Research Article

Psychological Science
1–11

DOI: 10.1177/0956797618778235
www.psychologicalscience.org/PS
SAGE

The Perception of Spontaneous and Volitional Laughter Across 21 Societies

Gregory A. Bryant[1,2], Daniel M. T. Fessler[2,3], Riccardo Fusaroli[4,5], Edward Clint[2,3], Dorsa Amir[6], Brenda Chávez[7], Kaleda K. Denton[8], Cinthya Díaz[7], Lealaiauloto Togiaso Duran[3], Jana Fanćovićová[9], Michal Fux[10], Erni Farida Ginting[11,12], Youssef Hasan[13], Anning Hu[14], Shanmukh V. Kamble[15], Tatsuya Kameda[16], Kiri Kuroda[16], Norman P. Li[17], Francesca R. Luberti[18], Raha Peyravi[19], Pavol Prokop[9,20], Katinka J. P. Quintelier[21], Hyun Jung Shin[22], Stefan Stieger[23,24], Lawrence S. Sugiyama[25], Ellis A. van den Hende[26], Hugo Viciana-Asensio[27], Saliha Elif Yildizhan[28], Jose C. Yong[17], Tessa Yuditha[11,29], and Yi Zhou[14]

[1]Department of Communication, University of California, Los Angeles; [2]UCLA Center for Behavior, Evolution, and Culture, University of California, Los Angeles; [3]Department of Anthropology, University of California, Los Angeles; [4]Interacting Minds Center, Aarhus University; [5]Department of Communication and Culture, Aarhus University; [6]Department of Anthropology, Yale University; [7]Department of Psychology, Pontificia Universidad Catolica del Peru; [8]Department of Ecology and Evolutionary Biology, University of California, Los Angeles; [9]Department of Biology, University of Trnava; [10]Department of Biblical and Ancient Studies, University of South Africa; [11]Jakarta Field Station, Max Planck Institute for Evolutionary Anthropology, Jakarta, Indonesia; [12]Faculty of Literature, Al Azhar Indonesia University; [13]Department of Social Sciences, Qatar University; [14]Department of Sociology, Fudan University; [15]Department of Psychology, Karnatak University; [16]Department of Social Psychology, University of Tokyo; [17]School of Social Sciences, Singapore Management University; [18]Evolution and Ecology Research Centre, University of New South Wales; [19]Department of Anthropology, Binghamton University; [20]Institute of Zoology, Slovak Academy of Sciences; [21]School of Business and Economics, Vrije Universiteit Amsterdam; [22]Department of Psychology, Pusan National University; [23]School of Psychology, University of Vienna; [24]Department of Psychology, Karl-Landsteiner University of Health Sciences; [25]Department of Anthropology, University of Oregon; [26]Department of Product Innovation and Management, Delft University of Technology; [27]Institute for Advanced Social Studies (IESA), CSIC, Córdoba, Spain; [28]Department of Biostatistics and Medical Informatics, Acıbadem Mehmet Ali Aydinlar University; and [29]Department of Education, Atma Jaya Catholic University

Abstract

Laughter is a nonverbal vocalization occurring in every known culture, ubiquitous across all forms of human social interaction. Here, we examined whether listeners around the world, irrespective of their own native language and culture, can distinguish between spontaneous laughter and volitional laughter—laugh types likely generated by different vocal-production systems. Using a set of 36 recorded laughs produced by female English speakers in tests involving 884 participants from 21 societies across six regions of the world, we asked listeners to determine whether each laugh was real or fake, and listeners differentiated between the two laugh types with an accuracy of 56% to 69%. Acoustic analysis revealed that sound features associated with arousal in vocal production predicted listeners' judgments fairly uniformly across societies. These results demonstrate high consistency across cultures in laughter judgments, underscoring the potential importance of nonverbal vocal communicative phenomena in human affiliation and cooperation.

Keywords

laughter, vocal communication, cross-cultural, emotion, speech, open data

Received 10/13/17; Revision accepted 4/17/18

Corresponding Author:
Gregory A. Bryant, University of California, Los Angeles, Department of Communication, 2225 Rolfe Hall, Los Angeles, CA 90095-1563
E-mail: gabryant@ucla.edu

Human social interaction relies on a complex suite of verbal and nonverbal communicative behaviors. Unlike language, across taxa, many nonverbal expressive behaviors have clear parallels in other species. Comparative analyses have revealed homologies in play vocalizations across mammals; in humans, this manifests as spontaneous laughter (Davila-Ross, Owren, & Zimmermann, 2009; Gervais & Wilson, 2005; Provine, 2000; Vettin & Todt, 2005). Consistent with this characterization of human laughter as a biologically evolved species-typical feature, laughter appears in every culture, evincing remarkable consistency in form (Provine, 2000). The functions of laughter are also plausibly universal. However, this is more difficult to determine, as laughter occurs embedded within a variety of social contexts, resulting in many laugh types. A growing research corpus potentially addresses questions of function by examining the contexts in which laughter is generated as well as laughter's social consequences (e.g., Otten, Mann, van Berkum, & Jonas, 2017; Scott, Lavan, Chen, & McGettigan, 2014). In contrast, much less is known about how laughter is perceived. Research has explored distinctions between spontaneous and volitional laughter (Bryant & Aktipis, 2014; Lavan, Scott, & McGettigan, 2016; McGettigan et al., 2013), judgments of affiliation in colaughter (Bryant et al., 2016), and how perceivers ascribe social functions to laughter (Wood, Martin, & Niedenthal, 2017). The phylogeny of laughter suggests an avenue through which, by investigating perceptions of laughter, one of the earliest functions of laughter can be explored. The human homologue of mammalian play vocalizations may have maintained the ancestral function of this trait, namely, to uniquely signal affiliation. If so, then listeners should be able to distinguish this signal from other forms of laughter—and, critically, this ability should be a species-typical trait, independent of the many facets of communication that differ across cultures.

