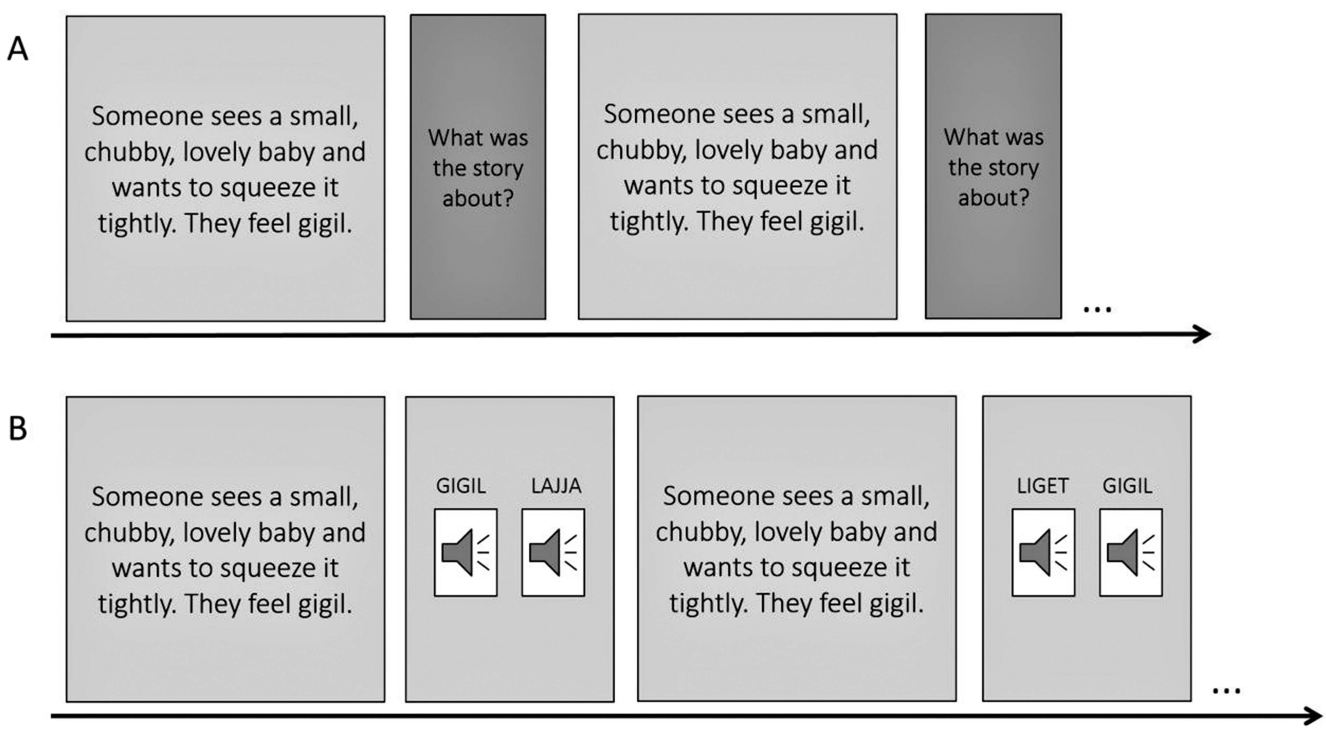

Figure 1. The classic, highly structured choice-from-array task with novel emotion categories. Prerecorded material presented over headphones is depicted in light gray boxes; verbal interactions with the experimenter are presented in dark gray boxes. (A) Manipulation check: Participants listened to a scenario in their native language and were then asked to describe how the protagonist in the story feels. (B) Perception trials blocked by emotion category. After the manipulation check for a given emotion category, participants completed a block of trials for that category. On every trial, participants listened to the scenario again followed by target and foil vocalizations. Vocalizations were played, one at a time, with an icon presented concurrently on either the left or right side of the screen (no words were present on screen). Participants touched an icon to select a vocalization. Once a trial was complete, participants completed another trial from the same emotion category. Targets were presented randomly on the left or right within a block.

again, followed by an invented *gigil* sound and another foil (e.g., an invented sound for the *liget* category). Once the block of trials for a given emotion category was complete, participants proceeded to another block of trials for a different emotion category. Emotion category blocks were presented in randomized order. Participants completed six emotion category blocks of five trials each, for a total of 30 trials.

Field site constraints dictated that each participant spend a similar amount of time in testing. As a result, all participants completed all experimental trials, but a participant's data for a given block of trials were removed before analysis if he or she failed the manipulation check for that emotion category three times. The sample size used in the analyses can be found in Figure 3.

Verifying emotion scenario novelty. After completion of the choice-from-array task, participants who freely labeled the vocalizations also freely labeled long-form versions of the scenarios that did not have any emotion word embedded. Only six participants were available to do this norming because of time constraints. As expected, participant-provided labels showed little agreement across participants and did not converge with known English concepts *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise* (see panel A of Figure 11 in the online supplemental material). A similar finding was observed for valence and arousal properties (see panel A of Figure 12 and 13 in the online supplemental material).

Results

Bayesian hypothesis testing. We conducted a Bayesian one sample *t* test to directly evaluate the degree of support for the null (universal) hypothesis (i.e., chance-level performance at .5), as well as to evaluate the sequentially accumulating support for the alternative (context) hypothesis (i.e., mean performance greater than chance; Rouder et al., 2009; Wetzels, Raaijmakers, Jakab, & Wagenmakers, 2009). Analyses were computed in JASP (JASP Team, 2017). Given that the effect sizes that support the context hypothesis might vary, a Bayesian approach evaluates a distribution of expected effect sizes rather than a single estimate. In JASP, we estimated the probability that the context hypothesis was true given a one-sided prior probability distribution where the median effect size was equivalent to Cohen's $\delta = .7$. This is consistent with strong evidence for the universality of emotion categories in choice-from-array tasks (Haidt & Keltner, 1999).[5]

A Bayesian one-sample *t* test conducted on aggregate response data demonstrated robust support for the hypothesis that participants would perform above chance relative to the null hypothesis. That is, Hadza participants performed at a level typically interpreted as evidence of universality (i.e., perceived the intended novel emotions) when tested using the highly structured version of the classic choice-from-array task used by Sauter et al. (2010). As can be seen in panel A of Figure 2, the Bayes factor for our Hadza participants was $BF_{10} = 187.905$, indicating that the observed data are 187.905 times more likely under the alternative (context) hypothesis than under the null (universal) hypothesis. This is considered "very strong" (Bayes factor greater than 150; Kass & Raftery, 1995) or "decisive" (Bayes factor greater than 10^2; Jeffreys, 1961) evidence for the context hypothesis.

These findings were robust, even when we varied the parameters of our analysis. For example, we varied the width of the prior probabilities for the effect size under the alternative (context) hypothesis. Narrower priors allow for a smaller range of expected effect sizes and, therefore, represent a more stringent test of the context hypothesis when compared with the null (universal) hypothesis (i.e., a prior width very close to zero most favors finding evidence for the null hypothesis). Under these conditions, the Bayes factor still indicates substantial support for the context hypothesis (i.e., $BF_{10} > 3$).

We also generated sequential analysis plots to examine the development of the Bayes factors as data accumulated. We observed that cumulative data increasingly provided evidence for the alternative (context) hypothesis when compared with the null (universal) hypothesis, lending confidence that our sample size was large enough and we were sufficiently powered to test our hypotheses (panel D in Figure 2).

Hierarchical generalized linear modeling (HGLM). In addition, we used hierarchical generalized linear modeling (HGLM; Raudenbush et al., 2004) to examine whether participants' performance was above chance-level responding when analyzed by novel emotion category. We chose HGLM for two reasons. First, the main dependent measure in the choice-from-array task is dichotomous: Did a participant pick the target vocalization on a given trial? (1 = *yes, correct*, 0 = *no, incorrect*). Participants' performance across multiple trials can be modeled as a binomial distribution, bounded at both ends (i.e., has values between 0 and 1). Traditional parametric approaches such as one-sample *t* tests and ANOVAs with binomial or categorical data cannot be used to analyze data from the choice-from-array task because binomial response data must be treated as proportions or percentages, resulting in confidence intervals that can extend beyond the interpretable values between 0 and 1, leading to spurious results (e.g., Agresti, 2002; Jaeger, 2008). HGLM accounts for data that are not normally distributed by using a nonlinear link function (Nelder & Wedderburn, 1972).

The second reason we chose HGLM is that nonparametric approaches such as χ^2 (e.g., Sauter, Eisner, Ekman, et al., 2010) and exact binomial *t* tests (e.g., Cordaro et al., 2016) do not account for the nested, nonindependent nature of the data in a repeated measures choice-from-array task. Trials are nonindependent because the probability of a correct response on one trial could influence the probability of a correct response on subsequent trials (e.g., because of perceptual similarity of the target vocalizations within a category, which are blocked together in the highly structured version of the task). HGLM is well suited for these dependencies because trials can be grouped into clusters, allowing for the error term to be partitioned, and increasing the power of the model to detect the effect (Guo & Zhao, 2000; Kenny, Korchmaros, & Bolger, 2003).

We analyzed the data using a Bernoulli multilevel model, which estimated a log-odds (i.e., the probability of performing above chance of .5) using a log-linear link function. Data were structured in a two-level model. Trials on which participants selected either

[5] In the analysis, we specified a Cauchy prior of $r = \sqrt{2}/2$ (i.e., .707; Morey, Rouder, & Jamil, 2015). A Cauchy prior width of $r = 1$ has also been recommended (Rouder et al., 2009; Wetzels et al., 2009). We used the smaller r value (i.e., $r = .707$) because it represented a more conservative test of our context hypothesis; however, we generated Bayes factors at $r = 1$ and $r = 1.414$ for sake of comparison (panel C in Figures 2, 4, and 5).

