American Psychologist

Mapping 24 Emotions Conveyed by Brief Human Vocalization

Alan S. Cowen, Hillary Anger Elfenbein, Petri Laukka, and Dacher Keltner

Mapping 24 Emotions Conveyed by Brief Human Vocalization

Alan S. Cowen
University of California, Berkeley

Hillary Anger Elfenbein
Washington University in St. Louis

Petri Laukka
Stockholm University

Dacher Keltner
University of California, Berkeley

Emotional vocalizations are central to human social life. Recent studies have documented that people recognize at least 13 emotions in brief vocalizations. This capacity emerges early in development, is preserved in some form across cultures, and informs how people respond emotionally to music. What is poorly understood is how emotion recognition from vocalization is structured within what we call a semantic space, the study of which addresses questions critical to the field: How many distinct kinds of emotions can be expressed? Do expressions convey emotion categories or affective appraisals (e.g., valence, arousal)? Is the recognition of emotion expressions discrete or continuous? Guided by a new theoretical approach to emotion taxonomies, we apply large-scale data collection and analysis techniques to judgments of 2,032 emotional vocal bursts produced in laboratory settings (Study 1) and 48 found in the real world (Study 2) by U.S. English speakers ($N = 1,105$). We find that vocal bursts convey at least 24 distinct kinds of emotion. Emotion categories (sympathy, awe), more so than affective appraisals (including valence and arousal), organize emotion recognition. In contrast to discrete emotion theories, the emotion categories conveyed by vocal bursts are bridged by smooth gradients with continuously varying meaning. We visualize the complex, high-dimensional space of emotion conveyed by brief human vocalization within an online interactive map.

Keywords: emotion, voice, affect, computational methods, semantic space

Supplemental materials: http://dx.doi.org/10.1037/amp0000399.supp

It is something of an anatomical wonder how humans communicate with the voice: The contraction of muscles surrounding the diaphragm produces bursts of air particles that are transformed into sound through vibrations of the vocal folds, and leave the mouth, depending on the position of the jaw, the tongue, and other implements of vocal control (Titze & Martin, 1998), in the form of words, laughter, playful intonation, crying, sarcastic tones, sighs, song, triumphant hollers, growls, or motherese. In essential ways, the voice makes humans human.

Recent studies are finding the human voice to be an extraordinarily rich and ubiquitous medium of the communication of emotion (Cordaro, Keltner, Tshering, Wangchuk, & Flynn, 2016; Juslin & Laukka, 2003; Kraus, 2017; Laukka et al., 2016; Provine & Fischer, 1989; Vidrascu & Devillers, 2005). In the present investigation, we apply a new theoretical model to address the following questions: How many emotions can we communicate with the voice? What drives vocal emotion recognition, emotion categories or more general affect dimensions (valence, arousal, etc.)? And is vocal emotion discrete or does it convey gradients of meaning?

Alan S. Cowen

The Richness of the Vocal Communication of Emotion

Humans communicate emotion through two different kinds of vocalization (Keltner, Tracy, Sauter, Cordaro, & McNeil, 2016; Scherer, 1986). One is *prosody*—the non-lexical patterns of tune, rhythm, and timbre in speech. Prosody interacts with words to convey feelings and attitudes, including dispositions felt toward objects and ideas described in speech (Mitchell & Ross, 2013; Scherer & Bänziger, 2004).

In the study of emotional prosody, participants are often recorded communicating different emotions while delivering sentences with neutral content (e.g., "Let me tell you something"). These recordings are then matched by new listeners to emotion words, definitions, or situations. An early review of 60 studies of this kind found that hearers can judge five different emotions in the prosody that accompanies speech—anger, fear, happiness, sadness, and tenderness—with accuracy rates approaching 70% (Juslin & Laukka, 2003; Scherer, Johnstone, & Klasmeyer, 2003). In a more recent study, listeners from four countries identified nine emotions with above-chance levels of recognition accuracy (Laukka et al., 2016).

A second way that humans communicate emotion in the voice is with *vocal bursts*, brief nonlinguistic sounds that occur in between speech incidents or in the absence of speech (Hawk, Van Kleef, Fischer, & Van der Schalk, 2009; Scherer & Wallbott, 1994). Examples include cries, sighs, laughs, shrieks, growls, hollers, roars, oohs, and ahhs (Banse & Scherer, 1996; Cordaro et al., 2016). Vocal bursts are thought to predate language and have precursors in

mammals: For example, primates emit vocalizations specific to predators, food, affiliation, care, sex, and aggression (Snowdon, 2002).

As in the literature on emotional prosody, recent studies have sought to document the range of emotions communicated by vocal bursts (for a summary, see Cordaro et al., 2016). In a paradigm typical in this endeavor, participants are given definitions (e.g., "Awe is the feeling of being in the presence of something vast that you don't immediately understand") and are asked to communicate that emotion with a brief sound that contains no words (Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009). Hearers are then presented with stories of emotion antecedents (e.g., "The person feels awestruck at viewing a large waterfall") and asked to choose from three to four vocal bursts the one that best matches the content of the story. Currently, it appears that 13 emotions can be identified from vocal bursts at rates substantially above chance (Cordaro et al., 2016; Laukka et al., 2013). The capacity to recognize vocal bursts emerges by 2 years of age (Wu, Muentener, & Schulz, 2017) and has been observed in more than 14 cultures, including two remote cultures with minimal Western influence (Cordaro et al., 2016; Sauter, Eisner, Ekman, & Scott, 2010; but see Gendron, Roberson, van der Vyver, & Barrett, 2014).

Vocal bursts are more than just fleeting ways through which we communicate emotion: They structure our social interactions (Keltner & Kring, 1998; Van Kleef, 2010). They convey information about features of the environment, which orients hearers' actions. For example, toddlers are four times less likely to play with a toy when a parent emits a disgust-like vocal burst—*yuck*—than a positive one (Hertenstein & Campos, 2004). They regulate relationships. From laughs, adults can infer a person's rank within a social hierarchy (Oveis, Spectre, Smith, Liu, & Keltner, 2013), and from shared laughs, whether two individuals are friends or strangers (Bryant et al., 2016; Smoski & Bachorowski, 2003). Finally, vocal bursts evoke specific brain responses (Frühholz, Trost, & Kotz, 2016; Scott, Lavan, Chen, & McGettigan, 2014). For example, screams comprise powerfully evocative acoustic signals that selectively activate the amygdala (Arnal, Flinker, Kleinschmidt, Giraud, & Poeppel, 2015). Within 50 ms, an infant's cry or laugh triggers activation in a parent's brain region—the periaqueductal gray—that promotes caring behavior (Parsons et al., 2014).

The Semantic Space of Emotion

What is less well understood is how the recognition of emotion from vocal bursts is structured. How do people infer meaning from brief vocalizations? As emotions unfold, people rely on emotion knowledge—hundreds and even thousands of concepts, metaphors, phrases, and sayings (Russell, 1991)—to describe the emotion-related re-

**Hillary Anger
Elfenbein**

sponse, be it a subjective experience, a physical sensation, or, the focus here, emotion-related expressive behavior. The meaning ascribed to any emotion-related response can be mapped to what we have called a *semantic space* (Cowen & Keltner, 2017, 2018), a multidimensional space that repre-

sents all responses within a modality (e.g., experience, expression). Semantic spaces of emotion are captured in analyses of judgments of the emotion-related response (see Figure 1A).

Semantic spaces are defined by three features. The first is its *dimensionality*—the number of distinct varieties of emotion that people represent within a response modality. In terms of the present investigation, when we perceive various emotional vocalizations, how many kinds of emotion do we recognize within this response modality?