Laughter is a family of vocalizations linked by a particular pattern of rhythmic respiratory and laryngeal activity (Bachorowski, Smoski, & Owren, 2001; Luschei, Ramig, Finnegan, Bakker, & Smith, 2006)—vocalizations that, with some notable exceptions (Provine, 2000), are often tied to feelings of mirth or joy. Laughs typically have a burstlike onset in which repeated oscillations of the glottis generate a series of bursts that decay over time in both energy and frequency (Provine & Yong, 1991). However, repetition is not essential, as a laugh can consist of only one burst as well. There is often, but not always, an associated perceived pitch in the bursts, resulting from the fundamental frequency (F_0) of vocal-fold vibration regimes during glottal oscillatory cycles. Laughter production in normal conversation exhibits systematic features, including constrained vowel and loudness patterning, consistent affective properties, and a rule-governed relationship between laugh bursts and speech (Bryant, 2011; Provine, 1993, 2000; Ruch & Ekman, 2001; Szameitat et al., 2009; Vettin & Todt, 2004).

In other mammals, play vocalizations are derived from ritualized breathing during rough-and-tumble play (Gervais & Wilson, 2005; Knutson, Burgdorf, & Panksepp, 1998; Provine, 2000). Although the rhythmic respiratory and laryngeal activity of human laughter constitute clear homologous aspects, human laughter differs from other primate play vocalizations in its higher proportion of voiced components—that is, more tonal, harmonically structured features attributable to vocal fold vibration (Davila-Ross et al., 2009). Intriguingly, voicing in laughter appears to be associated both with positive valence (Bachorowski & Owren, 2001) and with judgments of laughter as "fake" (Bryant & Aktipis, 2014). Such findings reveal the limited extent of knowledge regarding the relationships between physical properties of laughs and listeners' percepts. Although laughter's links to phylogenetically ancient play vocalizations indicate that some such perceptions should be independent of language, to date, only limited research has been conducted on laughter perception across cultures. Sauter, Eisner, Ekman, and Scott (2010) identified laughter as the most recognizable emotional vocalization across two disparate cultures (British and Himba). Bryant et al. (2016) found that listeners across 24 societies could detect friendship status on the basis of brief decontextualized clips of colaughter. These results reveal high perceptual sensitivity to this ubiquitous and ancient behavior.

Emotional vocal signals in humans are generated from a conserved production system shared by most social mammals (U. Jürgens, 2002). Humans also produce articulated speech using a largely distinct neural system (Ackermann, Hage, & Ziegler, 2014; Simonyan, 2014). Speech affords the imitation of a variety of sounds, including signals generated by the vocal emotion system such as laughter, crying, and pain shrieks. Nonverbal acted emotional vocalizations are acoustically distinct from their authentic counterparts, and the difference is perceptible (Anikin & Lima, 2018). However, cross-cultural findings are mixed, with some research reporting relatively low accuracy rates in discriminating play-acted vocal emotions from authentic expressions, as well as interactions between culture and emotion categories (R. Jürgens, Drolet, Pirow, Scheiner, & Fischer, 2013). Vocal emotion expressions are influenced by the vagal system, which extends to the recurrent laryngeal nerve (Ludlow, 2013). Thus, arousal in speakers can have direct effects on the vocal apparatus, including increased vocal fold tension, subglottal air

pressure, and glottal adduction rate, along with possible irregular vibration regimes of vocal fold tissue. Consequently, arousal in laughter is characterized by higher pitch, increased loudness, faster burst rate, and greater nontonal noise. The evolutionary introduction of volitional forms of expression that emulate genuine emotional signals created an arms race, pitting production dynamics against perceptual sensitivity: Vocalizers attempt to manipulate listeners by emitting sounds that falsely appear to reveal emotional states; in turn, listeners benefit from the ability to discriminate between honest indicators of vocalizers' emotional states and facsimiles thereof. We should therefore expect perceptual systems to strive to track relevant features to enhance the accuracy of social judgments.

We tested the above thesis by exploring cross-cultural recognition of dual vocal-production pathways in human laughter. Paralleling work on so-called Duchenne smiles (Gervais & Wilson, 2005), many proposed taxonomies of laughter distinguish between genuine and deliberate forms; this maps onto the aforementioned emotion-speech production distinction. Colingual listeners can discriminate between these basic laughter types (Bryant & Aktipis, 2014; Lavan, Rankin, Lorking, Scott, & McGettigan, 2017; Lavan et al., 2016; McGettigan et al., 2013), and neuroimaging work shows that these laugh types differentially activate brain regions during both production and perception (Lavan et al., 2017; McGettigan et al., 2013; Szameitat et al., 2010). Reflecting their respective production systems, spontaneous laughter and volitional laughter have different acoustic features. Spontaneous laughs have higher values on acoustic correlates of physical arousal, such as F_0, and shorter burst duration but also lower relative loudness, potentially because of the prolonged, regulated energy of volitional laughter produced by the speech system; they also often have fewer voiced elements, including a higher rate of intervoicing intervals (IVIs; Bryant & Aktipis, 2014; Lavan et al., 2016), which contributes to sound qualities that make them more similar to nonhuman animal vocalizations than volitional laughs (Bryant & Aktipis, 2014). The rate of IVI measures the proportion of the calls across a laugh not associated with voicing (i.e., nontonal), a ratio likely reflecting the extent of differential breath-control deployment during production. The percentage of unvoiced components per call is positively associated with colingual listeners' judgments of the laughs being real (Bryant & Aktipis, 2014; Wood et al., 2017) as well as with listeners' inability to distinguish slowed versions of spontaneous human laughs from nonhuman animal calls (Bryant & Aktipis, 2014).