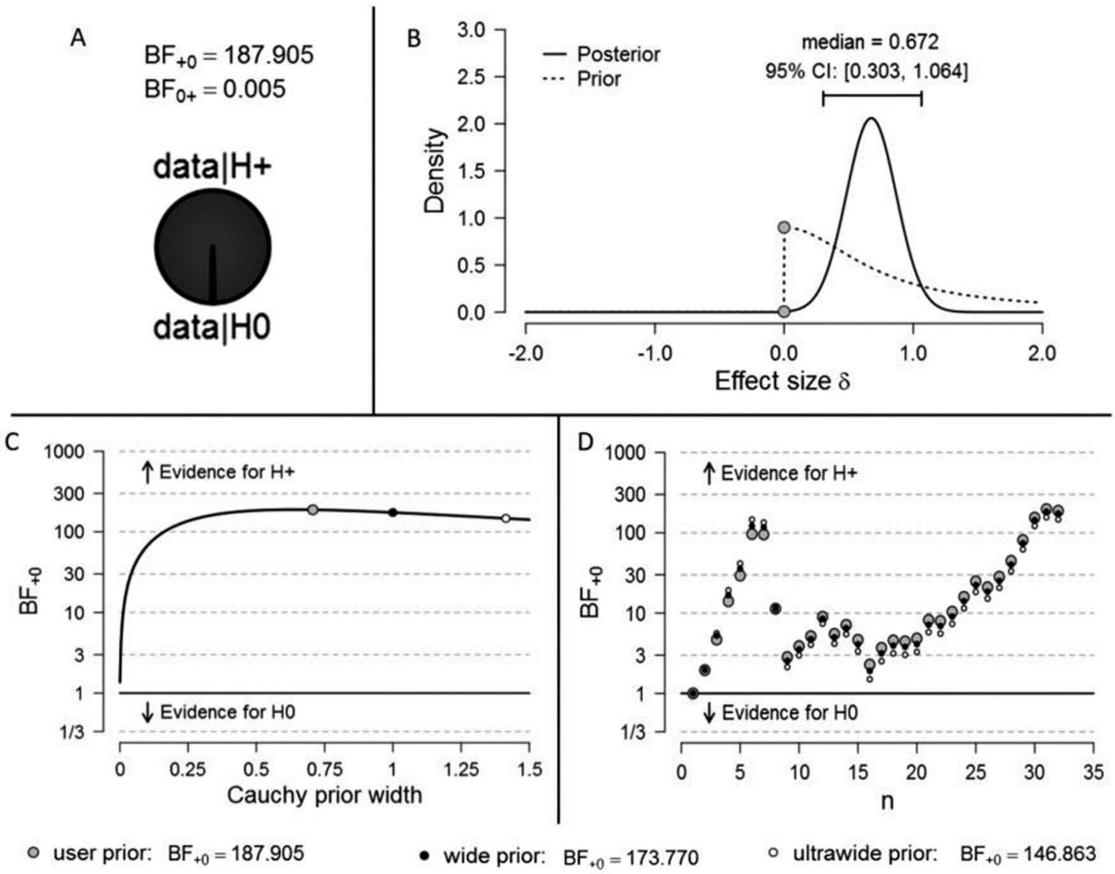

Figure 2. Inferential plots for Study 1, based on Bayesian one-sample *t* tests on overall performance of Hadza participants. (A) Pizza plot providing a proportional representation of the ratio of evidence for the alternative (context) hypothesis to evidence for null (universal) hypothesis. (B) Posterior distribution based on a one-sided prior distribution. Posterior mass to the right of zero indicates that participant performance is consistently above chance-level responding of .5. (C) Bayes factor robustness check. Narrower priors indicate a smaller range of expected effect sizes, favoring the null (universal) hypothesis. (D) Sequential analysis of evidence accumulated over the course of data collection. All figures adjusted from JASP (JASP Team, 2017).

the target or foil vocalization for six emotion categories (Level 1) were nested within individuals (Level 2). We used an intercept-as-outcome approach with dummy codes for each emotion category (Raudenbush et al., 2004). We have used the intercepts-as-outcomes approach to analyze other repeated measures data (e.g., Anderson, Siegel, & Barrett, 2011; L. F. Barrett & Niedenthal, 2004). We used a random effects model to compute the population-average estimates with robust *SE*s, allowing us to generalize the average probability of success beyond those individuals included in the sample. All HGLM analyses were conducted in HLM7 (SSI Inc., Lincolnwood, IL). See online supplemental material for model specifications.

The results are reported in Figure 3, panel A (see also Table 4 in the online supplemental material for detailed results). The analysis indicated that participants selected target vocalizations for *gigil*, *itoshii*, and *liget* at levels significantly above chance (*p*s range from .005 to .030). Participants selected targets for *lajja* and *glückschmerz* at levels approaching conventional levels of statistical significance (m = .58, p = .089 and m = .57, p = .093, respectively); notably, performance means were identical to *liget* (m = .58). The *glückschmerz* category obtained a bimodal distribution of responses: while 12 participants performed below chance, 17 participants performed above chance, including 10 who performed at or near ceiling. The *lajja* category was underpowered because many participants failed the manipulation check after describing the feeling as unpleasant, whereas it is experienced as pleasant in the original Oriya Hindu culture (Menon & Shweder, 1994). Of the 12 participants who passed the manipulation check for *lajja*, however, 8 selected target vocalizations at levels above chance. (See Figure 14 in the online supplemental material for distribution of performance above and below chance per novel emotion category.)

Comparison with Sauter et al. (2010). We used the highly structured choice-from-array task (i.e., including emotion words, repeated and blocked trials, and elaborate manipulation checks) from Sauter et al. (2010) to assess the ability of Hadza participants to match stipulated vocalizations to descriptions of novel emotion categories. As can be seen in panel B of Figure 3, the overall pattern of results resembled the pattern of findings in Sauter et al. (2010) for *anger*, *sadness*, *fear*, and other emotion categories that

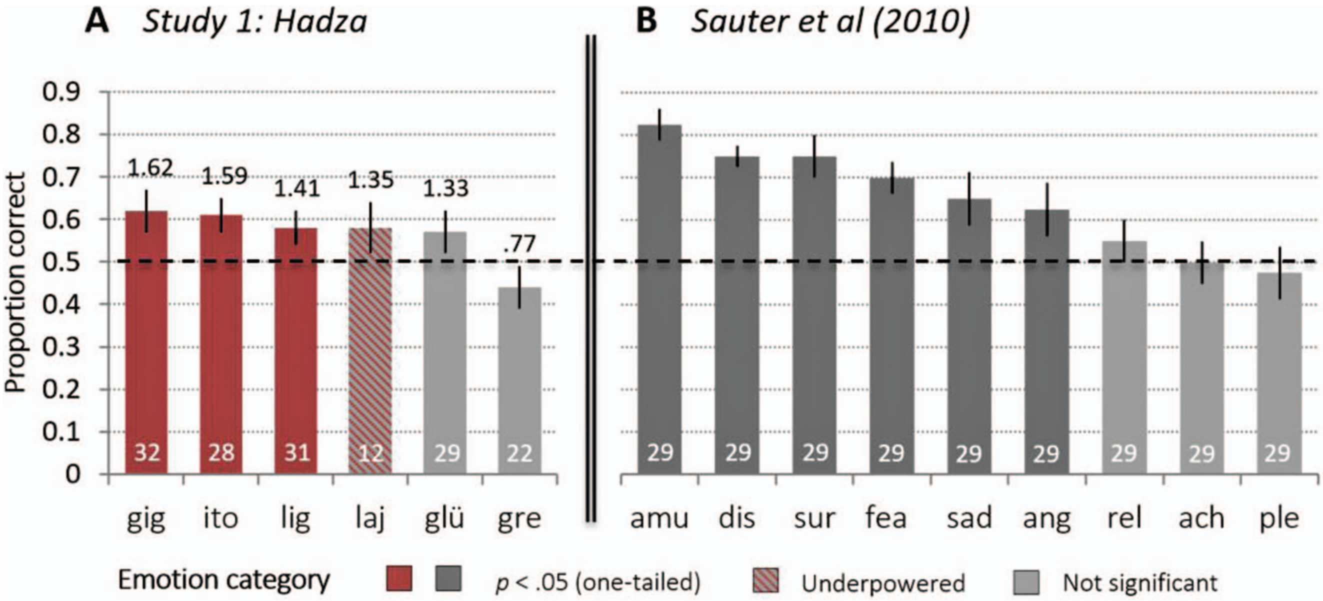

Figure 3. Results from Study 1 (panel A) compared against Sauter et al. (2010; panel B). Results data from Sauter et al. (2010) were originally presented in terms of mean number of correct responses (out of four trials per emotion category). These data have been replotted in descending order of proportion correct to facilitate direct comparison with the present study. Dashed line indicates chance-level performance (.5). Sample size per emotion category reported in white font at the bottom of each column. Standard error bars (±1 *SE*) are provided as distributional information only: the location of error bars above the chance line is not indicative of significant above-chance performance because these data were binomially rather than normally distributed. Effect sizes for Study 1 are reported above each column using the odds ratio (*OR*), which expresses group difference in probabilities when the outcome is dichotomous and the data are analyzed by logistic regression (Fleiss & Berlin, 2009). The *OR* is obtained by transforming the *b* weights using e^b (Feingold, 2013). In line with our strong a priori hypotheses, all tests were conducted using one-tailed probability thresholds to avoid Type II errors. Panel A: gig = gigil; gre = greng jai; glü = glückschmerz; ito = itoshii; laj = lajja; lig = liget; Panel B: ach = achievement; amu = amusement; ang = anger; dis = disgust; fea = fear; ple = sensual pleasure; rel = relief; sad = sadness; sur = surprise.

have been claimed as universal. In Sauter et al. (2010), individuals from the Himba culture in northwestern Namibia were asked to complete the same choice-from-array task, and chose the expected vocalizations at rates above chance, providing apparent support for the universality hypothesis. These results stand in contrast to those obtained when Himba participants freely labeled the vocalizations, which revealed little evidence that posed vocalizations were perceived universally as *anger*, *fear*, and so forth (Gendron et al., 2014a).