A second property of semantic spaces is what we call *conceptualization*: How do emotion categories (e.g., "sympathy") and domain-general affective appraisals such as valence, arousal, and other themes (e.g., control, unexpectedness) detailed in appraisal and componential theories of emotion (Lazarus, 1991; Scherer, 2009; Smith & Ellsworth, 1985) capture the individual's representation of an emotion-related response (Shuman, Clark-Polner, Meuleman, Sander, & Scherer, 2017)? When a parent hears a child cry, or a friend laugh, or people shouting in a bar in response to a buzzer-beating shot, are emotion categories necessary to capture the meaning the hearer recognizes in the vocalization, or is the meaning captured by broader affective appraisals? Studies have yet to formally address this question, though it is central to contrasting positions within several theories of emo-
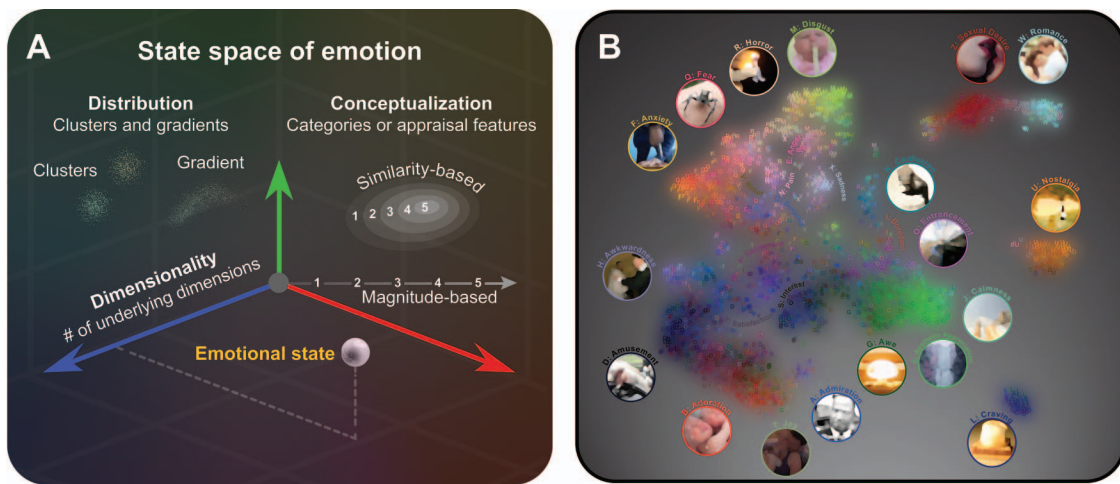

Figure 1. (A) Schematic representation of a semantic space of emotion. A semantic space of emotion is described by (a) its dimensionality, or the number of varieties of emotion within the space; (b) the conceptualization of the emotional states in the space in terms of categories and more general affective features—that is, the concepts which tile the space (upper right); and (c) the distribution of emotional states within the space. A semantic space is the subspace that is captured by concepts (e.g., categories, affective features). For example, a semantic space of expressive signals of the face is a space of facial movements along dimensions that can be conceptualized in terms of emotion categories and affective features, and that involves clusters or gradients of expression along those dimensions. (B) Twenty-seven-dimensional semantic space of reported emotional experience evoked by 2,185 short videos. Each letter represents a video. Videos are positioned and colored according to 27 distinct dimensions required to account for the emotions people reliably reported feeling in response to different videos, which corresponded to emotion categories. Within the space, we can see that there are gradients of emotion between categories traditionally thought of as discrete. Images from illustrative videos are shown for 18 of the dimensions. An interactive version of the figure is available at https://s3-us-west-1.amazonaws.com/emogifs/map.html. See the online article for the color version of this figure.

Petri Laukka

tion (Barrett, 2006; Ekman & Cordaro, 2011; Russell, 2003; Scherer, 2009). Here, we use statistical models to compare how emotion categories and more general affective appraisals capture the meaning people ascribe to vocal bursts.

Finally, semantic spaces of emotion are defined by how states are distributed within the space. Do the emotions people recognize in the human voice (or in facial expression, song, etc.) form clusters? What is the nature of the boundaries within that space? Are the categories really discrete, with sharp boundaries between them? Or not (Barrett, 2006)?

In a first study applying large-scale inference methods to derive a semantic space of emotion, we focused on reports of subjective emotional experience (Cowen & Keltner, 2017). Guided by recent empirical advances and relevant theory, we gathered participants' ratings (forced choice of over 30 emotions, free response, and 13 affective appraisals) of their self-reported experiences in response to 2,185 evocative short videos "scraped" from the Internet. New quantitative techniques allowed us to map a semantic space of subjective emotional experience, as visualized in Figure 1B. That study yielded three findings about the semantic space of emotional experience. First, videos reliably (across participants) elicited at least 27 distinct varieties of reported emotional experience; emotional experience is greater in variety—that is, dimensionality—than previously theorized (Keltner & Lerner, 2010; Shiota et al., 2017). With respect to the conceptualization of emotion, categorical labels such as "amusement," "fear," and "desire" captured, but could not be explained by, reports of affective appraisals (e.g., valence, arousal, agency, certainty, dominance). Thus, as

people represent their experiences with language, emotion categories cannot be reduced to a few broad appraisals such as valence and arousal. Finally, with respect to boundaries between emotion categories, we found little evidence of discreteness, but rather continuous gradients linking one category of experience to another.

Here, we extend our theorizing about semantic spaces to study the recognition of emotion from vocal bursts. This extension to emotion recognition is justified for empirical and theoretical reasons. Most notably, self-reports of subjective experience and emotional expression are only moderately correlated (for reviews, see Fernández-Dols & Crivelli, 2013; Matsumoto, Keltner, Shiota, O'Sullivan, & Frank, 2008). This suggests that semantic spaces of emotional experience and emotion recognition may have different properties, perhaps differing, for example, in their conceptualization or distribution. More generally, examining a semantic space of emotional expression offers new answers to central questions within emotion science: How many emotions have distinct signals? What drives the recognition of emotion, categories or affective appraisals? What is the nature of the boundaries between those categories?

The Present Investigation

The investigation of semantic spaces of emotion recognition requires methodological departures from how emotion recognition has been studied in the past. To capture the dimensionality of emotion recognition, it is critical to study as wide an array of expressive signals as possible. One way to achieve this is by sampling structural variations in expression, for instance, by attempting to reconstruct variations in facial expression using an artificial 3D model (Jack, Sun, Delis, Garrod, & Schyns, 2016). Another approach is to sample naturalistic expressive behaviors of as many emotion concepts as possible (Cowen & Keltner, 2017). Early studies of emotional vocalization focused primarily on emotions traditionally studied in facial expression—anger, sadness, fear, disgust, and surprise (Juslin & Laukka, 2003). More recent studies have expanded in focus to upward of 16 emotion concepts (Cordaro et al., 2016; Laukka et al., 2016). Here, we build upon these discoveries to consider how a wider array of emotion concepts—30 categories and 13 appraisals—may be distinguished in vocal bursts.

To capture the distribution of recognized emotions, it is critical to account for nuanced variation in expression, anticipated in early claims about expressive behavior (Ekman, 1993) but rarely studied. Most studies have focused on one to two prototypical vocal bursts for each emotion category (Cordaro et al., 2016; Gendron et al., 2014; Simon-Thomas et al., 2009). This focus on prototypical expressions fails to capture how vocal bursts vary within an emotion category

Dacher Keltner

and can yield erroneous claims about the nature of the boundaries between categories (Barrett, 2006).

With respect to conceptualization, it is critical to have independent samples of hearers rate the stimuli (vocal bursts) in terms of emotion categories *and* broader affective appraisals such as valence, arousal, certainty, and dominance. A more typical approach has been to match vocal bursts to discrete emotions, using words or brief stories depicting antecedents (Cordaro et al., 2016; Simon-Thomas et al., 2009). More recently, investigators have been gathering ratings of not only emotion categories but also inferred appraisals and intentions (Nordström, Laukka, Thingujam, Schubert, & Elfenbein, 2017; Shuman et al., 2017). By combining such data with statistical approaches developed in the study of emotional experience (Cowen & Keltner, 2017), we can compare how emotion categories and affect appraisals shape emotion recognition.

In sum, to characterize the semantic space of emotion recognition of vocal bursts, what is required is the study of a wide array of emotions, many examples of each emotion, and observer reports of both emotion categories and affective appraisals. Guided by these considerations, we present two studies that examine the semantic space of emotion recognition of vocal bursts. In the first, participants rated the largest array of vocal bursts studied to date—over 2,000—for the emotion categories they conveyed, for their affective appraisals, culled from dimensional and componential theories of emotion (Lazarus, 1991; Scherer, 2009; Smith & Ellsworth, 1985), and in a free-response format. Building upon the discoveries of Study 1, and given concerns about the nature of vocal bursts produced in laboratory settings, in Study 2, we gathered ratings of vocal bursts culled from

YouTube videos of people in naturalistic contexts. Guided by findings of candidate vocal expressions for a wide range of emotion categories, along with taxonomic principles derived in the study of emotional experience (Cowen & Keltner, 2017), we tested the following hypotheses: (a) With respect to the dimensionality of emotion recognition in vocal bursts, people will reliably distinguish upward of 20 distinct dimensions, or kinds, of emotion; (b) With respect to conceptualization, the recognition of vocal bursts in terms of emotion categories will explain the perception of affective appraisals such as valence and arousal, but not vice versa; and (c) With respect to the distribution of emotion, emotion categories (e.g., "awe" and "surprise") will be joined by gradients corresponding to smooth variations in how they are perceived (e.g., reliable judgments of intermediate valence along the gradient).