Research to date suggests that laughs produced by the two production systems are distinct. Because this reflects the activity of two different species-typical vocalization mechanisms, and selection will have consistently favored the ability to distinguish between the two types of laughter, we expected that this distinction would be universally recognizable. The strongest test of this prediction examines listeners who vary substantially in their degree of linguistic and cultural similarity to the laughers. Because language and other aspects of culture shape many features of verbal performance (Henrich, Heine, & Norenzayan, 2010), if the ability to distinguish between the two types of laughs is evident across a broad spectrum of commonality or difference between producer and receiver, then this capacity for discrimination likely constitutes a biologically evolved, species-typical trait.

We explored whether listeners from around the world (see Fig. 1) were able to distinguish between the two laugh types as produced by English speakers. We predicted that participants would reliably identify the laugh types and, as found in earlier work (Bryant & Aktipis, 2014), that acoustic features associated with spontaneous production (e.g., arousal-linked features such as higher F_0 and higher rate of IVI) would predict their judgments.

Method

Participants

Given previous work on listeners' discrimination of laughter types, we predicted a medium-sized effect. An average sample size per study site of 40 participants at a significance level (p) of less than .05 would be sufficient to detect an effect size (Cohen's d) of 0.25 with 88% power (R package *pwr*; Champely et al., 2017). We recruited 884 participants (500 women, 384 men; age: M = 26.6 years, SD = 7.0) from 21 different societies across six regions of the world (for full demographic information, see the Supplemental Material available online). Participant recruitment varied across study sites, but all were asked to volunteer, and no participants were paid.

Laughter stimuli

The stimulus set, used in a previous study (Bryant & Aktipis, 2014), consisted of 36 audio recordings of laughs. Eighteen spontaneous laughs were taken from 13 natural conversations between pairs of female young adult American English speakers who were friends at the time of the conversation; recordings were made in a laboratory (16-bit amplitude resolution, 44.1-kHz sampling rate, uncompressed WAV files; Sony DTC recorder, Sony ECM-77B microphones; Bryant, 2010).

Fig. 1. Map showing the 21 study sites.

Complementing this set, 18 volitional laughs, produced in response to the impromptu request "now laugh" during the course of an unrelated project, were collected from a different set of 18 female young adult American English speakers; these also were recorded in a laboratory (16-bit amplitude resolution, 44.1-kHz sampling rate, uncompressed WAV files; MicroTrack recorder, M-Audio, Cumberland, RI). The laughs were duration matched and amplitude normalized. For a full description of the stimulus set, see Bryant and Aktipis (2014).

Procedure

The 36 laughter samples were presented in random order using SuperLab 4.0 (Cedrus, San Pedro, CA) experiment software. For study sites in which a language other than English was used in conducting the experiment (16 of 21), the instructions were translated beforehand by the respective investigators or by native-language translators recruited by them for this purpose. Customized versions of the experiment were then created for each of the study sites using the translated instructions and a run-only version of the software. For study sites in which literacy was limited or absent, the experimenter read the instructions aloud to each participant in turn. Before each experiment and after obtaining informed consent, participants were told that they would be listening to recordings of women laughing and that, after each trial, they would be asked to determine whether the laugh was real or fake. Specifically, participants were told,

> In some of the recordings, the women were asked to laugh but were not given any other reason for laughing (we call these fake laughs). Other recordings are of women laughing naturally while talking to a friend (we call these real laughs).

Participants performed 1 practice trial and then completed the full experiment consisting of 36 trials. The study was approved for all sites by the University of California, Los Angeles Institutional Review Board. For the complete text of instructions and questions used in the experiment, see the Supplemental Material.

Results

Judgment task

To evaluate listener accuracy, we used a model-comparison approach in which variables were entered into generalized linear mixed models, and effects on model fit were measured using the Akaike information criterion. The data were modeled using the *glmer* procedure of the *lme4* package (Bates, Maechler, Bolker, & Walker, 2018) in the statistical platform R (R Core Team, 2014). The best-fitting model was a generalized linear mixed model using the Laplace approximation,

Table 1. Results From the Best-Fitting Model of Judgment Accuracy of Spontaneous and Volitional Laughter

Factor	Variance	*SD*	Estimate	*SE*	*z*	$p(>\|z\|)$
		Random effects				
Participant	0.03005	0.1733				
Laugh trial	1.53619	1.2394				
Society × Laugh Condition	0.08939	0.2990				
		Fixed effects				
Intercept			0.6252	0.2389	2.617	0.009*
Laugh condition			0.1908	0.2188	0.872	0.383

*$p < .01$.

with fixed effects of laugh condition (spontaneous or volitional) and random effects of participant, laugh trial, and an interaction between societies sampled and laugh condition (see Table 1). Accuracy (percentage correct) was the dependent measure. Across all participants, the overall rate of correct judgments was 64% ($SD = 0.48$, range = 56–69), a performance significantly better than chance ($z = 3.50$, $p < .001$), and spontaneous and volitional laughs were recognized overall at similar rates ($z = 0.872$, $p = .38$). There were no significant sex differences in listeners' judgments. Figure 2 shows the rates of correct judgments for each study site.

The best-fitting model included an interaction between societies sampled and laugh condition, with participants from some study sites showing a tendency to respond with "fake" more often than "real," and other participant groups showing the reverse pattern. Signal detection analysis was used to separate sensitivity from response bias in the task. Receiver-operating-characteristic (ROC) curves for each society were drawn using the *pROC* package in R (Robin et al., 2011; see Fig. 3). See Table S6 in the Supplemental Material for all signal detection values, including area-under-the-curve (AUC) values associated with the ROC figure.