Studies 2 and 3: Urban Industrialized Samples—China and United States

In Study 1, we predicted and found support for the hypothesis that the classic choice-from-array task provides an experimental context that helps participants choose the target stimuli for emotion categories. A remote sample of Hadza hunter-gatherers selected stipulated (i.e., made-up) target vocalizations for novel emotion categories at levels significantly above chance, appearing to provide evidence of cross-cultural emotion perception despite verified lack of exposure. In Studies 2 and 3, we replicated the study in samples of Chinese and United States participants from urban industrialized cultural contexts.

Method

Participants. Study 2 participants were 34 native Mandarin speakers (24 female) tested in Dalian, China. Participants were recruited through student networks at Liaoning Normal University, and were required to be native Mandarin Chinese speakers, over 18 years of age, and have normal or corrected-to-normal hearing and vision. All testing was completed at the university. Participants ranged in age from 18 to 27, with a median age of 21 years. Study 2 participants provided written consent before participation and were remunerated with ¥20.

Participants for Study 3 were 42 native English speakers (21 female) tested in Boston, MA. Participants were recruited through both the psychology department and the broader community at Northeastern University, and were required to be native English speakers, over 18 years of age, and have normal or corrected-to-normal hearing and vision. All testing was completed at the university. Participants ranged in age from 18 to 67, with a median age of 19 years. Study 3 participants provided written consent before participation and were remunerated with study credit or $10.

As in Study 1, participants completed all experimental trials, even those that were later removed from analysis because of failure to pass a manipulation check. The final sample sizes available for analysis are reported in Figure 6.

Stimuli. Vocalizations were the same as used for Study 1. Scenarios for Study 2 were translated into Mandarin Chinese by a native speaker of the language who is also fluent in English, and were then back-translated into English by a second bilingual speaker to confirm translational equivalency (as recommended by Brislin, 1970). The fifth author (C.L.) served as the Chinese translator for Study 2. Scenarios were recorded by native speakers of Mandarin Chinese (Study 2) and North American English (Study 3).

Procedure. The procedure was identical to Study 1, including the assessments of vocalization and emotion concept novelty. We also provided written instructions in addition to verbal instructions and feedback because, unlike the Hadza, both Chinese and English are associated with written text. All participants were literate in their native language. Written instructions for Mandarin Chinese were translated and back-translated.

Verifying vocalization novelty. All Study 2 and 3 participants completed the vocalization free-labeling task before choice-from-array task. Once again, we found that a substantial percentage of labels did not represent known emotion categories in English (Study 2, 43% of responses; Study 3, 36% of responses). In Study 2, consistency increased as compared with Study 1, but specificity continued to be low: for example, while 47% of participants associated *gigil* with *surprise*, this label was also applied to every other vocalization, including 37% of the labels for *lajja* (panel B of Figure 8 in the online supplemental material). In Study 3, participants used more emotion terms to label the scenarios, although these were wide-ranging and were neither consistently nor specifically associated with *anger*, *disgust*, *fear*, *happiness*, *sadness*, *surprise*, or other English emotion categories (e.g., only 11% of labels for *gigil* were associated with *surprise*, along with 21% of the labels for *lajja*; panel C of Figure 6 in the online supplemental material). High overall variation in valence and arousal properties were observed for each novel emotion category, indicating little agreement on their affective features within the cultural samples (panels B and C of Figure 9 and 10 in the online supplemental material).

Verifying emotion category novelty. All participants from Studies 2 and 3 also completed the scenario free-labeling task after completion of the choice-from-array task. As with the vocalization free-labeling task, many participants provided labels that did not represent a known emotion category (Study 2, 33% of responses; Study 3, 41% of responses). Nonetheless, the labels for the scenarios were more consistent than for the vocalizations. For example, *gigil* was the most consistently labeled category in Study 2, with 61% of participants associating it with *happiness* (although 26% labeled it as another emotion category, and 13% did not associate it with a discrete emotion). In Study 3, *itoshii* was the most consistently labeled category, with 35% of participants associating it with *sadness* (although 48% labeled it as another emotion category, and 16% did not associate it with a discrete emotion). See panels B and C of Figure 11 in the online supplemental material for the full distribution of results. The affective properties of the scenario labels also evidenced greater consistency in comparison to those for the vocalizations, both within and across cultural groups (panels B and C of Figure 12 in the online supplemental material and 13). However, valence and arousal did not correspond for the vocalization and the scenario associated each novel emotion category, suggesting that participants did not have a preexisting association between the two in terms of core affective features (see online supplemental material for further details and discussion).

Results

Bayesian hypothesis testing. Bayes factors for one-sample t tests indicated "extreme" evidence in favor of the alternative (context) hypothesis over the null (universal) hypothesis in both Studies 2 and 3, replicating findings from Study 1, meaning that participants were able to choose the stipulated target vocalization over the foil at significant levels. Proportional representations of the ratio of evidence for the context hypothesis to evidence for the universal hypothesis are provided in panel A of Figures 4 (Study 2) and 5 (Study 3). Once again, we examined the robustness of our conclusions by comparing support for the context hypothesis according to varying widths of the prior for effect size. The Bayes factors continued to indicate extreme support for the context hypothesis even in more stringent tests using prior widths close to zero (panel C of Figures 4 and 5). Further, we observed that cumulative data for Studies 2 and 3 provided clear, increasing support for the context hypothesis (panel D of Figures 4 and 5), once again confirming that our sample sizes provided adequate power.

Hierarchical generalized linear modeling. In Study 2, Chinese participants selected the made-up vocalizations at a level significantly above chance for four of six novel emotion categories: *gigil*, *glückschmerz*, *greng jai*, and *itoshii* (Figure 6, panel A; see also Figure 15 in the online supplemental material for performance distributions and Table 5 for detailed results). All Chinese participants' data were excluded from the *lajja* category because they misunderstood *lajja*'s valence as unpleasant.

In Study 3, U.S. participants selected the made-up vocalizations at a level significantly above chance for all six novel emotion categories (Figure 6, panel B; see also Figure 16 in the online supplemental material for performance distributions and Table 6 for detailed results). This is despite the fact that, for the category *lajja*, only 10 U.S. participants were retained for analysis for understanding it as pleasant.

Comparing performance across Studies 1 through 3. Across all three samples, there is no emotion category on which participants performed consistently at chance. Participants in all three studies performed above chance or approaching conventional levels of significance in choosing made-up target vocalizations for three novel emotion categories that they had never before been exposed to: *gigil*, *itoshii*, and *glückschmerz* (Figure 3, panel A; Figure 6, panels A and B).

General Discussion

The choice-from-array task remains widely used in psychological research and is the most common task design in studies of emotion perception. In three studies, we demonstrated support for our hypothesis that a classic, highly structured choice-from-array task creates a context that encourages emotional meaning-making and in so doing may provide stronger evidence of cross-cultural emotion perception than would otherwise be observed. Indeed,

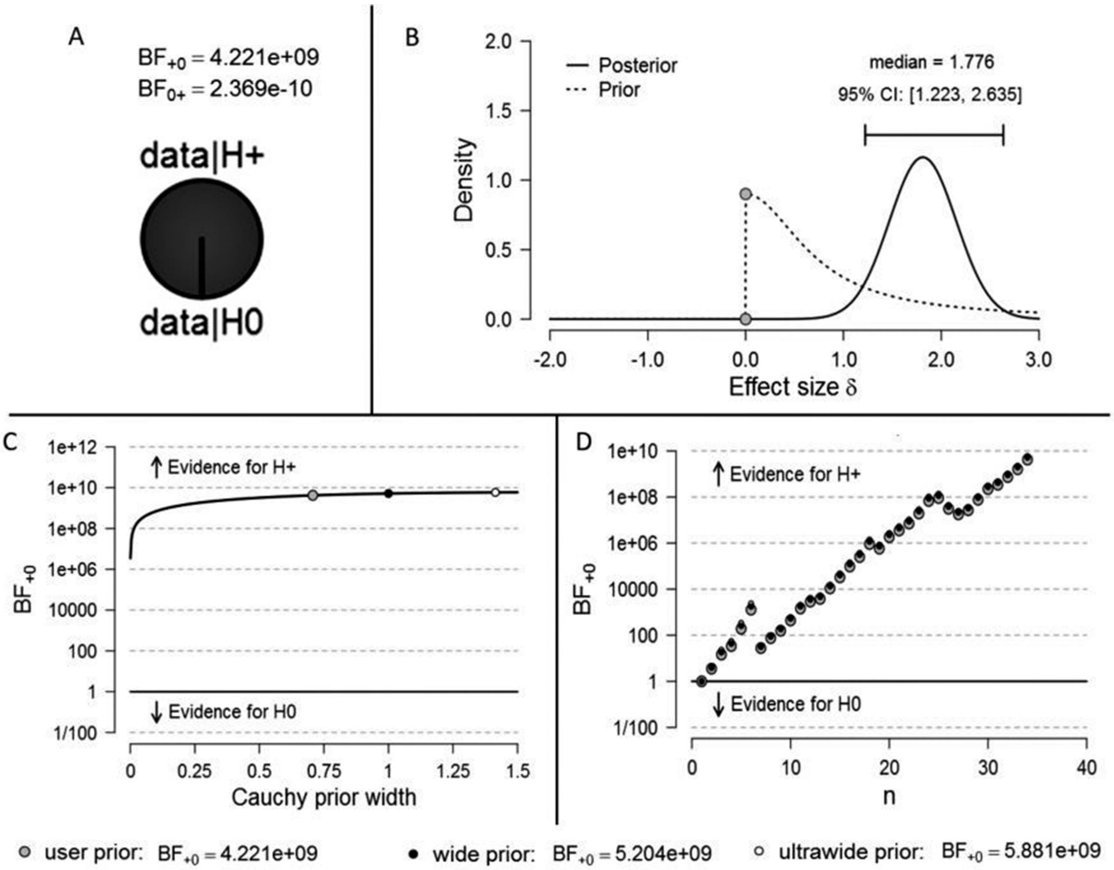

Figure 4. Inferential plots for Study 1, based on Bayesian one-sample *t* tests on overall performance of Chinese participants. (A) Pizza plot providing a proportional representation of the ratio of evidence for the alternative (context) hypothesis to evidence for null (universal) hypothesis. (B) Posterior distribution based on a one-sided prior distribution. Posterior mass to the right of zero indicates that participant performance is consistently above chance-level responding of .5. (C) Bayes factor robustness check. Narrower priors indicate a smaller range of expected effect sizes, favoring the null (universal) hypothesis. (D) Sequential analysis of evidence accumulated over the course of data collection. All figures adjusted from JASP (JASP Team, 2017).