Experiment 1: The Dimensionality, Conceptualization, and Distribution of Emotion Recognition of Voluntarily Produced Vocal Bursts

Characterizing emotion-related semantic spaces requires (a) the study of as wide an array of emotions as possible; (b) the study of a rich array of expressions, to expand beyond the study of expression prototypes; and (c) judgments of emotion categories and affective appraisals. Toward these ends, we captured 2,032 vocal bursts in the laboratory. One set of participants then judged these vocal bursts in terms of emotion categories. A second set of participants judged the vocal bursts using a free-response format. A final set of participants judged the vocal bursts for 13 scales of affect, guided by the latest advances in appraisal and componential theories, which reflect the most systematic efforts to characterize the full dimensionality of emotion using domain-general concepts. With new statistical techniques, we tested hypotheses concerning the dimensionality, conceptualization, and distribution of emotions as recognized in the richest array of vocal bursts studied to date.

Method

Creation of library of vocal bursts. Guided by past methods (Simon-Thomas et al., 2009), we recorded 2,032 vocal bursts by asking 56 individuals (26 females, Ages 18–35) to express their emotions as they imagined being in scenarios that were roughly balanced in terms of 30 emotion categories and richly varying along 13 commonly measured affective appraisals (Tables S1 and S2 of the online supplemental materials). The individuals were recruited from four countries (27 United States, 9 India, 13 Kenya, 7 Singapore) and including professional actors and amateur volunteers (see online supplemental materials for details). The emotion categories were derived from recent studies demonstrating

that a number of categories of emotion are reliably conveyed by vocal bursts (Cordaro et al., 2016; Laukka et al., 2013; Simon-Thomas et al., 2009; for summary, see Table S3 of the online supplemental materials and Keltner et al., 2016) and from findings of states found to reliably occur in daily interactions (Rozin & Cohen, 2003). the categories adoration, amusement, anger, awe, confusion, contempt, contentment, desire, disappointment, disgust, distress, ecstasy, elation, embarrassment, fear, guilt, interest, love, neutral, pain, pride, realization, relief, sadness, serenity, shame, surprise (negative), surprise (positive), sympathy, and triumph. The affective scales were culled from dimensional and componential theories of emotion (for summary, see Table S4 of the online supplemental materials). The affective scales valence, arousal, approach/avoidance, attention, certainty versus uncertainty, commitment-to-an-individual, control, dominance versus submissiveness, fairness versus unfairness, identity-with-a-group, improvement/worsening, obstruction, and safety versus unsafety (Mehrabian & Russell, 1974; Russell, 2003; Scherer, 2009; Smith & Ellsworth, 1985). Four hundred twenty-five of the stimuli were recorded as part of the Vocal Expressions of Nineteen Emotions across Cultures (VENEC) corpus (Laukka et al., 2013); 1,607 are introduced here. In light of potential limitations both in the expressive abilities of the individuals who produced the vocal bursts and the range of emotional scenarios used to elicit the vocal bursts, our results should not be taken as evidence against the possibility of other emotions that might be expressed with vocal bursts—that is, we do not claim that the expressions studied here are fully exhaustive of human vocal expression. Our intention was to create as complex an array of emotional vocal bursts as warranted by the current scientific literature.

Categorical, affective appraisal, and free-response judgments of vocal bursts. In our study of emotional experience, collecting judgments from approximately 10 participants has proven sufficient to closely approximate the population mean in terms of reliability (Cowen & Keltner, 2017). Presuming that this would extend to vocal bursts, we used Amazon Mechanical Turk to obtain repeated (9–12) judgments of each vocal burst in three randomly assigned formats. A total of 1,017 U.S. English-speaking raters Ages 18 to 76 (545 female, mean age = 36 years) participated. Figure S1 of the online supplemental materials provides a breakdown of participant age and gender by response format. Participants were randomly assigned to one of three emotion recognition conditions. In the first, participants judged each sound in terms of the 30 aforementioned emotion categories, choosing the category that best matched the sound that they heard from a list of 30 presented alphabetically, alongside relevant synonyms. A second group of participants rated each sound in terms of the 13 affective scales, covering not only valence and arousal but also other affective appraisals required to differentiate more complex arrays of emotions (Lazarus, 1991; Scherer, 2009; Smith &

Ellsworth, 1985). Given the limitations of forced choice formats (DiGirolamo & Russell, 2017; Russell, 1994), a third group of participants provided free-response descriptions of the emotion conveyed by each sound. For this, they were instructed to type in whatever interpretation of the vocal burst they felt appropriate. After typing each letter, a set of all possible completions from a comprehensive list of 600 emotion terms appeared (Data Set S1 of the online supplemental materials). This search-based format is not entirely unconstrained, but it addresses the critique that multiple choice formats prime participants with a list of concepts and has the advantage of ruling out ambiguities from multiple spellings and conjugations of each term. Examples of each judgment format are given in Figure 2A–2C. The experimental procedures were approved by the Institutional Review Board at the University of California, Berkeley. All participants gave their informed consent.

On average, a participant judged 279.8 vocal bursts. With these methods, a total of 284,512 individual judgments (24,384 forced-choice categorical judgments, 24,384 free-response judgments, and 237,744 9-point affective scale judgments) were gathered to enable us to characterize the semantic space of vocal emotion recognition. These methods were found to capture an estimated 90.8% and 91.0% of the variance in the population mean categorical and affective scale judgments, respectively, indicating that they accurately characterized the population average responses to each individual stimulus, and demonstrating that they did so with roughly equivalent precision for each judgment type. (See online supplemental materials for details regarding explainable variance calculation. Our methods captured 58.2% of the variance in population mean free-response judgments, as a result of the multiplicity of distinct free responses, but none of our analyses rely on the estimation of mean free-response judgments of individual stimuli.)

Results

How many emotions can vocal bursts communicate? Our first hypothesis predicted that people recognize upward of 20 distinct emotions in the 2,032 vocal bursts. We tested this hypothesis in two ways: by assessing the degree to which participants agreed in their labeling of the vocal bursts, and by analyzing commonalities in forced choice and free-response labeling of the vocal bursts to uncover the number of emotions recognized with both methods.