An exploratory analysis of the possible impacts of six estimated demographic variables (English fluency, mass media exposure, mass media exposure in English, education, community size, and economic mode; see the Supplemental Material) on participants' response patterns revealed that societies' economic mode was most associated with a tendency to judge laughs as being real. Economic mode refers to a rough categorization based on principal economic activities and market integration. For example, the Shuar in Ecuador live in small villages and have minimal dependence on market exchanges, whereas highly industrialized societies such as the United States have maximal dependence on market exchanges. Figure S2 in the Supplemental Material reveals an overall pattern of increased responses of "real" in societies with greater industrialization and more reliance on skilled professionals. For all model comparisons and complete demographic analysis, see Table S7 in the Supplemental Material.

Acoustic analysis

Acoustic features, including the frequency and temporal dynamics of voiced and unvoiced segments, were automatically extracted from the individual laugh segments, following a procedure analogous to that used by Bryant et al. (2016). The acoustic features were used to statistically reconstruct (a) the distinction between spontaneous and volitional laughs and (b) the rate at which participants judged each laugh as real (i.e., spontaneous). We used a 5-fold cross-validated process wherein a Lasso algorithm (Tibshirani, 1996) first individuated key features (see Table S4 in the Supplemental Material); then, these were assessed in multiple logistic (for judgments of real vs. fake) and linear (for judgment rate) regressions. Because cross-validation is a stochastic process, we repeated the process 100 times to ensure stability of the results. We report cross-validated performance of the model (adjusted R^2 for linear regression and ROC curve for logistic regression), including 95% confidence intervals (CIs) on the repetitions and beta coefficients for the same models fitted on the full data set.

The acoustic-based model reliably predicted participants' judgments, employing coefficient of variation of intensity ($\beta = 0.5$, $SE = 0.09$, $p < .001$), mean pitch ($\beta = 0.41$, $SE = 0.09$, $p < .001$), and the mean absolute deviation of harmonics-to-noise ratio ($\beta = -0.46$, $SE = 0.09$, $p < .001$). The model could explain 63.9% of the variance (R^2; 95% CI = [55.5, 69.8]). Figure 4 displays the cross-validated model predictions (*x*-axis) against the actual mean judgments reported by participants (*y*-axis).

We were also able to reliably discriminate spontaneous from volitional laughs independent of participants' judgments, employing the rate of IVI ($\beta = 2.14$, $SE = 0.89$, $p = .016$), harmonics-to-noise ratio interquartile range ($\beta = -0.97$, $SE = 0.68$, $p = .15$), and median ($\beta =$