such evidence is not observed when using other methods for assessing emotion perception (e.g., Crivelli, Jarillo, Russell, & Fernandez-Dols, 2016; Gendron, Hoemann, et al., 2018; Gendron et al., 2014a). It has long been known that telling participants an emotion story and asking them to select an emotion cue from a small set of options facilitates more consistent performance than less constrained experimental tasks, such as asking participants to freely label emotion cues. That may be why the method is so popular in the first place (Gendron & Barrett, 2009, 2017; Widen & Russell, 2013). Dashiell (1927) pioneered the choice-from-array task to overcome the comprehension and compliance issues associated with collecting data in preliterate communities who are unfamiliar with standard laboratory methods and did not provide strong evidence of cross-cultural emotion perception (Russell, 1994). Our studies show for the first time that contrived (i.e., made-up) vocalizations for nonuniversal emotion categories that are novel in three cultural contexts are made to appear universal when tested using a choice-from-array task.

The present studies did not separately manipulate each psychologically potent feature of the classic choice-from-array task (e.g., the presence of emotion words, repeated and blocked trials, and elaborate manipulation checks) to examine their independent or synergistic effects. This could be an avenue for future research. For example, previous evidence indicates that a choice-from-array task without the blocked trial structure still encourages above-chance performance (Gendron et al., 2014a). Our studies did not include explicit control conditions because of field site constraints (e.g., limited access to Study 1 participants). Nonetheless, our procedure for verifying vocalization novelty was a free-label task that was conducted before the choice-from-array task and did not include emotion words, blocked trials, or an elaborate manipulation check. Data from this task, therefore, give an estimate of participants' perception of the novel vocalizations without any additional experimental context. We found that participants within a cultural sample did not label individual vocalizations with a high degree of agreement, and they often used the same or similar words to label multiple vocalizations (for details, see results and online supplemental material). Thus, participants' freely generated labels did not provide the evidence of cross-cultural emotion perception observed in the choice-from-array task. These findings

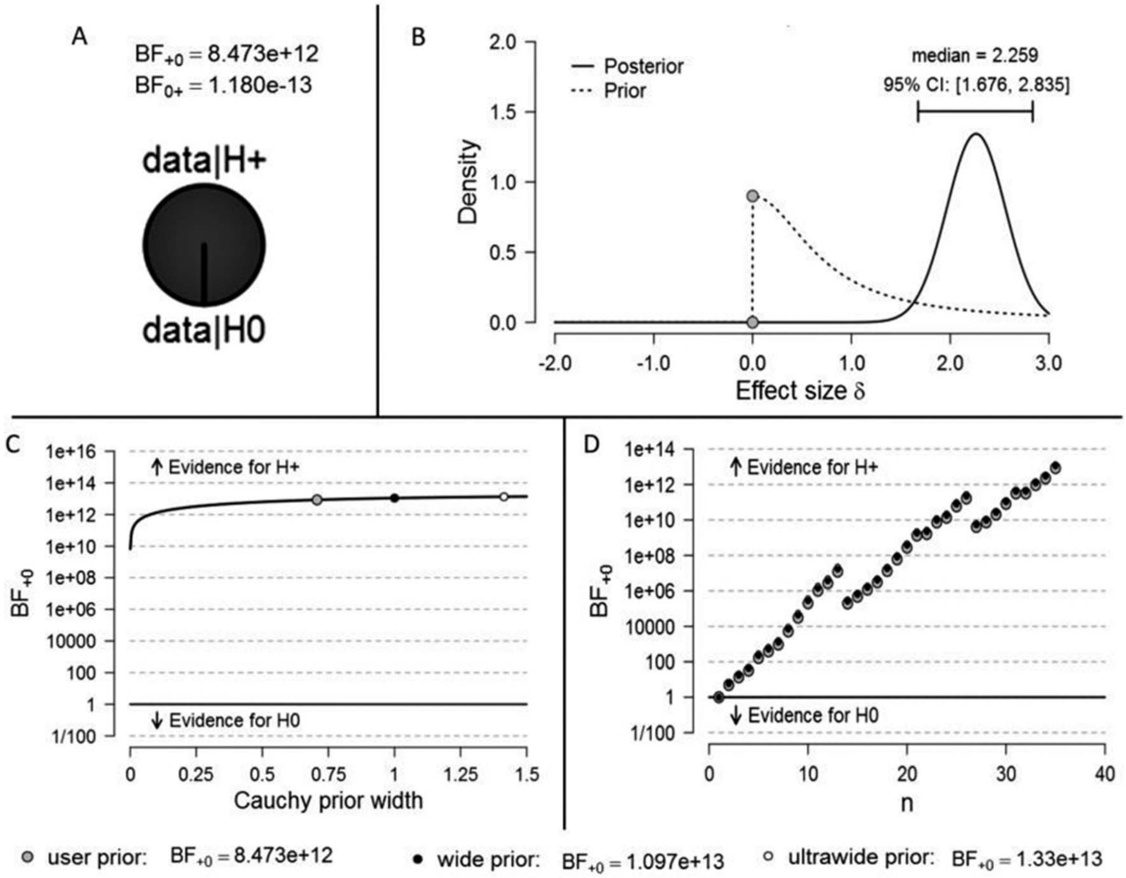

Figure 5. Inferential plots for Study 1, based on Bayesian one-sample *t* tests on overall performance of U.S. participants. (A) Pizza plot providing a proportional representation of the ratio of evidence for the alternative (context) hypothesis to evidence for null (universal) hypothesis. (B) Posterior distribution based on a one-sided prior distribution. Posterior mass to the right of zero indicates that participant performance is consistently above chance-level responding of .5. (C) Bayes factor robustness check. Narrower priors indicate a smaller range of expected effect sizes, favoring the null (universal) hypothesis. (D) Sequential analysis of evidence accumulated over the course of data collection. All figures adjusted from JASP (JASP Team, 2017).

parallel prior studies of *anger*, *fear*, *disgust*, *happiness*, *sadness*, and so forth, in which the classic choice-from-array task produced more consistent evidence of cross-cultural emotion perception than free-labeling and other less constrained tasks (see Gendron, Crivelli, et al., 2018, for a review).

Our findings are consistent with other scientific domains, where it is well-known that the experimental context influences what is observed (e.g., physics: Gleiser, 2015; biology: Lewontin, 2001). Here, as in those scientific domains, the observation is not that context is a contaminating factor that produces demand characteristics, but that contextual factors are authentically part of the phenomena in question. In psychology, emotion perception is typically assumed to be a simple matter of registering or detecting emotional information contained in physical cues such as facial configurations and vocalizations. In contrast, our findings, along with recent published evidence, suggest that perceivers are active meaning makers who infer the emotional meaning in faces and voices, and that context is a crucial part of this process. Experimenters may not intend for their choice of task to be a meaningful part of the context, but mounting evidence suggests that a perceiver's brain treats it this way, nonetheless.

Alternative Interpretations

It may be tempting to interpret our findings as evidence that *gigil*, *glückschmerz*, *greng jai*, *itoshii*, *lajja*, and *liget* are, in fact, universal emotion categories. For example, it has recently been suggested that *fiero*, an Italian concept similar to *liget*, may be universal (Ekman & Cordaro, 2011), and the number of putative universal emotion categories is continually on the rise (Cordaro et al., 2016; Sauter, 2017; see also Cowen & Keltner, 2017, but also L. F. Barrett, Khan, Dy, & Brooks, 2018). Participants' completion of the manipulation check procedure could indeed be seen as demonstrating the universality of these emotion categories. However, such an interpretation fails to consider conceptual combination (Barsalou, 1987), the process by which instances of novel categories can be constructed online by integrating existing knowledge of other emotion concepts that participants possess from their

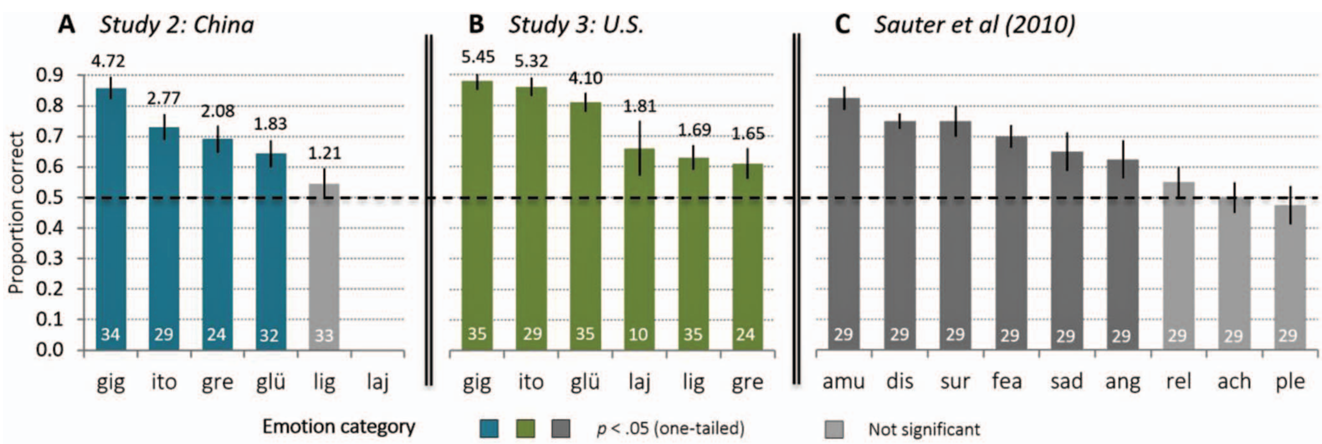

Figure 6. Results from Study 2 (panel A), Study 3 (panel B), compared against Sauter et al. (2010; panel C). Results data from Sauter et al. (2010) were originally presented in terms of mean number of correct responses (out of four trials per emotion category). These data have been replotted in descending order of proportion correct to facilitate direct comparison with the present study. Dashed line indicates chance-level performance (.5). Sample size per emotion category reported in white font at the bottom of each column. Standard error bars (± 1 *SE*) are provided as distributional information only: the location of error bars above the chance line is not indicative of significant above-chance performance because these data were binomially rather than normally distributed. Effect sizes for Studies 2 and 3 are reported above each column using the odds ratio (*OR*). In line with our strong a priori hypotheses, all tests were conducted using one-tailed probability thresholds to avoid Type II errors. Panel A: gig = gigil; gre = greng jai; glü = glückschmerz; ito = itoshii; laj = lajja; lig = liget; Panel B: ach = achievement; amu = amusement; ang = anger; dis = disgust; fea = fear; ple = sensual pleasure; rel = relief; sad = sadness; sur = surprise.