Forced-choice labeling. We first analyzed how many emotion categories were recognized at above-chance levels in judgments of the 2,032 vocal bursts. Most typically, emotion recognition has been assessed in the degree to which participants' judgments conform to experimenters' expectations regarding the emotions conveyed by each expression (see Laukka et al., 2013, for such results from an overlapping set of vocal bursts). We depart from this con-
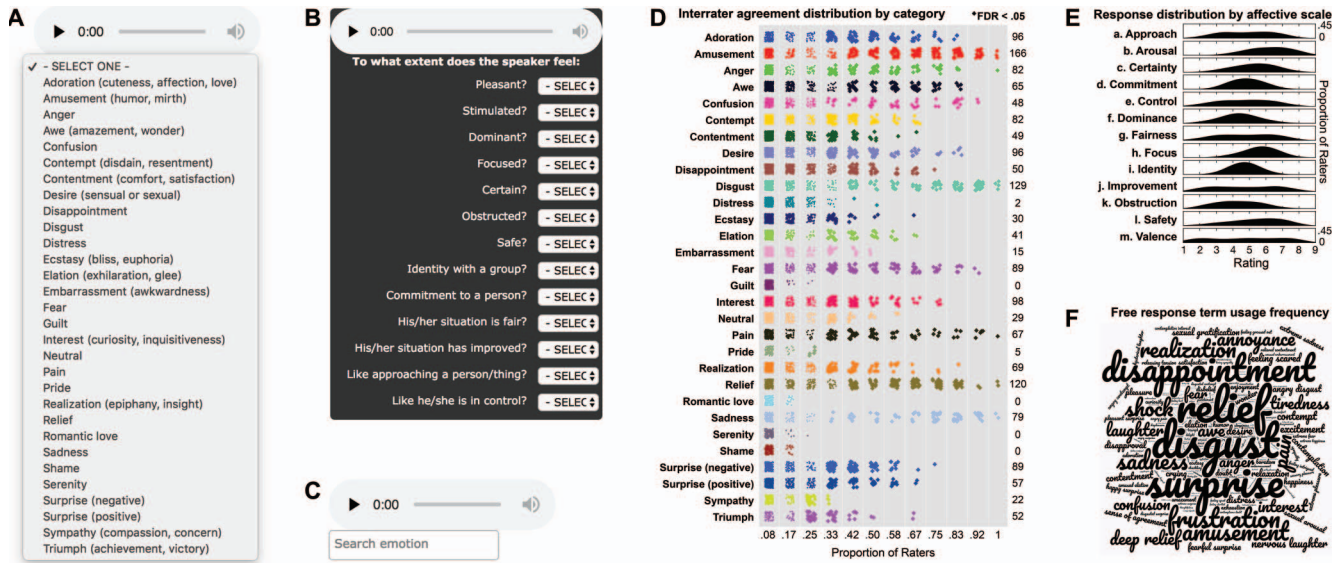
Figure 2. Judgment formats and response distributions. (A–C) Example survey segments are shown from the (A) categorical, (B) affective scale, and (C) free-response surveys. (D) Interrater agreement levels per vocal burst, per category for all 30 categories. Nonzero interrater agreement rates are shown for each category across all 2,032 vocal bursts. The number of vocal bursts in each category with significant interrater agreement rates (FDR < .05, simulation test detailed in online supplemental materials) is shown to the right of each dot plot and varies from 0 (*guilt, romantic love, serenity, and shame*) to 166 (*amusement*), $M = 57.6$, $SD = 43.0$. Dots have been jittered for clarity. (E) Response distribution by affective scale. Smooth histograms indicate the distribution of average judgments of the vocal bursts for each affective scale. (F) Free-response term usage frequency. The height of each term is directly proportion to its usage frequency. The most commonly used term, *disgust* was applied 1,074 times. See the online article for the color version of this figure.

firmatory approach for two reasons: It assumes that each scenario that prompted the production of a vocal burst can elicit only one emotion (rather than blends of emotions and/or emotions different from experimenters' predictions), and the vocal bursts recorded during each session were not always attributable to a single scenario (see online supplemental materials for details). To capture emotion recognition, we instead focus on the reliable perception (recognition) of emotion in each vocal burst, operationalized as interrater agreement—the number of raters who chose the most frequently chosen emotion category for each vocal burst. This approach is motivated by arguments that the signal value of an expression is determined by how it is normally perceived (Jack et al., 2016).

In this preliminary test of our first hypothesis, we found that 26 of the 30 emotion categories were recognized at above-chance rates from vocal bursts (mean 57.6 significant vocal bursts per category; see Figure 2D). Overall, 77% of the 2,032 vocal bursts were reliably identified with at least one category (false discovery rate [FDR] = $q <$ .05, simulation test, detailed in online supplemental materials). For interrater agreement levels per vocal burst, as well as the number of significantly recognized vocal bursts in each category, see Figure 2D. In terms more typically used in the field, the average rate of interrater agreement—that is, the

average number of raters who chose the maximally chosen category for each vocal burst—was 47.7%, comparable with recognition levels observed in past vocal burst studies (Cordaro et al., 2016; Elfenbein & Ambady, 2002; Laukka et al., 2013; Sauter et al., 2010; Simon-Thomas et al., 2009). (Note that interrater agreement rates did not differ substantially by culture in which vocal bursts were recorded; see Figure 2A of the online supplemental materials).

Relating categories to free-response judgments. Our first finding, then, is that as many as 26 emotion categories are recognized at above-chance levels in vocal bursts in a forced choice format. Such univariate accuracy metrics, however, *do not* determine the number of categories that are truly distinct in the recognition of emotion, despite their centrality to past studies of emotion recognition. For example, any two categories, such as *sadness* and *disappointment*, may both have been reported reliably even if they were used interchangeably (e.g., as synonyms). Given this issue, and potential limitations of forced choice judgments (DiGirolamo & Russell, 2017; Russell, 1994), we tested our first hypothesis in a second way, by determining how many distinct dimensions, or patterns of responses, within the category judgments were also reliably associated with distinct patterns of responses within the free-response judgments.

To assess the number of dimensions of emotion, or kinds of emotion, that participants reliably *distinguished* from the 2,032 vocal bursts across the two methods, we used canonical correlations analysis (CCA; Hardoon, Szedmak, & Shawe-Taylor, 2004; see online supplemental materials for discussion of limitations of factor analytic approaches used by past studies). This analysis uncovered dimensions in the category judgments that correlated significantly with dimensions in the free-response judgments. As one can see in Figure 3A, we found that 24 distinct dimensions, or kinds of emotion

($p = .014$), were required to explain the reliable correlations between the category and free-response judgments of the 2,032 vocal bursts. What this means is that there were 24 distinct kinds of emotion evident in the patterns of covariation between forced choice and free-response judgments, a finding that goes beyond the interrater agreement findings in showing that people not only reliably identify but also reliably distinguish a wide variety of emotions from the 2,032 vocal bursts, and that is in keeping with our first hypothesis. Vocal bursts can communicate at least 24 distinct emotions, nearly dou-
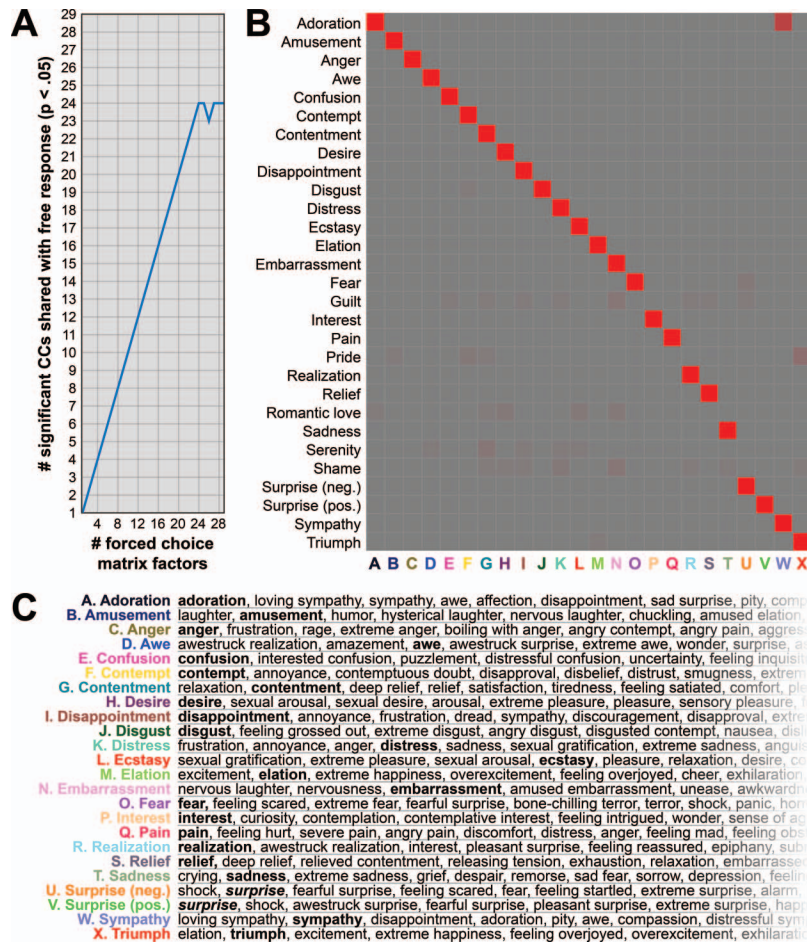

Figure 3. Vocal bursts reliably and accurately convey 24 distinct dimensions of emotion. (A) The application of canonical correlations analysis (CCA) between the categorical and free-response judgments revealed 24 shared dimensions ($p < .05$). Here, the results of a subsequent analysis reveal that this 24-dimensional space can be derived by applying non-negative matrix factorization (NNMF) to the categorical judgments. NNMF was applied iteratively with increasing numbers of factors, from 1 to 29. The first 24 all span separate dimensions of meaning shared with the free-response judgments when subsequently applying CCA ($p < .05$). (B) Shown here are the 24 dimensions of emotion conveyed by vocal bursts, extracted by applying NNMF to the category judgments. We can see that they each load maximally on a distinct category. (C) The free-response terms most closely correlated with each of the 24 dimensions consistently include this maximally loading category and related terms (for corresponding correlation values, see Table S5 of the online supplemental materials). Together, these results reveal that vocal bursts can reliably (in terms of consistency across raters) and accurately (in terms of recognized meaning) convey at least 24 semantically distinct semantic dimensions of emotion. See the online article for the color version of this figure.

bling the number of emotions recognized from vocal bursts documented in past studies (e.g., Cordaro et al., 2016).