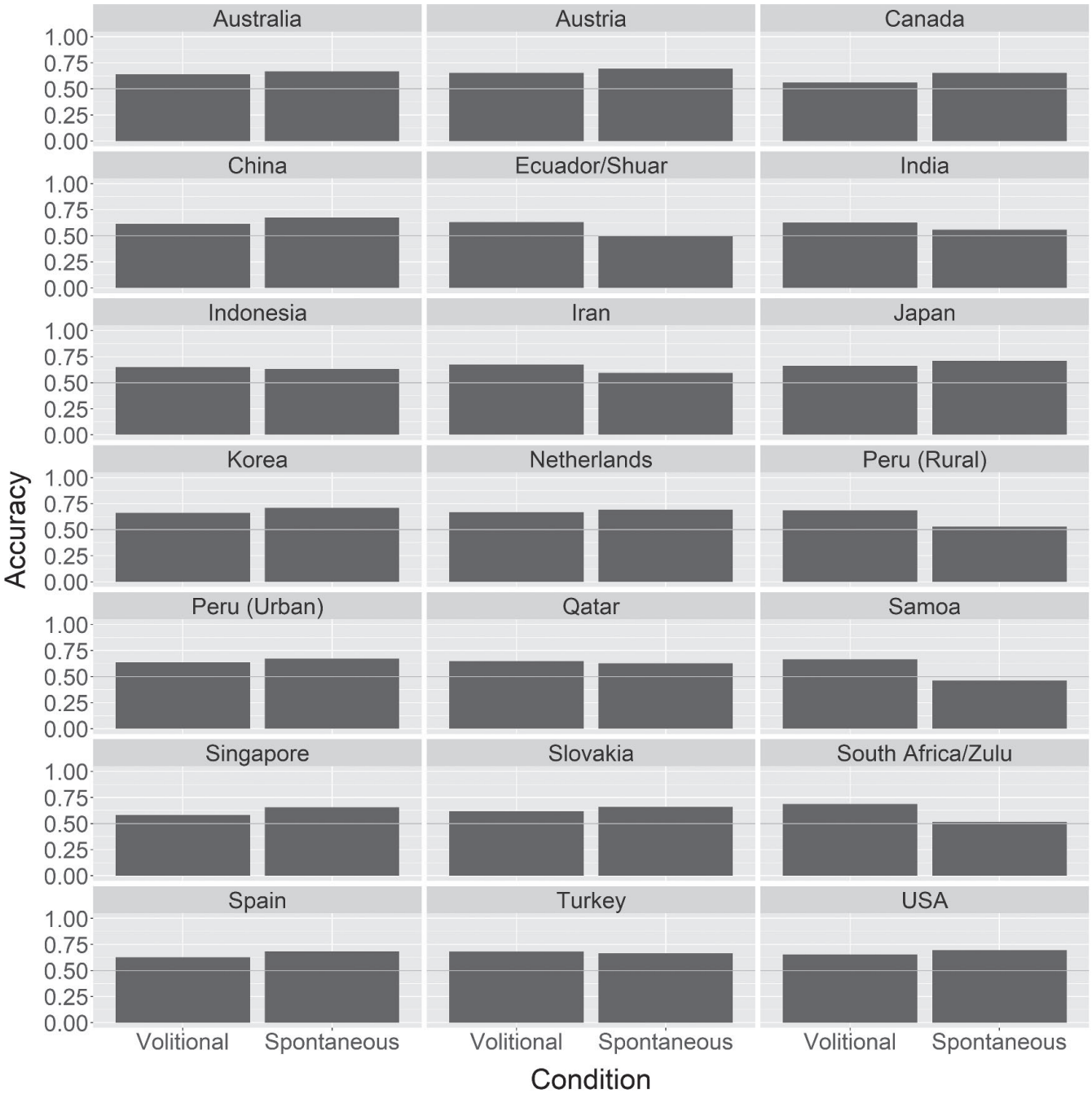


Fig. 2. Accuracy (overall proportion of correct judgments) in each study site broken down by laugh condition (volitional and spontaneous). Chance performance is represented by 0.50. In every society sampled, overall accuracy, collapsing across categories, was significantly better than chance.

−1.14, $SE = 0.66$, $p = .087$). The model had an estimated AUC of 83.32% (95% CI = [69.91, 89.51]), with an accuracy of 76.97% (95% CI = [63.89, 86.11]), a sensitivity of 79.61% (95% CI = [66.67, 88.89]), and a specificity of 74.33% (95% CI = [61.11, 83.33]).

Across societies, laughs that had higher intensity variability, higher pitch, and lower harmonics-to-noise-ratio variability were more likely to be judged as real. These features could also accurately discriminate spontaneous and volitional laughs (AUC: 64.79%, 95% CI = [52.16, 75]; accuracy: 64.44%, 95% CI = [55.56, 75]), although not as accurately as the optimal features identified by our analysis. For complete details of the acoustic analysis, see the Supplemental Material.