own culture. In addition, the hypothesis that *gigil*, *glückschmerz*, *greng jai*, *itoshii*, *lajja*, and *liget* are universal also fails to explain how participants were able to select the appropriate vocalizations for each category, given that these vocalizations were invented by the experimenters.

It is also possible that our participants have biologically basic emotion concepts for *anger*, *sadness*, *fear*, and so forth, and that they were combining them to perform well on the choice-from-array task (e.g., creating culture-specific emotion blends; Ekman & Cordaro, 2011; Shao, Doucet, & Caruso, 2015). Yet, data collected using less constrained experimental tasks in other small-scale societies (including the same group of Hadza hunter-gatherers) suggests that these emotion categories are not universal (e.g., Crivelli, Jarillo, et al., 2016; Gendron, Hoemann, et al., 2018; Gendron et al., 2014a). Nevertheless, conceptual combination may allow people to experience and perceive emotions across cultural boundaries: it may be a means of *creating* universality by sharing (L. F. Barrett, 2017a). For example, it is plausible that participants in the original studies of cross-cultural emotion perception (e.g., Ekman et al., 1969) were combining concepts from their own cultures to complete choice-from-array tasks. While conceptual combination is not inherently at odds with cross-cultural emotion perception, it is inconsistent with the strongest, traditional version of the universality hypothesis. In that view, emotion perception is an inborn or early-to-develop capacity that is independent of emotion concepts (e.g., Izard, 1994). From this perspective, conceptual combination would not be needed for emotion categories whose universality derives from their biologically basicness and innateness. Although recent accounts of universality have discussed it as a graded phenomenon (e.g., Cordaro et al., 2018; Keltner & Haidt, 1999) that can vary based on culture-specific display rules, decoding rules, and dialects of nonverbal behaviors, even discussions that relax the assumptions of universality still assume that culture-specific experience tunes inborn, fixed action programs. A constructionist account, by contrast, posits conceptual combination as a mechanism by which people acquire emotion concepts and become emotionally acculturated (L. F. Barrett, 2017b), which guides their expressive behaviors from the outset. This hypothesis awaits experimental testing.

Further, it is important to consider the implications of a hypothesis that *gigil*, *glückschmerz*, *greng jai*, *itoshii*, *lajja*, and *liget* are sufficiently translatable from other emotion concepts that participants already possess from their own culture. If we are willing to infer from task performance that emotion categories are "sufficiently translatable," then this interpretation is equally applicable to *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*. Put plainly, if "universal" emotion categories cannot be distinguished from "novel" emotion categories on the basis of performance on the classic choice-from-array task, then this task is not an adequate method for assessing universality, calling the interpretation of hundreds of prior studies into question.

A related line of interpretation is that participants were familiar enough with the emotional experiences associated with the novel emotion categories that they could successfully complete the task, even though the six concepts we used are not encoded in Hadzane, Mandarin Chinese, or English by unique words. That participants were able to bring their prior experience to bear in completing the task is not at issue: it is precisely through mechanisms such as conceptual combination that, we propose, humans are able to gain a basic understanding of each other's emotional state. Therefore, the weakest possible interpretation of the current findings is that the classic choice-from-array task faithfully primes (but does not enhance) participants' ability to make appropriate meaning out of situated nonverbal cues, and that we have only succeeded in extending the range of situations and cues. The history of published studies shows, however, that highly structured choice-from-array tasks provide support for cross-cultural emotion perception when other methods do not, calling into question the robustness and replicability of evidence for universal emotions.

A final alternative interpretation of the current findings is that the six novel emotion categories, along with their stipulated vocalizations, are in fact subordinate members of so-called "basic" emotions (e.g., *gigil* is a form of *happiness*). The results of our free-labeling data for vocalizations and long-form scenarios across all three studies do not immediately support this interpretation. Overall, participants in all three cultural samples provided labels for the novel vocalizations that were general affective descriptions such as "good" or "bad," or offered words such as "love" that do not correspond with *anger*, *disgust*, *fear*, *happiness*, *sadness*, or *surprise*. In contrast, a notable proportion of participants consistently provided labels for three novel emotion scenarios that corresponded with presumed universal categories: *gigil* with *happiness* (Hadza, 80% of labels; China, 61%; United States, 29%), *itoshii* with *sadness* (Hadza, 70% of labels; China, 47%; United States, 35%), and *glückschmerz* with *anger* (Hadza, 33% of labels; China, 23%; United States, 24%). These labels were not necessarily used in a specific way, however. *Happiness* was also frequently associated with other novel emotion scenarios (e.g., Hadza, 83% of *liget* labels; China, 13% of *lajja* labels), as was *sadness* (e.g., Hadza, 40% of *greng jai* labels; China, 8% of *gluckschmerz* labels; see Figure 11 in the online supplemental material for details). More important, this pattern of findings is also consonant with an interpretation of conceptual combination: the labels for novel vocalizations and novel scenarios did not consistently correspond with the same emotion categories, implying that performance on the choice-from-array task may not be driven by conceptual labeling of the scenario alone. To perform well on the task, participants would still have needed to extend any preexisting categories (perhaps by using conceptual combination) to accommodate novel vocalizations.

Processes Supporting Cross-Cultural Emotion Perception

If the experimental context is full of psychologically potent features that can influence how people infer meaning in vocalizations, then our findings have broader implications for the study of emotion perception. Certainly our findings join others in casting doubt on the claim that cross-cultural emotion perception is "an established axiom of behavioral science" (Izard & Saxton, 1988, pp. 651–652). However, beyond the potential inadequacy of the classic choice-from-array task for testing the universality hypothesis, our findings offer indirect support for the hypothesis that emotion perception in the real world is the result of multiple processes, such as identifying similarities between someone's physical changes (facial movements, vocal acoustics, etc.) in a particular situational context and prior experiences from the past, using knowledge of emotion words and concepts, using process-of-elimination strategies, as well as perceiving affect and learning categories (for a discussion, see L. F. Barrett, 2017a). These processes may function like ingredients that contribute to different recipes for emotion perception in different cultures.

The current studies do not provide systematic evidence for what these processes are or how they work, individually or synergistically. Instead, they can be thought of as "proof of concept" that such studies are sorely needed and worth the investment. This was also the conclusion of a recent review of scientific evidence on inferring emotion in human facial movements, to be published in *Psychological Science in the Public Interest*:

> The science of emotion expression and emotion perception has been more a science of stereotypes rather than a science of how people actually move their faces to express emotion and the processes by which those movements carry information about emotion to someone else (a perceiver). . . . In reality, emotions are expressed with facial movements that are more variable and context-dependent. . . . Their context-dependence goes well beyond display rules or cultural accents. As a consequence, the stereotypes . . . must be replaced by a thriving scientific effort to observe and describe the lexicon of context-sensitive ways in which people move their facial muscles to express emotion, and the discovery of when and how people infer emotions in other people's facial movements. (L. F. Barrett et al., in press, p. 114 of the article draft)

Following published research, we would suggest the same insights hold true for vocalizations. The present studies, while not conclusively revealing which processes should be the target of empirical focus, do make several suggestions.

Cognitive bootstrapping. The experimental features of forced-choice designs, including the classic choice-from-array task, can be cognitively bootstrapped (e.g., by identifying perceptual similarities or using process-of-elimination strategies; Russell, 1994) to promote online category learning. Developmental studies suggest that cognitive bootstrapping underlies successful performance on a wide range of experimental tasks (e.g., Cassels & Birch, 2014; Diesendruck, Hall, & Graham, 2006; Haryu, Imai, & Okada, 2011; Markman & Wachtel, 1988; Waxman & Booth, 2001). In recent studies of emotion perception, participants have been shown to use process-of-elimination strategies when presented with novel emotion words and facial configurations. These strategies are used in selecting a response option within a given trial (Nelson & Russell, 2016a), in tracking previously selected response options across trials (DiGirolamo & Russell, 2017), and in freely labeling stimuli previously presented in a separate task (Nelson & Russell, 2016b).