In our next analysis, we determined the *meaning* of the 24 dimensions of emotion conveyed by the vocal bursts. We did so with a technique that reduces the patterns of category judgments to their underlying dimensions, or what one might think of as core meaning. Although different techniques extract rotations of the same dimensional space, we chose non-negative matrix factorization (NNMF; Lee & Seung, 1999) because it produces dimensions that are easily interpretable, having only positive loadings on each category. We applied NNMF with an increasing number of dimensions, ranging from 1 to the maximum, 29 (the number of emotion categories after excluding "neutral," which is redundant given the forced-choice nature of the judgments). We then subsequently performed CCA between the resulting dimensions of the emotion category judgments and the free-response judgments. The results of this analysis (Figure 3A) reveal that NNMF extracted 24 dimensions that were reliably shared with the free-response judgments ($p = .011$). In other words, when NNMF extracted 24 dimensions from participants' emotion category judgments of the vocal bursts, it captured the 24 distinct kinds of emotion required to explain correlations between category and free-response judgments.

To interpret the 24 kinds of emotion conveyed by the vocal bursts, we assessed the weights of each dimension extracted by NNMF on the 30 categories of emotion with which we began. This analysis determines how each emotion category judgment contributes to each of the 24 dimensions we have uncovered in patterns of covariation across the forced-choice and free-response methods. Which categories do each of the 24 dimensions represent? The answer to this question is represented in Figure 3B. In this chart, the colored letters on the horizontal axis refer to the 24 dimensions, and the category names are presented on the vertical axis. The dimensions closely corresponded to adoration, amusement, anger, awe, confusion, contempt, contentment, desire, disappointment, disgust, distress, ecstasy, elation, embarrassment, fear, interest, pain, realization, relief, sadness, surprise (negative), surprise (positive), sympathy, and triumph. (The categories that did not correspond to distinct dimensions—guilt, pride, romantic love, serenity, and shame—were among the least frequently chosen, as shown in Figure 2B of the online supplemental materials. Hence, the categories people chose more regularly tended to be used reliably.)

To verify that the 24 dimensions of emotion conveyed by vocal bursts can be accurately interpreted in terms of the 24 categories, we examined whether participants used similar emotion terms in the free-response format to describe each vocal burst. The results of this analysis are presented in Figure 3C, which shows how the dimensions, labeled in

colored terms to the left of each row, correspond to the free-response terms, in black, chosen to label the 2,032 vocal bursts, listed in order of descending correlation. These results indicate that each of the 24 dimensions, or kinds of emotion, was closely correlated with the use of the free-response term identical to its maximally loading category found in the forced choice judgments, along with synonyms and related terms. To illustrate, awe emerged as a kind of emotion conveyed with vocal bursts, and in free-response format, people were most likely to label vocal bursts conveying awe with "awestruck, amazement, and awe." Thus, the dimensions accurately capture the meaning attributed to each vocal burst as opposed to reflecting methodological artifacts of forced choice, such as process of elimination (DiGirolamo & Russell, 2017).

Taken together, the results thus far lend support to our first hypothesis: 24 dimensions of emotion were reliably identified in categorical and free-response judgments of 2,032 vocal bursts.

The conceptualization of emotion: Are emotion categories, affective appraisals, or both necessary to capture the recognition of emotion? What captures the recognition of emotion from vocal bursts: emotion categories, more general affective appraisals (such as valence, arousal, certainty, effort, and dominance), or both? Our second hypothesis derived from our past study of emotional experience (Cowen & Keltner, 2017), and held that participants' categorical judgments of vocal bursts would explain their judgments along the 13 affective scales, but not vice versa. Emotion categories, this hypothesis holds, offer a broader conceptualization of emotion recognition than the most widely studied affective appraisals.

This prediction can be formally tested using statistical models. With the present data, we used cross-validated predictive models, which determined the extent to which participants' affective scale judgments (of valence, arousal and so on) explained participants' categorical judgments, and vice versa. In keeping with our second hypothesis, we found that using both linear (ordinary least squares) and nonlinear (nearest neighbors) models, the categories explain almost all of the explainable variance in the scales of affect (90% and 97%), whereas the scales of affect explain at most around half of the explainable variance in the categories (32% and 52%). (See online supplemental materials for more details of this analysis.) In Figure 4A, we illustrate how the information carried in participants' ratings using the 13 affective scales—valence, arousal, dominance, and so forth, as represented in green—is a subspace of the information carried by the emotion categories, as represented in orange. What this means is that how people conceptualize emotion in vocal bursts in terms of valence, arousal, certainty, and so on can be explained by how they categorize the emotion with emotion categories, but there is variance in the emotions conveyed by vocal bursts that is not explained by valence, arousal, dominance, and other affective
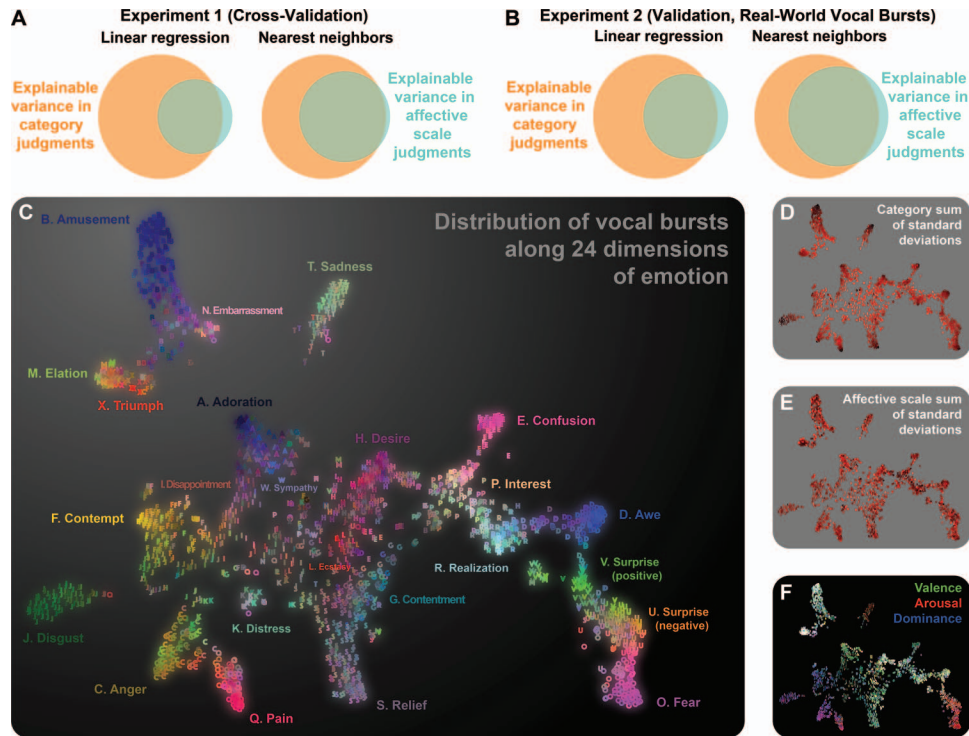
Figure 4. (A) Affective scale judgments of the vocal bursts are explained by, but cannot explain, the category judgments. Both linear and nonlinear models reveal that the explainable variance in the 13 affective scale judgments is almost entirely a subspace of the explainable variance in the forced-choice categorical judgments (see online supplemental materials for details.) Thus, the categories have more explanatory value than the scales of affect in explicating how emotions are reliably recognized from vocal bursts. (B). These results extend to real-world expressions. Models were trained on data from Experiment 1 and tested on ratings of the naturalistic vocal bursts gathered in Experiment 2. Both linear and nonlinear models reveal that the explainable variance in the 13 affective scale judgments is almost entirely a subspace of the explainable variance in the forced-choice categorical judgment. The categories have more explanatory value than the scales of affect in explicating how emotions are reliably recognized from vocal bursts. (C) Map of 2,032 vocal bursts within a 24-dimensional space of vocal emotion generated with t-distributed stochastic neighbor embedding (t-SNE). T-SNE approximates local distances between vocal bursts without assuming linearity (factor analysis) or discreteness (clustering). Each letter corresponds to a vocal burst and reflects its maximal loading on 24 dimensions extracted from the category judgments. Unique colors assigned to each category are interpolated to smoothly reflect the loadings of each vocal burst on each dimension. The map reveals smooth structure in the data, with variations in the proximity of vocal bursts within each category to other categories. See the interactive map (available at https://s3-us-west-1.amazonaws.com/vocs/map.html) to explore every sound and view its ratings. (D–E) Standard deviations in judgments of the categories and affective scales. A black-to-red (or dark to light) scale represents the sum, across categories or affective scales, of the standard deviations in judgments across participants. For example, vocal bursts plotted in black had lower standard deviations in categorical judgments (i.e., high interrater agreements) and thus were further from the perceptual boundaries between categories. There is little relationship between affective scale ambiguity and proximity to category boundaries, suggesting that the categories are confused primarily because they are bridged by smooth gradients in meaning rather than as a result of ambiguity. See Figure S3 of the online supplemental materials for further analyses confirming this for specific gradients. (F) Cross-category gradients correlate with variations in valence (green, highest in center), arousal (red, highest on bottom right), and dominance (blue, highest on bottom left). Despite the weak relationship between affective scale ambiguity and proximity to category boundaries, categories predict smooth gradients in affective meaning, further establishing that these gradients are not a byproduct of ambiguity. We can also see why the affective scales are insufficient to explain the category judgments, given that almost every unique color—that is, combination of valence, arousal, and dominance—is associated with multiple regions in the semantic space of emotion recognized in vocal bursts. See the online article for the color version of this figure.