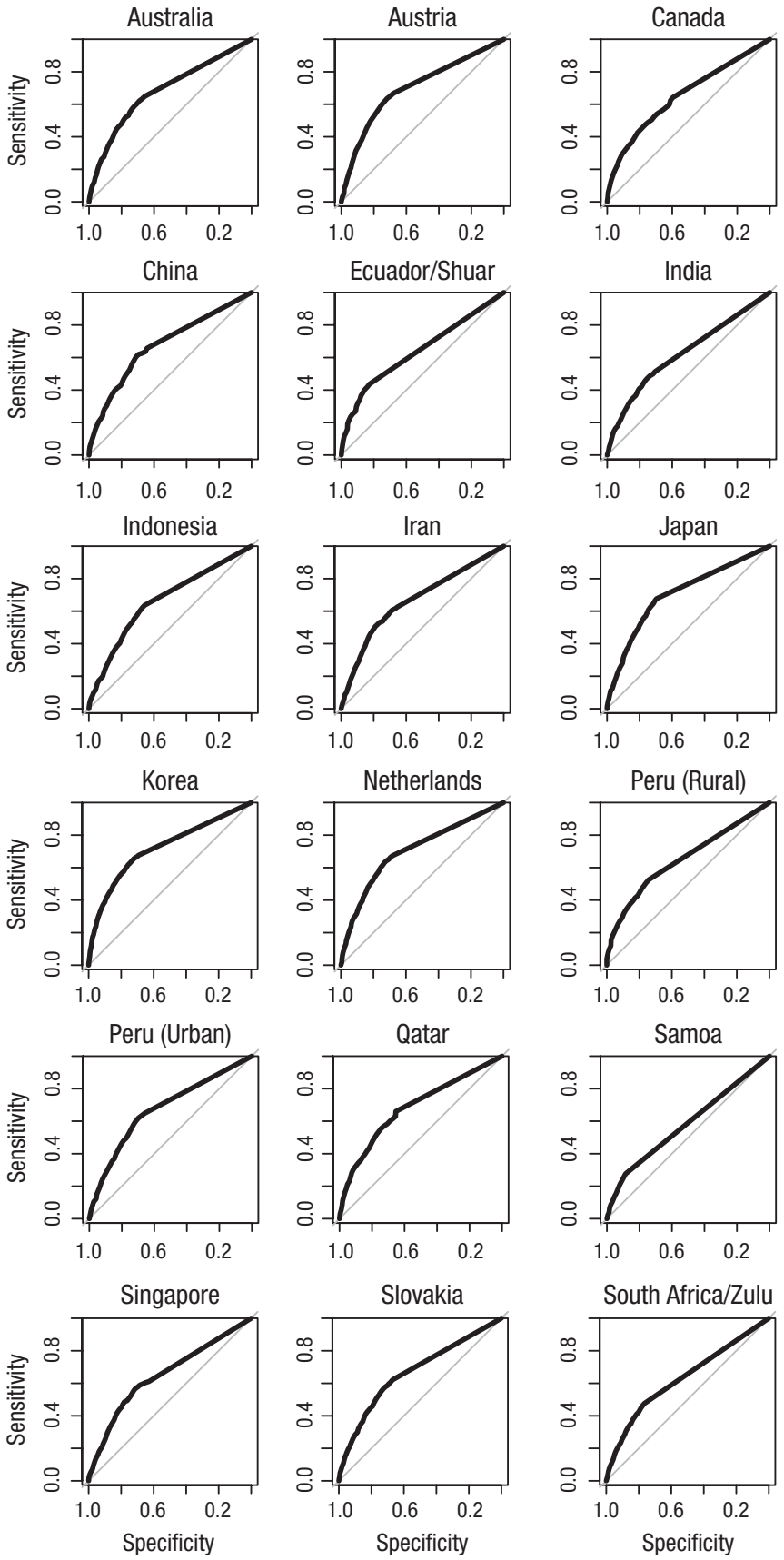


Fig. 3. *(continued on next page)*

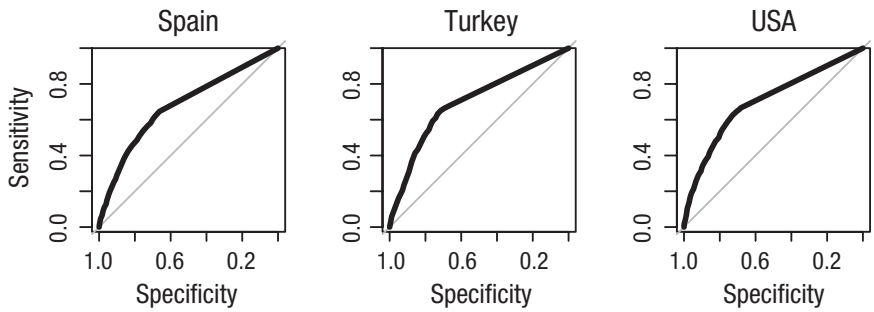


Fig. 3. Sensitivity as a function of specificity in each study site. Sensitivity and specificity are measures of how well a binary classification test performs. Sensitivity is the ability of the test to correctly identify participants exhibiting a given behavior (i.e., the true positive rate), whereas specificity is the ability of the test to correctly identify participants not exhibiting the given behavior (i.e., the true negative rate). Arbitrarily setting spontaneous laughter as the condition of interest, we defined true positives as correctly identifying spontaneous laughs and true negatives as correctly identifying volitional laughs. Receiver-operating-characteristic (ROC) curves represent the trade-off between sensitivity and specificity, as the cutoff point was systematically varied. Thus, the area between the ROC curve and the main diagonal (the area under the curve) represents overall performance of the decision-making process independent of response bias (i.e., independent of bias on the cutoff point). The bigger the area, the better-performing the model.

Discussion

Our results show that, around the world, regardless of their culture, native language, or cultural or linguistic similarity to the vocalizers, people reliably distinguished spontaneous and volitional laughter. In every society, participants correctly identified laugh types above chance, and judgments of spontaneity were associated with acoustic features likely tied to arousal in the vocalizers—specifically, greater intensity variability, higher pitch, and increased noisy features. These results are highly consistent with studies to date examining the perception of spontaneous and volitional laughter

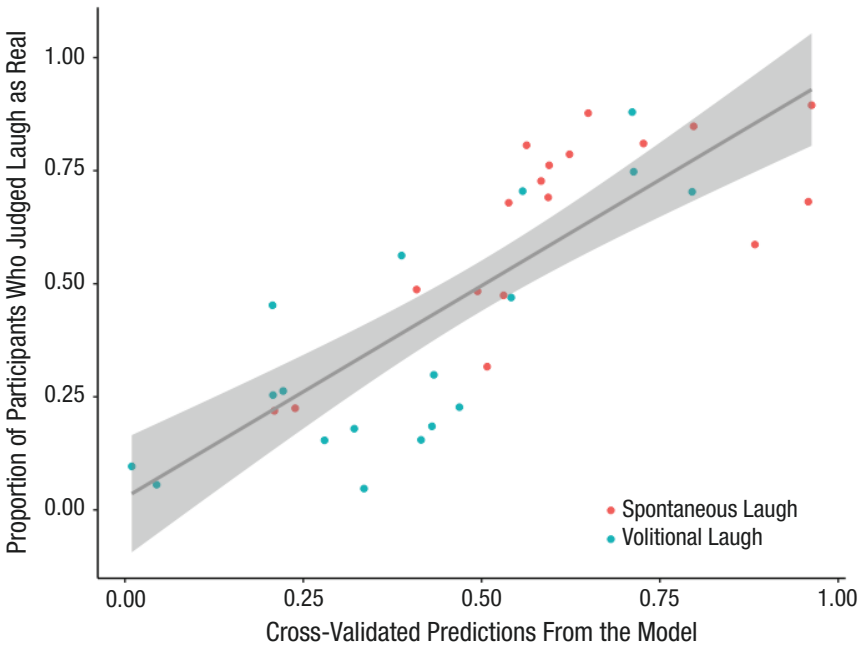


Fig. 4. Scatterplot (with best-fitting regression line) showing the correlation between participants' judgments (collapsed across all societies) of a laugh as being real and predicted values using the acoustic features selected by the statistical model. The error band indicates the 95% confidence interval.

within cultures; acoustic correlates of arousal have been previously shown to be associated with judgments of laughter genuineness (e.g., Bryant & Aktipis, 2014; Lavan et al., 2017; Lavan et al., 2016; McGettigan et al., 2013; Wood et al., 2017). But we also found some differences across cultures in judgment patterns of spontaneous and volitional forms, with small-scale societies, in particular, tending to judge tokens overall as more likely to be fake (for details and discussion, see the Supplemental Material). Other recent work (R. Jürgens et al., 2013) also found interesting interactions between encoding conditions (authentic emotional expressions vs. play-acted expressions) and culture, an issue that deserves more attention.