Affect perception. The inference of affective meaning may have contributed to our observed effects and may also be an important feature in emotion perception. There is ample evidence that facial and vocal cues are perceived in terms of the valence and the level of arousal that they communicate (L. F. Barrett & Bliss-Moreau, 2009; Russell & Barrett, 1999). Affect perception is robust across cultures (Russell, 1991; Russell et al., 2003; Russell & Barrett, 1999), in children who do not possess explicit emotion concept knowledge (Widen, 2016), and in patients who have lost emotion knowledge because of semantic dementia (Lindquist et al., 2014). For vocalizations, specific acoustic features (e.g., fun-

[6] Note that the vocalizations tested were not normed for actual physiological activation; instead, level of arousal was inferred based on the context in which the vocalizations were produced.

damental frequency and amplitude) are reliably associated with the perception of arousal (Bachorowski, 1999; Bachorowski & Owren, 2008; but see Scherer, Johnstone, & Klasmeyer, 2003 for a review of a discrete emotions account of acoustic features). Recent work on arousal perception suggests that these acoustic features hold across species (Filippi et al., 2017).[6]

Category learning. Participants may also have leveraged conceptual features of the experimental context, such as emotion words, to complete the task of emotion perception. In this way, our findings hold clues to improving cross-cultural emotion communication, in that they suggest a view of emotion perception as culturally dependent upon concepts that are acquired through category learning. A growing body of work demonstrates that words serve as invitations to form abstract categories with limited perceptual regularity across instances (e.g., Ferry et al., 2010). Mounting evidence from studies of the face indicate that emotion categories are abstract in that their instances are highly variable across situations (L. F. Barrett et al., in press), as do studies of psychophysiology and brain imaging (e.g., Wilson-Mendenhall, Barrett, & Barsalou, 2015). Our choice-from-array task, based on that used by Sauter et al. (2010, 2015), contained design features that may have allowed participants to quickly learn novel emotion categories when labeled with words, such that they achieved levels of performance equivalent to those reported in support of cross-cultural emotion perception. This pattern of performance is consistent with a large body of findings from the developmental psychology literature showing that children and even young infants can learn novel, abstract categories with the help of words (e.g., Ferry et al., 2010; Waxman & Booth, 2001; Xu, Cote, & Baker, 2005; Yin & Csibra, 2015).

Conceptual combination. We hypothesize that category learning may occur in the context of a brief experimental task, as in real life, through the process of conceptual combination. In our study, better performance may have been observed for those categories that were easier to construct via combination of the knowledge and experiences promoted by a given culture. Ease of conceptual combination may also be reflected by the number of participants who passed the manipulation check for a novel emotion category. In such cases, participants will not necessarily understand the concept exactly as a native would; however, conceptual combination may allow for some cross-cultural communication, albeit imperfect. As such, our findings are consistent with the hypothesis that conceptual combination may be the foundation of category learning and, thereby, of cross-cultural emotion communication. This is a hypothesis in need of further scientific investigation.

The specific pattern of performance observed across the three samples suggests that preexisting cultural knowledge may have played a role in task performance. We speculate that participants were more easily able to learn novel emotion categories that fit local cultural values and practices, consistent with research on cultural fit and emotional values (Richerson & Boyd, 2005; Tamir et al., 2016). For example, the concept of *greng jai*, which describes the combination of gratitude and social guilt one feels when offered an overly generous gift or burdensome favor, likely does not fit as well with Hadza cultural practices of resource sharing and expectations of communal collaboration (Apicella, Marlowe, Fowler, & Christakis, 2012), as it does with Chinese norms associated with maintaining and saving face (Chang & Holt, 1994). Likewise, we can hypothesize that U.S. participants are anchoring on how the experience of *greng jai* clashes with the core cultural value of personal independence (Markus & Kitayama, 1991). Future research could further develop a priori hypotheses to test how category learning is impacted by conceptual fit along these and other cultural dimensions.

Conclusion

The patterns of performance we observed in the present studies suggest a new context within which to integrate the hundreds of published studies using a choice-from-array task to test, and ultimately provide support for, the hypothesis that certain emotion categories are universally perceived. Participants from three cultural samples selected stipulated (i.e., made-up) target vocalizations for unfamiliar emotion categories at levels exceeding chance, suggesting that certain experimental design features may facilitate emotional meaning-making, even when the emotion concepts and the vocalizations are novel (i.e., not consistently and specifically associated with preexisting emotion categories). Accordingly, our findings invite discussions about the psychological potency of experimental design features in a task that is pervasive in psychological research, as well as the meaning-making processes that undergird emotion perception. Furthermore, our findings build on previously published studies in suggesting that the variety of processes that contribute to emotion perception, and social perception more generally, may be differentially recruited across cultural contexts.

Our findings also suggest a deeper point about ecological validity. Humans make meaning of their environment, usually as an automatic, effortless, and obligatory consequence of the way they process information (L. F. Barrett, 2017b). This meaning-making is not suspended in the context of an experiment; rather, features of this context may facilitate it, becoming psychologically potent in a way that may or may not be representative of everyday life. To acknowledge and account for this, studies of cross-cultural emotion perception must compare findings across multiple methods (e.g., Crivelli, Jarillo, et al., 2016; Crivelli, Russell, Jarillo, & Fernández-Dols, 2017; Gendron et al., 2014a), sample spontaneous behavior from naturally occurring interactions (e.g., Crivelli, Carrera, & Fernández-Dols, 2015; Fernández-Dols & Ruiz-Belda, 1995; Tracy & Matsumoto, 2008), and explore how domain-general processes such as category learning and conceptual combination may influence performance. Together, these steps will lead to a more robust, nuanced, and replicable science of human behavior, including emotion perception.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York, NY: Wiley. http://dx.doi.org/10.1002/0471249688

Anderson, E., Siegel, E. H., & Barrett, L. F. (2011). What you feel influences what you see: The role of affective feelings in resolving binocular rivalry. *Journal of Experimental Social Psychology, 47,* 856–860. http://dx.doi.org/10.1016/j.jesp.2011.02.009

Apicella, C. L., & Crittenden, A. N. (2015). Hunter-gatherer families and parenting. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (Vol. IV, pp. 578–597). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9781119125563.evpsych123

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature, 481,* 497–501. http://dx.doi.org/10.1038/nature10736

Audacity Team. (2015). Audacity (R): Free audio edition and recorder (Version 2. 1.2) [Computer software]. Retrieved from http://www.audacityteam.org/

Aviezer, H., & Hassin, R. (2017). Inherently ambiguous: An argument for contextualized emotion perception. In J. M. Fernandez-Dols & J. A. Russell (Eds.), *The science of facial expression*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780190613501.003.0018

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., . . . Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science, 19,* 724–732. http://dx.doi.org/10.1111/j.1467-9280.2008.02148.x

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

Bachorowski, J.-A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8,* 53–57. http://dx.doi.org/10.1111/1467-8721.00013

Bachorowski, J.-A., & Owren, M. J. (2008). Vocal expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 196–210). New York, NY: Guilford Press.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5,* 617–629. http://dx.doi.org/10.1038/nrn1476

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., . . . Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 4688–4693. http://dx.doi.org/10.1073/pnas.1522070113

Barrett, L. F. (2017a). *How emotions are made: The secret life the brain and what it means for your health, the law, and human nature*. New York, NY: Houghton Mifflin Harcourt.

Barrett, L. F. (2017b). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12,* 11833. http://dx.doi.org/10.1093/scan/nsx060

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A., & Pollak, S. (in press). Emotional expressions reconsidered: Challenges to inferring emotion in human facial movements. *Psychological Science in the Public Interest*.

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in Experimental Social Psychology, 41,* 167–218. http://dx.doi.org/10.1016/S0065-2601(08)00404-8

Barrett, L. F., Khan, Z., Dy, J., & Brooks, D. (2018). Nature of emotion categories: Comment on Cowen and Keltner. *Trends in Cognitive Sciences, 22,* 97–99. http://dx.doi.org/10.1016/j.tics.2017.12.004

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science, 20,* 286–290. http://dx.doi.org/10.1177/0963721411422522

Barrett, L. F., & Niedenthal, P. M. (2004). Valence focus and the perception of facial affect. *Emotion, 4,* 266–274. http://dx.doi.org/10.1037/1528-3542.4.3.266

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts reconsidered: The ecological and intellectual bases of categories* (pp. 389–412). New York, NY: Cambridge University Press.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40,* 531–539. http://dx.doi.org/10.3758/BRM.40.2.531

Berbesque, J. C., Wood, B. M., Crittenden, A. N., Mabulla, A., & Marlowe, F. W. (2016). Eat first, share later: Hadza hunter-gatherer men consume more while foraging than in central places. *Evolution and Human Behavior, 37,* 281–286. http://dx.doi.org/10.1016/j.evolhumbehav.2016.01.003

Berger, J. O., & Mortera, J. (1999). Default Bayes factors for nonnested hypothesis testing. *Journal of the American Statistical Association, 94,* 542–554. http://dx.doi.org/10.1080/01621459.1999.10474149

Berger, J. O., & Wolpert, R. L. (1988). *The likelihood principle* (2nd ed.). Haywood, CA: The Institute of Mathematical Sciences.

BetterThanEnglish. (2012). *Better than English* [Online database]. Retrieved from http://betterthanenglish.com/

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1,* 185–216. http://dx.doi.org/10.1177/135910457000100301

Bruner, J. S. (1957). Going beyond the information given. *Contemporary Approaches to Cognition, 1,* 119–160.

Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture, 8,* 135–148. http://dx.doi.org/10.1163/156770908X289242

Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., . . . Zhou, Y. (2016). Detecting affiliation in colaughter across 24 societies. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 4682–4687. http://dx.doi.org/10.1073/pnas.1524993113

Calbi, M., Angelini, M., Gallese, V., & Umiltà, M. A. (2017). "Embodied Body Language": An electrical neuroimaging study with emotional faces and bodies. *Scientific Reports, 7,* 6875. http://dx.doi.org/10.1038/s41598-017-07262-0

Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology, 70,* 205–218. http://dx.doi.org/10.1037/0022-3514.70.2.205

Cassels, T. G., & Birch, S. A. (2014). Comparisons of an open-ended vs. forced-choice 'mind reading' task: Implications for measuring perspective-taking and emotion recognition. *PLoS ONE, 9,* e93653. http://dx.doi.org/10.1371/journal.pone.0093653

Chanes, L., Wormwood, J. B., Betz, N., & Barrett, L. F. (2018). Facial expression predictions as drivers of social perception. *Journal of Personality and Social Psychology, 114,* 380–396. http://dx.doi.org/10.1037/pspa0000108

Chang, H.-C., & Holt, G. R. (1994). A Chinese perspective on face as inter-relational concern. In S. Ting-Toomey (Ed.), *The challenge of facework: Cross-cultural and interpersonal issues* (pp. 95–132). Albany, NY: State University of New York Press.

Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology, 53,* 751–766. http://dx.doi.org/10.1037/0022-3514.53.4.751

Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion, 16,* 117–128. http://dx.doi.org/10.1037/emo0000100

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18,* 75–93. http://dx.doi.org/10.1037/emo0000302

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences of the United States of America, 114,* E7900–E7909. http://dx.doi.org/10.1073/pnas.1702247114

Crivelli, C., Carrera, P., & Fernández-Dols, J.-M. (2015). Are smiles a sign of happiness? Spontaneous expressions of judo winners. *Evolution and Human Behavior, 36,* 52–58. http://dx.doi.org/10.1016/j.evolhumbehav.2014.08.009

Crivelli, C., & Gendron, M. (2017a). Facial expressions and emotions in indigenous societies. In J. A. Russell & J.-M. Fernández-Dols (Eds.), *The science of facial expression* (pp. 497–515). Oxford, UK: Oxford

University Press. http://dx.doi.org/10.1093/acprof:oso/9780190613501.003.0026

Crivelli, C., & Gendron, M. (2017b). Facial expressions and emotions in indigenous societies. In J. M. Fernandez-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 497–515). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780190613501.003.0026

Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J. M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General, 145,* 830–843. http://dx.doi.org/10.1037/xge0000172

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J.-M. (2016). The fear gasping face as a threat display in a Melanesian society. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 12403–12407. http://dx.doi.org/10.1073/pnas.1611622113

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J.-M. (2017). Recognizing spontaneous facial expressions of emotion in a small-scale society of Papua New Guinea. *Emotion, 17,* 337–347. http://dx.doi.org/10.1037/emo0000236

Darwin, C. (2005). *The expression of the emotions in man and animals*. London, UK: John Murray. Digireads.com Publishing. Retrieved from http://darwin-online.org.uk/ (Original work published 1872)

Dashiell, J. F. (1927). A new method of measuring reactions to facial expression of emotion. *Psychological Bulletin, 24,* 174–175.

De Boinod, A. J. (2007). *The meaning of tingo: And other extraordinary words from around the world*. London, UK: Penguin.

de Gelder, B. (2016). Emotional Body Perception in the Wild. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (4 ed., pp. 483–494): New York, NY: Guilford Press Publications.

Diesendruck, G., Hall, D. G., & Graham, S. A. (2006). Children's use of syntactic and pragmatic knowledge in the interpretation of novel adjectives. *Child Development, 77,* 16–30. http://dx.doi.org/10.1111/j.1467-8624.2006.00853.x

DiGirolamo, M. A., & Russell, J. A. (2017). The emotion seen in a face can be a methodological artifact: The process of elimination hypothesis. *Emotion, 17,* 538–546. http://dx.doi.org/10.1037/emo0000247

Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General, 147,* 62–73. http://dx.doi.org/10.1037/xge0000361

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242. http://dx.doi.org/10.1037/h0044139

Ekman, P. (1972). *Universals and cultural differences in facial expressions of emotion*. Paper presented at the Nebraska Symposium on Motivation, 1971.

Ekman, P. (2007). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York, NY: Macmillan.

Ekman, P. (2017). Facial expressions. In J. M. Fernandez-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 39–56). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780190613501.003.0003

Ekman, P., & Cordaro, D. T. (2011). What is meant by calling emotions basic. *Emotion Review, 3,* 364–370. http://dx.doi.org/10.1177/1754073911410740

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124–129. http://dx.doi.org/10.1037/h0030377

Ekman, P., Heider, E., Friesen, W. V., & Heider, K. (1972). Facial expression in a preliterate culture. *Psychological Bulletin, 115,* 268–287. http://dx.doi.org/10.1037/0033-2909.115.2.268

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221,* 1208–1210. http://dx.doi.org/10.1126/science.6612338

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164,* 86–88. http://dx.doi.org/10.1126/science.164.3875.86

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 203–235. http://dx.doi.org/10.1037/0033-2909.128.2.203

Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modeling of group differences. *Review of General Psychology, 17,* 111–121. http://dx.doi.org/10.1037/a0030048

Fernández-Dols, J. M., & Ruiz-Belda, M.-A. (1995). Are smiles a sign of happiness? Gold medal winners at the Olympic Games. *Journal of Personality and Social Psychology, 69,* 1113–1119. http://dx.doi.org/10.1037/0022-3514.69.6.1113

Fernández-Dols, J. M., & Ruiz-Belda, M.-A. (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. *The Psychology of Facial Expression, 2,* 255–274.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development, 81,* 472–479. http://dx.doi.org/10.1111/j.1467-8624.2009.01408.x

Filippi, P., Congdon, J., Hoang, J., Bowling, D., Reber, S., PAšukonis, A., . . . Sturdy, C. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings of the Royal Society Section B: Biological Sciences, 284,* 20170990. http://dx.doi.org/10.1098/rspb.2017.0990

Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 237–253). New York, NY: Russell Sage Foundation.

Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology, 60,* 229–240. http://dx.doi.org/10.1037/0022-3514.60.2.229

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.

Fugate, J., Gendron, M., Nakashima, S., & Barrett, L. F. (2018). Emotion words: Adding face value. *Emotion, 18,* 693–706. http://dx.doi.org/10.1037/emo0000330

Fugate, J. M., Gouzoules, H., & Barrett, L. F. (2010). Reading chimpanzee faces: Evidence for the role of verbal labels in categorical perception of emotion. *Emotion, 10,* 544–554. http://dx.doi.org/10.1037/a0019017

Gendron, M. (2017). Revisiting diversity: Cultural variation reveals the constructed nature of emotion perception. *Current Opinion in Psychology, 17,* 145–150. http://dx.doi.org/10.1016/j.copsyc.2017.07.014

Gendron, M., & Barrett, L. F. (2009). Reconstructing the past: A century of ideas about emotion in psychology. *Emotion Review, 1,* 316–339. http://dx.doi.org/10.1177/1754073909338877

Gendron, M., & Barrett, L. F. (2017). Facing the past: A history of the face in psychological research on emotion perception. In J. M. Fernandez-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 15–36). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780190613501.003.0002

Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality reconsidered: Diversity in meaning making about facial expressions. *Current Directions in Psychological Science*. Advance online publication. http://dx.doi.org/10.1177/0963721417746794

Gendron, M., Hoemann, K., Crittenden, A. N., Msafiri, S., Ruark, G., & Barrett, L. F. (2018). *Emotion perception in Hadza hunter-gatherers: A test of evolutionary hypotheses about universal emotions*. Manuscript submitted for publication.

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

Gendron, M., Mesquita, B., & Barrett, L. F. (2013). Emotion perception:

Putting the face in context. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 379–389). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780195376746.013.0034

Gendron, M., Roberson, D., & Barrett, L. F. (2015). Cultural variation in emotion perception is real: A response to Sauter, Eisner, Ekman, and Scott (2015). *Psychological Science, 26,* 357–359. http://dx.doi.org/10.1177/0956797614566659

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science, 25,* 911–920. http://dx.doi.org/10.1177/0956797613517239

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

Gewald, J.-B. (2010). Remote but in contact with history and the world. *Proceedings of the National Academy of Sciences of the United States of America, 107,* E75. http://dx.doi.org/10.1073/pnas.1001284107

Gleiser, M. (2015). *The island of knowledge: The limits of science and the search for meaning*. New York, NY: Basic Books.

Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology, 26,* 441–462. http://dx.doi.org/10.1146/annurev.soc.26.1.441

Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics, 69,* 839–855. http://dx.doi.org/10.3758/BF03193921

Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition and Emotion, 13,* 225–266. http://dx.doi.org/10.1080/026999399379267

Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action-based verb extension. *Child Development, 82,* 674–686. http://dx.doi.org/10.1111/j.1467-8624.2010.01567.x

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466,* 29. http://dx.doi.org/10.1038/466029a

Hess, U., & Hareli, S. (2017). The social signal value of emotions: The role of contextual factors in social inferences drawn from emotion displays. In J. M. Fernandez-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 375–393). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780190613501.003.0020

Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *WIREs, 4,* 93–103. http://dx.doi.org/10.1002/wcs.1203

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115,* 288–299. http://dx.doi.org/10.1037/0033-2909.115.2.288

Izard, C. E., & Saxton, P. M. (1988). Emotions. In R. C. Atkinson, R. J. Hernstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology: Perception and motivation; learning and cognition* (2 ed., Vol.1–2, pp. 627–676). Oxford, England: Wiley.

Jack, R. E., Sun, W., Delis, I., Garrod, O. G., & Schyns, P. G. (2016). Four not six: Revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General, 145,* 708–730. http://dx.doi.org/10.1037/xge0000162

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007

JASP Team. (2017). JASP (Version 0. 8.1.2) [Computer software]. Retrieved from https://jasp-stats.org/

Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology, 5,* 1–10. http://dx.doi.org/10.20982/tqmp.05.1.p001

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Jones, N. B. (2016). *Demography and evolutionary ecology of Hadza hunter-gatherers*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781107707030

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion, 13,* 505–521. http://dx.doi.org/10.1080/026999399379168

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods, 8,* 115–128. http://dx.doi.org/10.1037/1082-989X.8.2.115

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W., . . . Thingujam, N. S. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology, 4,* 353.

Lewontin, R. C. (2001). *The triple helix: Gene, organism, and environment*. Cambridge, MA: Harvard University Press.