appraisals. This finding supports our second hypothesis—that emotion categories have greater explanatory value than the affective scales in representing how people recognize emotion in vocal bursts.

Are emotion categories discrete or not? The distribution of emotion categories. To test our third hypothesis—that we would find smooth gradients between the different categories of emotion recognized in vocal bursts—we first visualized the distribution of the category judgments of vocal bursts within the 24-dimensional space we have uncovered. We did so using a method called t-distributed stochastic neighbor embedding (t-SNE; Maaten & Hinton, 2008). This method projects high-dimensional data into a two-dimensional space in a way that attempts to preserve the local distances between data points, in this case, separate vocal bursts, as much as possible, but allows for distortions in longer distances between data points, with the goal of illustrating the shape of the distribution of the data. We applied t-SNE to the projections of the vocal burst categorical judgments into the 24-dimensional space extracted using NNMF (see Figure 4). As expected, t-SNE distorted larger distances between vocal bursts in the 24-dimensional space but preserved more fine-grained similarities between vocal bursts, such as continuous gradients of perceived emotional states, illustrated by smooth variations in color in Figure 4. In an online interactive version of Figure 4 (https:// s3-us-west-1.amazonaws.com/vocs/map.html), each of the 2,032 sounds can be played in view of their relative positions within the 24-dimensional space of vocal emotion. This visualization simultaneously represents the distribution of the 24 distinct emotion categories uncovered in prior analyses within a semantic space, the variants of vocal bursts within each emotion category, and the gradients that bridge categories of emotion.

Visual and auditory inspection of this representation of the semantic space of emotion recognition in vocal bursts yields several insights that converge with claims in the science of emotion: that the emotions of moral disapproval—anger, disgust, and contempt (see bottom left of figure) are in a similar general space but are distinct (Rozin, Lowery, Imada, & Haidt, 1999); that one set of emotions—interest, awe, realization, and confusion (see middle to the right of the figure)—track the individual's understanding of the world (Shiota et al., 2017); that expressions of love (adoration) and desire are distinct, as anticipated by attachment theory (Gonzaga, Keltner, Londahl, & Smith, 2001); and that embarrassment and amusement (see upper left of figure) are close in meaning (Keltner, 1996), as are surprise and fear (Ekman, 1993).

As evident in these figures, and in keeping with our third hypothesis, the semantic space of emotion recognition from the human voice is a complex, nondiscrete structure with smooth within-category variations in the relative location of each individual vocal burst. For example, some vocal bursts

most frequently labeled "negative surprise" were placed further away from vocal bursts labeled fear, whereas others were placed adjacent to fear. Listening to these different surprise vocal bursts gives the impression that the ones placed closer to fear signal a greater potential for danger than those placed further away, suggesting that surprise and fear may occupy a continuous gradient despite typically being considered discrete categories. Similar gradients were revealed between a variety of other category pairs, bridging anger with pain, elation with triumph, relief with contentment, desire with ecstasy, and more (Figure 4, left).

One explanation for these results is that many of the vocalizations were perceived one way by some participants (e.g., as anger) and differently by others (e.g., as pain). An alternative possibility is that although some vocal bursts clearly signal one emotion, others reliably convey intermediate or mixed emotional states. To address these contrasting possibilities, we compared proximity to category boundaries (operationalized as the standard deviation in category ratings) with the standard deviation in ratings of the affective scales. If the smooth gradients between categories were caused by ambiguity in meaning, then vocal bursts that fell near the boundary between two categories that consistently differed in their affective appraisals would also have received highly variable affective scale judgments across participants.

We find that gradients between categories correspond to smooth variations in meaning. In many cases, neighboring categories (e.g., adoration, sympathy) differed dramatically in their affective appraisals, but intermediacy between categories did not relate to *ambiguity* in affective appraisals such as valence, as seen in Figure 4 (right). Overall, the correlation between category and affective scale ambiguity—the sum of the standard deviations of each attribute—was quite weak (Pearson's $r = .14$, Spearman's $r = .13$). These results suggest that the smooth gradients between categories largely reflect what might be thought of as emotional blends of categories as opposed to ambiguity in meaning. These findings align with the recent interest in blended emotions captured in subjective experience (Cowen & Keltner, 2017; Watson & Stanton, 2017) and facial expression (Du, Tao, & Martinez, 2014). For further analyses confirming that smooth gradients between specific categories correspond to smooth variations in meaning, see Figure S3.

Experiment 2: The Conceptualization of Emotion in Real-World Vocal Bursts

Method

A shortcoming in our first study, common to the study of emotional vocalization, is that the vocal bursts were produced in laboratory settings. This raises the question of

whether our findings extend to vocalizations found in the real world (Anikin & Lima, 2017; Juslin, Laukka, & Bän-ziger, 2018; Sauter & Fischer, 2018; Scherer, 2013). With respect to the *conceptualization* of emotion recognition in vocal bursts, concerns regarding ecological validity may limit our finding that emotion categories explain the signaling of affective scales such as valence and arousal, given that signals of valence, arousal, and so on may vary more in ecological contexts (e.g., Russell, 1994). To address this possibility, we obtained judgments of a set of 48 vocal bursts recorded in more naturalistic settings and used these data to further examine whether categories judgments reliably capture scales of affect in real-life vocal bursts.

Guided by the 24 kinds of emotion recognized in vocal bursts documented in our first study, we extracted a library of 48 vocal bursts from richly varying emotional situations captured in naturalistic YouTube videos. These contexts were intended to span the 24 dimensions of emotion observed in Study 1 and to richly vary along the 13 affective scales (see Table S6 of the online supplemental materials for description of the contexts in which the sounds were extracted, predicted categorical and affective appraisals, as well as links to each sound). The vocal bursts were found in emotionally evocative situations—watching a baby's pratfalls, viewing a magic trick, disclosures of infidelity or the loss of virginity, observing physical pain, eating good food, or receiving a massage. Videos were located using search terms associated with of emotional content (e.g., *puppy*, *hugged*, *falling*, *magic trick*) and vocal bursts were extracted from emotional situations within each video. We obtained repeated (18) judgments of each vocal burst in terms of the 30 categories and the 13 affective scales, as in Experiment 1, using Amazon Mechanical Turk. A total of 88 additional U.S. English-speaking raters (46 female, mean age = 36) participated in these surveys.