Our group has shown previously that, in 24 societies, listeners were able to determine, on the basis of brief clips of colaughter, whether dyads of native speakers of American English were friends or strangers (Bryant et al., 2016). Consonant with the thesis that, reflecting genuine prosocial emotions, spontaneous laughter constitutes an honest signal of affiliation—one imperfectly emulated in volitional laughter—the acoustic features associated with identifying friends in that study were similar to the features of spontaneous laughs described here, namely, features associated with speaker arousal. Taken together, these findings demonstrate that listeners are sensitive to acoustic features indicating emotional arousal in speakers and suggest an adaptive laughter-signaling system that inherently involves the triggering of emotional vocalizations with arousal-linked acoustic properties. Listeners everywhere can discriminate between two broad laughter categories; however, a fuller taxonomy of laughter types is needed. Moreover, it is possible that we inflated the distinctiveness of our general categories by using volitional laughs that did not originate in natural social contexts (i.e., they were produced on command). As Provine (2012) noted, voluntary productions of laughter differ in many ways from spontaneous laughs. Our stimulus set also included only female laughers. Future work should examine the dynamics of cross-sex laugh perception across disparate cultures as well as potential affective properties of low-pitched, aggressive laughter afforded by male vocalizers.

The social ecology of nonverbal expression within a dual-vocal-systems framework requires a designation not only of which system produces a vocalization but also of how it is deployed in social interaction (see also Wood et al., 2017). A laugh generated by the speech system is not necessarily a selfish manipulation; indeed, as suggested above, in many contexts, such laughs indicate cooperative intent. A brief volitional laugh that signals, for instance, a conversational turn or the recognition of some encrypted (i.e., implicit) content is cooperative in both the Gricean-conversational and the biological sense (Flamson & Bryant, 2013). Future work should therefore examine the complexities of how laughter signals interact with language use. Much of what people laugh about in social interaction is tied to what people are saying; variations in the production and interactive timing of laughter can reveal rich information regarding the underlying cognitive processes in conversation. Finally, there is much to learn about how laughing fits into the multimodal contexts of ordinary interpersonal communication. The more closely we examine laughter, the more evident are its intricacies.

Action Editor

Ayse K. Uskul served as action editor for this article.

Author Contributions

The first four authors are listed in order of the importance of their contributions. G. A. Bryant designed the hypothesis and methods, conducted the core analyses, and wrote and revised the manuscript. D. M. T. Fessler envisioned the cross-cultural component, organized the cross-cultural research, and assisted in writing and revising the manuscript. R. Fusaroli conducted the acoustic and signal detection analyses and contributed the corresponding draft text. E. Clint managed the cross-cultural research. All remaining authors contributed data and are listed in alphabetical order. All the authors approved the final manuscript for submission.

ORCID iDs

Riccardo Fusaroli https://orcid.org/0000-0003-4775-5219
Kiri Kuroda https://orcid.org/0000-0001-7278-1430
Norman P. Li https://orcid.org/0000-0002-4059-1613

Acknowledgments

We thank our participants and Alex Espinosa-Fornieles, Solomua Fiamatai, Togiaso Fiamatai, Thenjiwe Hlatshwayo, Yoon Kyoung Lee, Sarah Popelka, Shaudee Salari, and Pipi'imaleeleele Tonumaipe'a for assistance.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797618778235

Open Practices

All data have been made publicly available via Harvard Dataverse and can be accessed at https://dataverse.harvard.edu/dataverse/laughterperception. Materials for this experiment have not been made publicly available, and the design and analysis plans were not preregistered.

References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral & Brain Sciences*, *37*, 529–546.

Anikin, A., & Lima, C. F. (2018). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *The Quarterly Journal of Experimental Psychology*, *71*, 622–641.

Bachorowski, J. A., & Owren, M. J. (2001). Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science*, *12*, 252–257.

Bachorowski, J. A., Smoski, M. J., & Owren, M. J. (2001). The acoustic features of human laughter. *Journal of the Acoustical Society of America*, *110*, 1581–1597.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2018). *lme4: Linear mixed-effects models using 'Eigen' and S4* (R package version 1.1-7) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/lme4/index.html

Bryant, G. A. (2010). Prosodic contrasts in ironic speech. *Discourse Processes*, *47*, 545–566.

Bryant, G. A. (2011). Verbal irony in the wild. *Pragmatics & Cognition*, *19*, 291–309.

Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evolution & Human Behavior*, *35*, 327–335.

Bryant, G. A., Fessler, D. M. T., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. A., . . . Zhou, Y. (2016). Detecting affiliation in colaughter across 24 societies. *Proceedings of the National Academy of Sciences, USA*, *113*, 4682–4687.

Champely, S. (2017). *pwr: Basic functions for power analysis* (Version 1.2-2) [Computer software]. Retrieved from http://cran.r-project.org/package=pwr

Davila-Ross, M., Owren, M., & Zimmermann, E. (2009). Reconstructing the evolution of laughter in great apes and humans. *Current Biology*, *19*, 1106–1111.

Flamson, T. J., & Bryant, G. A. (2013). Signals of humor: Encryption and laughter in social interaction. In M. Dynel (Ed.), *Developments in linguistic humour theory* (pp. 49–74). Amsterdam, The Netherlands: John Benjamins.

Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *Quarterly Review of Biology*, *80*, 395–430.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral & Brain Sciences*, *33*, 61–83.

Jürgens, R., Drolet, M., Pirow, R., Scheiner, E., & Fischer, J. (2013). Encoding conditions affect recognition of vocally expressed emotions across cultures. *Frontiers in Psychology*, *4*, Article 111. doi:10.3389/fpsyg.2013.00111

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, *26*, 235–238.

Knutson, B., Burgdorf, J., & Panksepp, J. (1998). Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *Journal of Comparative Psychology*, *112*, 65–73.

Lavan, N., Rankin, G., Lorking, N., Scott, S., & McGettigan, C. (2017). Neural correlates of the affective properties of spontaneous and volitional laughter types. *Neuropsychologia*, *95*, 30–39.

Lavan, N., Scott, S. K., & McGettigan, C. (2016). Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter. *Journal of Nonverbal Behavior*, *40*, 133–149.