Lin, P.-Y. (2013, January 23). *Unspeakableness: An intervention of language evolution and human communication*. Retrieved from http://uniquelang.peiyinglin.net/01untranslatable.html

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion, 14,* 375–387. http://dx.doi.org/10.1037/a0035293

Lomas, T. (2016). Towards a positive cross-cultural lexicography: Enriching our emotional landscape through 216 'untranslatable' words pertaining to well-being. *The Journal of Positive Psychology, 11,* 546–558. http://dx.doi.org/10.1080/17439760.2015.1127993

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science, 18,* 1077–1083. http://dx.doi.org/10.1111/j.1467-9280.2007.02028.x

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20,* 121–157. http://dx.doi.org/10.1016/0010-0285(88)90017-5

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98,* 224–253. http://dx.doi.org/10.1037/0033-295X.98.2.224

Marlowe, F. W. (2010). *The Hadza: Hunter-gatherers of Tanzania* (Vol. 3). Berkeley, CA: University of California Press.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics, 34,* 338–348. http://dx.doi.org/10.3758/BF03203046

Menon, U., & Shweder, R. A. (1994). Kali's tongue: Cultural psychology and the power of shame in Orissa, India. In S. Kitayama & H. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 241–284). Washington, DC: American Psychological Association.

Mobbs, D., Weiskopf, N., Lau, H. C., Featherstone, E., Dolan, R. J., & Frith, C. D. (2006). The Kuleshov Effect: The influence of contextual framing on emotional attributions. *Social Cognitive and Affective Neuroscience, 1,* 95–106. http://dx.doi.org/10.1093/scan/nsl014

Mohanty, A., & Sussman, T. J. (2013). Top-down modulation of attention by emotion. *Frontiers in Human Neuroscience, 7,* 102. http://dx.doi.org/10.3389/fnhum.2013.00102

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). *BayesFactor: Computation of Bayes factors for common designs* (R package version 0.9.12–2). Retrived from https://rdrr.io/cran/BayesFactor/man/ttestBF.html

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A, 135,* 370–384. http://dx.doi.org/10.2307/2344614

Nelson, N. L., & Russell, J. A. (2016a). Building emotion categories: Children use a process of elimination when they encounter novel expressions. *Journal of Experimental Child Psychology, 151,* 120–130. http://dx.doi.org/10.1016/j.jecp.2016.02.012

Nelson, N. L., & Russell, J. A. (2016b). A facial expression of pax: Assessing children's "recognition" of emotion from faces. *Journal of Experimental Child Psychology, 141,* 49–64. http://dx.doi.org/10.1016/j.jecp.2015.07.016

Norenzayan, A., & Heine, S. J. (2005). Psychological universals: What are they and how can we know? *Psychological Bulletin, 131,* 763–784. http://dx.doi.org/10.1037/0033-2909.131.5.763

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology, 52,* 629–651. http://dx.doi.org/10.1146/annurev.psych.52.1.629

Pinker, S. (1997). *How the mind works*. New York, NY: Norton.

Raichlen, D. A., Pontzer, H., Harris, J. A., Mabulla, A. Z., Marlowe, F. W., Josh Snodgrass, J., . . . Wood, B. M. (2017). Physical activity patterns and biomarkers of cardiovascular disease risk in hunter-gatherers. *American Journal of Human Biology, 29,* e22919. http://dx.doi.org/10.1002/ajhb.22919

Raudenbush, S., Bryk, A., Cheong, Y., Congdon, R., & du Toit, M. (2004). HLM 6 hierarchical linear and nonlinear modeling [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Raudenbush, S., Bryk, A., & Congdon, R. (2011). HLM 7.00 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago, IL: University of Chicago Press.

Rosaldo, M. Z. (1980). *Knowledge and passion*. New York, NY: Cambridge University Press.

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review, 21,* 301–308. http://dx.doi.org/10.3758/s13423-014-0595-4

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56,* 356–374. http://dx.doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

Ruiz-Belda, M.-A., Fernández-Dols, J.-M., Carrera, P., & Barchard, K. (2003). Spontaneous facial expressions of happy bowlers and soccer fans. *Cognition and Emotion, 17,* 315–326. http://dx.doi.org/10.1080/02699930302288

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178. http://dx.doi.org/10.1037/h0077714

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin, 110,* 426–450. http://dx.doi.org/10.1037/0033-2909.110.3.426

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115,* 102–141. http://dx.doi.org/10.1037/0033-2909.115.1.102

Russell, J. A. (1995). Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin, 118,* 379–391. http://dx.doi.org/10.1037/0033-2909.118.3.379

Russell, J. A., Bachorowski, J. A., & Fernandez-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54,* 329–349. http://dx.doi.org/10.1146/annurev.psych.54.101601.145102

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76,* 805–819. http://dx.doi.org/10.1037/0022-3514.76.5.805

Sands, B. (1998). The linguistic relationship between Hadza and Khoisan. In M. Schladt (Ed.), *Language, identity, and conceptualization among the Khoisan* (pp. 265–283). Köln, Germany: Rudiger Kupper Verlag.

Sauter, D. A. (2017). The nonverbal communication of positive emotions: An emotion family approach. *Emotion Review, 9,* 222–234. http://dx.doi.org/10.1177/1754073916667236

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology, 63,* 2251–2272. http://dx.doi.org/10.1080/17470211003721642

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,* 2408–2412. http://dx.doi.org/10.1073/pnas.0908239106

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2015). Emotional vocalizations are recognized across cultures regardless of the valence of distractors. *Psychological Science, 26,* 354–356. http://dx.doi.org/10.1177/0956797614560771

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433–456). New York, NY: Oxford University Press.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication, 40,* 99–116. http://dx.doi.org/10.1016/S0167-6393(02)00078-X

Shao, B., Doucet, L., & Caruso, D. R. (2015). Universality versus cultural specificity of three emotion domains: Some evidence based on the cascading model of emotional intelligence. *Journal of Cross-Cultural Psychology, 46,* 229–251. http://dx.doi.org/10.1177/0022022114557479

Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for? *Current Directions in Psychological Science, 20,* 395–399. http://dx.doi.org/10.1177/0963721411424739

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52,* 1061–1086. http://dx.doi.org/10.1037/0022-3514.52.6.1061

Sorenson, E. R. (1975). Culture and the expression of emotion. In T. R. Williams (Ed.), *Psychological anthropology* (pp. 361–372). Chicago, IL: Aldine. http://dx.doi.org/10.1515/9783110802818.361

Tamir, M., Schwartz, S. H., Cieciuch, J., Riediger, M., Torres, C., Scollon, C., . . . Vishkin, A. (2016). Desired emotions across cultures: A value-based account. *Journal of Personality and Social Psychology, 111,* 67–82. http://dx.doi.org/10.1037/pspp0000072

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of emotions and their relationship to internal regulatory variables. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3 ed., pp. 114–137). New York, NY: Guilford Press.

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 11655–11660. http://dx.doi.org/10.1073/pnas.0802686105

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3,* 397–405. http://dx.doi.org/10.1177/1754073911410747

Tracy, J. L., & Robins, R. W. (2008). The nonverbal expression of pride: Evidence for cross-cultural recognition. *Journal of Personality and Social Psychology, 94,* 516–530. http://dx.doi.org/10.1037/0022-3514.94.3.516

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion, 7,* 487–494. http://dx.doi.org/10.1037/1528-3542.7.3.487

Vogt, J., De Houwer, J., Moors, A., Van Damme, S., & Crombez, G. (2010). The automatic orienting of attention to goal-relevant stimuli. *Acta Psychologica, 134,* 61–69. http://dx.doi.org/10.1016/j.actpsy.2009.12.006

Wagenmakers, E.-J., Beek, T. F., Rotteveel, M., Gierholz, A., Matzke, D., Steingroever, H., . . . Pinto, Y. (2015). Turning the hands of time again: A purely confirmatory replication study and a Bayesian analysis. *Frontiers in Psychology, 6,* 494. http://dx.doi.org/10.3389/fpsyg.2015.00494

Wallbott, H. G. (1988). In and out of context: Influences of facial expression and context information on emotion attributions. *British Journal of*

Social Psychology, 27, 357–369. http://dx.doi.org/10.1111/j.2044-8309.1988.tb00837.x

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45,* 1191–1207. http://dx.doi.org/10.3758/s13428-012-0314-x

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology, 43,* 217–242. http://dx.doi.org/10.1006/cogp.2001.0764

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6,* 291–298. http://dx.doi.org/10.1177/1745691611406923

Wetzels, R., Raaijmakers, J. G., Jakab, E., & Wagenmakers, E.-J. (2009). How to quantify support for and against the null hypothesis: A flexible WinBUGS implementation of a default Bayesian t test. *Psychonomic Bulletin & Review, 16,* 752–760. http://dx.doi.org/10.3758/PBR.16.4.752

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

Widen, S. C., & Russell, J. A. (2013). Children's recognition of disgust in others. *Psychological Bulletin, 139,* 271–299. http://dx.doi.org/10.1037/a0031640

Wieser, M. J., & Brosch, T. (2012). Faces in context: A review and systematization of contextual influences on affective face processing. *Frontiers in Psychology, 3,* 471. http://dx.doi.org/10.3389/fpsyg.2012.00471

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2015). Variety in emotional life: Within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Social Cognitive and Affective Neuroscience, 10,* 62–71. http://dx.doi.org/10.1093/scan/nsu037

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica, 132,* 173–189. http://dx.doi.org/10.1016/j.actpsy.2009.02.002

Xu, F., Cote, M., & Baker, A. (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science, 16,* 372–377. http://dx.doi.org/10.1111/j.0956-7976.2005.01543.x

Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. *Psychological Science, 26,* 1316–1324. http://dx.doi.org/10.1177/0956797615588753

Young, F. W., & Lewyckyj, R. (1979). *ALSCAL user's guide: A guide for users of ALSCAL: A nonmetric multidimensional scaling and unfolding program with several individual differences options*. Chapel Hill, NC: Data Analysis and Theory Associates.

Received February 23, 2017
Revision received July 3, 2018
Accepted July 9, 2018 ■