Results

Interrater agreement for the naturalistic vocal bursts averaged 42.6%, comparable with those observed for laboratory-induced vocal bursts in Experiment 1 (Figures S2A and S4 of the online supplemental materials), despite potential limitations in the quality of recordings found on YouTube. Using predictive models estimated on data from Experiment 1 and tested on data from Experiment 2, we determined the extent to which categorical judgments explained affective appraisal judgments and vice versa in the naturalistic vocal bursts. Replicating results from Experiment 1, and in keeping with Hypothesis 2, we found that the categories explained almost all of the explainable variance in the affective scales (89% and 92%), whereas the affective scales explained around half of the explainable variance in the categories (41% and 59%), as shown in Figure 4B. These results extend our findings from vocal bursts collected in the laboratory to samples from the real world, confirming that emotion categories capture, but cannot be captured by, affective scales representing core affect themes such as valence, arousal, dominance, and so on.

General Discussion

The human voice is a complex medium of emotional communication that richly structures social interactions. Recent studies have made progress in understanding which emotions are recognized in the voice, how this capacity emerges in development and is preserved across cultures, and parallels between emotional communication in the voice and feelings induced by music. Given past methodological tendencies (e.g., the focus on prototypical vocalizations of a limited number of emotions), what is only beginning to be understood is how the emotions conveyed by the voice are represented within a multidimensional space.

Toward this end, we have proposed a theoretical approach to capturing how people represent emotion-related experience and expression within a semantic space (Cowen & Keltner, 2017, 2018). Semantic spaces are defined by their dimensionality, or number of emotions, their conceptualization of emotional states (e.g., specific emotion categories or domain-general affective appraisals), and the distribution of emotion-related responses, in all their variation, within that space. Building upon a study of emotional experience (Cowen & Keltner, 2017), in the present study, we examined the semantic space of emotion recognition of vocal bursts. We gathered vocal bursts from 56 individuals imagining over 100 scenarios richly varying in terms of 30 categories of emotion and 13 affective scales, yielding the widest array of emotion-related vocalizations studied to date. Participants judged these 2,032 vocal bursts with items derived from categorical and appraisal/constructionist approaches as well as in free-response format. These methods captured a semantic space of emotion recognition in the human voice, revealing the kinds of emotion signaled and their organization within a multidimensional space.

Taken together, the results of our two studies yield data-driven insights into important questions within emotion science. Vocal bursts are richer and more nuanced than typically thought, reliably conveying 24 dimensions of emotion that can be conceptualized in terms of emotion categories. In contrast to many constructivist and appraisal theories, these dimensions cannot be explained in terms of a set of domain-general appraisals, most notably valence and arousal, that are posited to underlie the recognition of emotion and commonly used in the measurement of emotion-related response. However, in contrast to discrete emotion theories, the emotions conveyed by vocal bursts vary continuously in meaning along gradients between categories. These results converge with doubts that emotion

categories "carve nature at its joints" (Barrett, 2006). Visualizing the distribution of 2,032 vocal bursts along 24 continuous semantic dimensions (see https://s3-us-west-1.amazonaws.com/vocs/map.html) demonstrates the variety and nuance of emotions they signal.

Vocal bursts signal myriad positive emotions that have recently garnered scientific attention, such as adoration, amusement, awe, contentment, desire, ecstasy, elation, interest, and triumph (Shiota et al., 2017; Tracy & Matsumoto, 2008). Evidence for expressions distinguishing these states validates the recent expansion in the range of emotions investigated empirically. So, too, do findings of distinctions between nuanced states relevant to more specific theoretical claims: adoration/sympathy, anger/disappointment, distress/fear, and negative/positive surprise (Egner, 2011; Johnson & Connelly, 2014; Reiss, 1991; Shiota et al., 2017).

That many categories were bridged by smooth gradients converges with recent studies of emotional blends in subjective experience and facial expression (Cowen & Keltner, 2017; Du et al., 2014; Watson & Stanton, 2017). Across gradients between categories, vocal signals vary continuously in meaning. For example, as they traverse the gradient from "disappointment" to "sympathy" to "love," vocal bursts increasingly signal the desire to approach (Figure S3 of the online supplemental materials). These findings point to a rich landscape of emotional blends in vocalization warranting further study.

The present findings also inform progress in related fields, including the neural basis of emotion recognition and the training of machines to decode emotion. Future neuroscience studies will need to expand in focus beyond discrete prototypes of a few emotion categories to explain how the rich variety of categories are distinguished by the brain and how continuous variation between categories may be represented in continuously varying patterns of brain activity (for indications of this, see Harris, Young, & Andrews, 2012). Similarly, machine learning efforts to decode emotion from audio will need to expand in focus beyond a small set of discrete categories or two affective appraisals (Schüller, Elkins, & Scherer, 2017) to account for how a wide array of categories and continuous gradients between them are conveyed by acoustic features (for relationship with duration, fundamental frequency, and harmonic-to-noise ratio, see Figure S5 of the online supplemental materials).

In considering these findings, it is worth noting that they are based on vocal emotion recognition by U.S. participants. Further work is needed to examine variation in the semantic space of recognition in other cultures, though preliminary studies have suggested that upward of 20 emotion categories may be recognized even in remote cultures (Cordaro et al., 2016), supporting possible universals in the semantic space of vocal emotion. Likewise, given reports of developmental change and age-related decline in vocal emotion

recognition (Lima, Alves, Scott, & Castro, 2014; Sauter, Panattoni, & Happé, 2013), it would also be interesting to examine variation in the semantic space of vocal emotion recognition across the life span. Finally, given the acoustic similarities between emotion recognition in vocalization and in music (Juslin & Laukka, 2003; Laukka & Juslin, 2007), it would be interesting to compare these semantic spaces.

The present findings dovetail with recent inquiries into the semantic space of emotional experience (Cowen & Keltner, 2017). There is substantial overlap between the varieties of emotion recognition uncovered in the present study and those identified in reported emotional experience, including adoration, amusement, anger, awe, confusion, contentment/peacefulness, desire, disgust, distress/anxiety, elation/joy, embarrassment/awkwardness, fear, interest, pain, relief, sadness, and surprise. In both studies, categories captured a broad space of emotion recognition, and continuous gradients bridged categories such as interest and awe. Together, these results converge on a high-dimensional taxonomy of emotion defined by a rich array of categories bridged by smooth gradients.

References

Anikin, A., & Lima, C. F. (2017). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *The Quarterly Journal of Experimental, 71,* 1–21. http://dx.doi.org/10.1080/17470218.2016.1270976

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology, 25,* 2051–2056. http://dx.doi.org/10.1016/j.cub.2015.06.043

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614–636. http://dx.doi.org/10.1037/0022-3514.70.3.614

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science, 1,* 28–58. http://dx.doi.org/10.1111/j.1745-6916.2006.00003.x

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

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

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

Cowen, A. S., & Keltner, D. (2018). Clarifying the conceptualization, dimensionality, and structure of emotion: Response to Barrett and colleagues. *Trends in Cognitive Sciences, 22,* 274–276. http://dx.doi.org/10.1016/j.tics.2018.02.003

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

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America, 111,* E1454–E1462. http://dx.doi.org/10.1073/pnas.1322355111

Egner, T. (2011). Surprise! A unifying model of dorsal anterior cingulate function? *Nature Neuroscience, 14,* 1219–1220. http://dx.doi.org/10.1038/nn.2932

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48,* 384–392. http://dx.doi.org/10.1037/0003-066X.48.4.384

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

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

Fernández-Dols, J. M., & Crivelli, C. (2013). Emotion and expression: Naturalistic studies. *Emotion Review, 5,* 24–29. http://dx.doi.org/10.1177/1754073912457229

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews, 68,* 96–110. http://dx.doi.org/10.1016/j.neubiorev.2016.05.002

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). 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

Gonzaga, G. C., Keltner, D., Londahl, E. A., & Smith, M. D. (2001). Love and the commitment problem in romantic relations and friendship. *Journal of Personality and Social Psychology, 81,* 247–262. http://dx.doi.org/10.1037/0022-3514.81.2.247

Hardoon, D. R., Szedmak, S., & Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural Computation, 16,* 2639–2664. http://dx.doi.org/10.1162/0899766042321814

Harris, R. J., Young, A. W., & Andrews, T. J. (2012). Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 21164–21169. http://dx.doi.org/10.1073/pnas.1212207110

Hawk, S. T., Van Kleef, G. A., Fischer, A. H., & Van der Schalk, J. (2009). Worth a thousand words: Absolute and relative decodability of nonlinguistic affect vocalizations. *Emotion, 9,* 293–305.

Hertenstein, M. J., & Campos, J. J. (2004). The retention effects of an adult's emotional displays on infant behavior. *Child Development, 75,* 595–613. http://dx.doi.org/10.1111/j.1467-8624.2004.00695.x

Jack, R. E., Sun, W., Delis, I., Garrod, O. G. B., & 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

Johnson, G., & Connelly, S. (2014). Negative emotions in informal feedback: The benefits of disappointment and drawbacks of anger. *Human Relations, 67,* 1265–1290. http://dx.doi.org/10.1177/0018726714532856

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814. http://dx.doi.org/10.1037/0033-2909.129.5.770

Juslin, P. N., Laukka, P., & Bänziger, T. (2018). The mirror to our soul? Comparisons of spontaneous and posed vocal expression of emotion. *Journal of Nonverbal Behavior, 42,* 1–40. http://dx.doi.org/10.1007/s10919-017-0268-x

Keltner, D. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion, 10,* 155–172. http://dx.doi.org/10.1080/026999396380312

Keltner, D., & Kring, A. (1998). Emotion, social function, and psychopathology. *Review of General Psychology, 2,* 320–342. http://dx.doi.org/10.1037/1089-2680.2.3.320

Keltner, D., & Lerner, J. S. (2010). Emotion. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (5th ed., pp. 317–352). New York, NY: McGraw-Hill. http://dx.doi.org/10.1002/9780470561119.socpsy001009

Keltner, D., Tracy, J., Sauter, D. A., Cordaro, D. C., & McNeil, G. (2016). Expression of emotion. In L. F. Barrett & M. Lewis (Eds.), *Handbook of emotions* (pp. 467–482). New York, NY: Guilford Press.

Kraus, M. W. (2017). Voice-only communication enhances empathic accuracy. *American Psychologist, 72,* 644–654. http://dx.doi.org/10.1037/amp0000147

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. http://dx.doi.org/10.3389/fpsyg.2013.00353

Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology, 111,* 686–705. http://dx.doi.org/10.1037/pspi0000066

Laukka, P., & Juslin, P. N. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion, 31,* 182–191. http://dx.doi.org/10.1007/s11031-007-9063-z

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist, 46,* 819–834. http://dx.doi.org/10.1037/0003-066X.46.8.819

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401,* 788–791. http://dx.doi.org/10.1038/44565

Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (2014). In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. *Emotion, 14,* 145–160. http://dx.doi.org/10.1037/a0034287

Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9,* 2579–2605.

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (Vol. 3, pp. 211–234). New York, NY: Guilford Press.

Mehrabian, A., & Russell, J. (1974). *An approach to environmental psychology.* Cambridge, MA: MIT Press.

Mitchell, R. L. C., & Ross, E. D. (2013). Attitudinal prosody: What we know and directions for future study. *Neuroscience and Biobehavioral Reviews, 37,* 471–479. http://dx.doi.org/10.1016/j.neubiorev.2013.01.027

Nordström, H., Laukka, P., Thingujam, N. S., Schubert, E., & Elfenbein, H. A. (2017). Emotion appraisal dimensions inferred from vocal expressions are consistent across cultures: A comparison between Australia and India. *Royal Society Open Science, 4,* 170912. http://dx.doi.org/10.1098/rsos.170912

Oveis, C., Spectre, A., Smith, P. K., Liu, M. Y., & Keltner, D. (2013). Laughter conveys status. *Journal of Experimental Social Psychology, 65,* 109–115. http://dx.doi.org/10.1016/j.jesp.2016.04.005

Parsons, C. E., Young, K. S., Joensson, M., Brattico, E., Hyam, J. A., Stein, A., . . . Kringelbach, M. L. (2014). Ready for action: A role for the human midbrain in responding to infant vocalizations. *Social Cognitive and Affective Neuroscience, 9,* 977–984. http://dx.doi.org/10.1093/scan/nst076

Provine, R. R., & Fischer, K. R. (1989). Laughing, smiling, and talking: Relation to sleeping and social context in humans. *Ethology, 83,* 295–305. http://dx.doi.org/10.1111/j.1439-0310.1989.tb00536.x

Reiss, S. (1991). Expectancy model of fear, anxiety, and panic. *Clinical Psychology Review, 11,* 141–153. http://dx.doi.org/10.1016/0272-7358(91)90092-9

Rozin, P., & Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion, 3,* 68–75. http://dx.doi.org/10.1037/1528-3542.3.1.68

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76,* 574–586. http://dx.doi.org/10.1037/0022-3514.76.4.574

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. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110,* 145–172. http://dx.doi.org/10.1037/0033-295X.110.1.145

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., & Fischer, A. H. (2018). Can perceivers recognise emotions from spontaneous expressions? *Cognition and Emotion, 32,* 504–515. http://dx.doi.org/10.1080/02699931.2017.1320978

Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology, 31,* 97–113.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99,* 143–165. http://dx.doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23,* 1307–1351. http://dx.doi.org/10.1080/02699930902928969

Scherer, K. R. (2013). Vocal markers of emotion: Comparing induction and acting elicitation. *Computer Speech & Language, 27,* 40–58. http://dx.doi.org/10.1016/j.csl.2011.11.003

Scherer, K. R., & Bänziger, T. (2004). Emotional expression in prosody: A review and an agenda for future research. *Proceedings of Speech Prosody, 2004,* 359–366. Retrieved from http://isle.illinois.edu/sprosig/sp2004/PDF/Scherer-Baenziger.pdf

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

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology, 66,* 310–328. http://dx.doi.org/10.1037/0022-3514.66.2.310

Schüller, B., Elkins, A., & Scherer, K. (2017). Computational analysis of vocal expression of affect: Trends and challenges. In J. Burgoon, N. Magnenat-Thalmann, M. Pantic, A. Vinciarelli (Eds.), *Social signal processing* (pp. 56–58). Massachusetts, England: Cambridge University Press.

Scott, S. K., Lavan, N., Chen, S., & McGettigan, C. (2014). The social life of laughter. *Trends in Cognitive Sciences, 18,* 618–620. http://dx.doi.org/10.1016/j.tics.2014.09.002

Shiota, M. N., Campos, B., Oveis, C., Hertenstein, M. J., Simon-Thomas, E., & Keltner, D. (2017). Beyond happiness: Building a science of discrete positive emotions. *American Psychologist, 72,* 617–643. http://dx.doi.org/10.1037/a0040456

Shuman, V., Clark-Polner, E., Meuleman, B., Sander, D., & Scherer, K. R. (2017). Emotion perception from a componential perspective. *Cognition and Emotion, 31,* 47–56. http://dx.doi.org/10.1080/02699931.2015.1075964

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion, 9,* 838–846. http://dx.doi.org/10.1037/a0017810

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48,* 813–838. http://dx.doi.org/10.1037/0022-3514.48.4.813

Smoski, M., & Bachorowski, J. A. (2003). Antiphonal laughter between friends and strangers. *Cognition and Emotion, 17,* 327–340. http://dx.doi.org/10.1080/02699930302296

Snowdon, C. T. (2002). Expression of emotion in nonhuman animals. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 457–480). New York, NY: Oxford University Press.

Titze, I. R., & Martin, D. (1998). Principles of voice production. *The Journal of the Acoustical Society of America, 104,* 1148. http://dx.doi.org/10.1121/1.424266

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

Van Kleef, G. A. (2010). The emerging view of emotion as social information. *Social and Personality Psychology Compass, 4,* 331–343. http://dx.doi.org/10.1111/j.1751-9004.2010.00262.x

Vidrascu, L., & Devillers, L. (2005). Real-life emotion representation and detection in call centers data. In J. Tao, T. Tan, & R. W. Picard (Eds.), *Affective computing and intelligent interaction* (pp. 739–746). Berlin, Germany: Springer.

Watson, D., & Stanton, K. (2017). Emotion blends and mixed emotions in the hierarchical structure of affect. *Emotion Review, 9,* 99–104. http://dx.doi.org/10.1177/1754073916639659

Wu, Y., Muentener, P., & Schulz, L. E. (2017). One- to four-year-olds connect diverse positive emotional vocalizations to their probable causes. *Proceedings of the National Academy of Sciences of the United States of America, 114,* 11896–11901. http://dx.doi.org/10.1073/pnas.1707715114