Ludlow, C. L. (2013). Motor control and biomechanics of laryngeal and pharyngeal muscles. In L. K. McLoon & F. H. Andrade (Eds.), *Craniofacial muscles: A new framework for understanding the effector side of craniofacial muscle control* (pp. 167–183). New York, NY: Springer.

Luschei, E. S., Ramig, L. O., Finnegan, E. M., Bakker, K. K., & Smith, M. E. (2006). Patterns of laryngeal EMG and the activity of the respiratory system during spontaneous laughter. *Journal of Neurophysiology*, *96*, 442–450.

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2013). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral Cortex*, *25*, 246–257.

Otten, M., Mann, L., van Berkum, J. J., & Jonas, K. J. (2017). No laughing matter: How the presence of laughing witnesses changes the perception of insults. *Social Neuroscience*, *12*, 182–193.

Provine, R. R. (1993). Laughter punctuates speech: Linguistic, social and gender contexts of laughter. *Ethology*, *95*, 291–298.

Provine, R. R. (2000). *Laughter: A scientific investigation*. New York, NY: Penguin.

Provine, R. R. (2012). *Curious behavior: Yawning, laughing, hiccupping, and beyond*. Cambridge, MA: Harvard University Press.

Provine, R. R., & Yong, Y. L. (1991). Laughter: A stereotyped vocalization. *Ethology*, *89*, 115–124.

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), Article 77. doi:10.1186/1471-2105-12-77

Ruch, W., & Ekman, P. (2001). The expressive pattern of laughter. In A. Kaszniak (Ed.), *Emotion, qualia, and consciousness* (pp. 426–443). Tokyo, Japan: Word Scientific.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, USA*, *107*, 2408–2412.

Scott, S. K., Lavan, N., Chen, S., & McGettigan, C. (2014). The social life of laughter. *Trends in Cognitive Sciences*, *18*, 618–620.

Simonyan, K. (2014). The laryngeal motor cortex: Its organization and connectivity. *Current Opinion in Neurobiology*, *28*, 15–21.

Szameitat, D. P., Alter, K., Szameitat, A. J., Wildgruber, D., Sterr, A., & Darwin, C. J. (2009). Acoustic profiles of

distinct emotional expressions in laughter. *Journal of the Acoustical Society of America*, *126*, 354–366.

Szameitat, D. P., Kreifelts, B., Alter, K., Szameitat, A. J., Sterr, A., Grodd, W., & Wildgruber, D. (2010). It is not always tickling: Distinct cerebral responses during perception of different laughter types. *NeuroImage*, *53*, 1264–1271.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, *58*, 267–288.

Vettin, J., & Todt, D. (2004). Laughter in conversation: Features of occurrence and acoustic structure. *Journal of Nonverbal Behavior*, *28*, 93–115.

Vettin, J., & Todt, D. (2005). Human laughter, social play, and play vocalizations of nonhuman primates: An evolutionary approach. *Behaviour*, *142*, 217–240.

Wood, A., Martin, J., & Niedenthal, P. (2017). Towards a social functional account of laughter: Acoustic features convey reward, affiliation, and dominance. *PLOS ONE*, *12*(8), Article e0183811. doi:10.1371/journal.pone.0183811

SUPPLEMENTARY MATERIALS

The perception of spontaneous and volitional laughter across 21 societies

Gregory A. Bryant, Daniel M. T. Fessler, Riccardo Fusaroli, Edward Clint, Dorsa Amir, Brenda Chavez, Kaleda K. Denton, Cinthya Diaz, Lealaiauloto Togiaso Duran, Jana Fančovičová, Michal Fux, Youssef Hasan, Anning Hu, Shanmukh V. Kamble, Tatsuya Kameda, Norman P. Li, Francesca R. Luberti, Raha Peyravi, Pavol Prokop, Katinka Quintelier, Hyun Jung Shin, Stefan Stieger, Lawrence S. Sugiyama, Wataru Toyokawa, Ellis A. van den Hende, Hugo Viciana-Asensio, Saliha Elif Yildizhan, Jose C. Yong, Tessa Yuditha, Yi Zhou

I. Demographic information and analysis

Table S1. Sex and age breakdown by study site.

Group	Region	N	Men	Women	Mean Age	Age SD	Age Range
S. Africa/Zulu	Africa	100	50	50	35	14.9	18-73
China	Asia	33	15	18	23.3	2.6	20-32
India	Asia	34	14	20	24.1	3	20-30
Indonesia	Asia	39	19	20	20.5	3.1	19-39
Iran	Asia	39	20	19	30	10	19-63
Japan	Asia	57	30	27	21.8	2.3	18-34
Korea	Asia	58	28	30	20.7	2	18-28
Qatar	Asia	31	0	31	22.4	2.4	19-27
Singapore	Asia	40	22	18	22.8	3.5	18-33
Samoa	Oceania	42	23	19	40.5	16.8	18-80
Australia	Oceania	32	18	14	28.8	13.4	19-72
Austria	Europe	29	8	21	30.5	6.7	22-48
Netherlands	Europe	43	25	18	23	2.3	18-27
Slovakia	Europe	71	13	58	24.4	7.4	18-60
Spain	Europe	35	18	17	22.8	4.1	18-33
Turkey	Europe	29	14	15	26.3	11.2	18-70
Canada	N. America	30	16	14	36.6	14.3	18-63
USA	N. America	48	16	32	19.4	2.6	18-24
Ecuador/Shuar	S. America	33	14	19	33.5	11.7	18-60
Peru (Urban)	S. America	30	9	21	21	2.5	18-27
Peru (Rural)	S. America	31	12	19	31.1	11	18-58
TOTALS	**6**	**884**	**384**	**500**	**26.6**	**7.0**	**18-80**

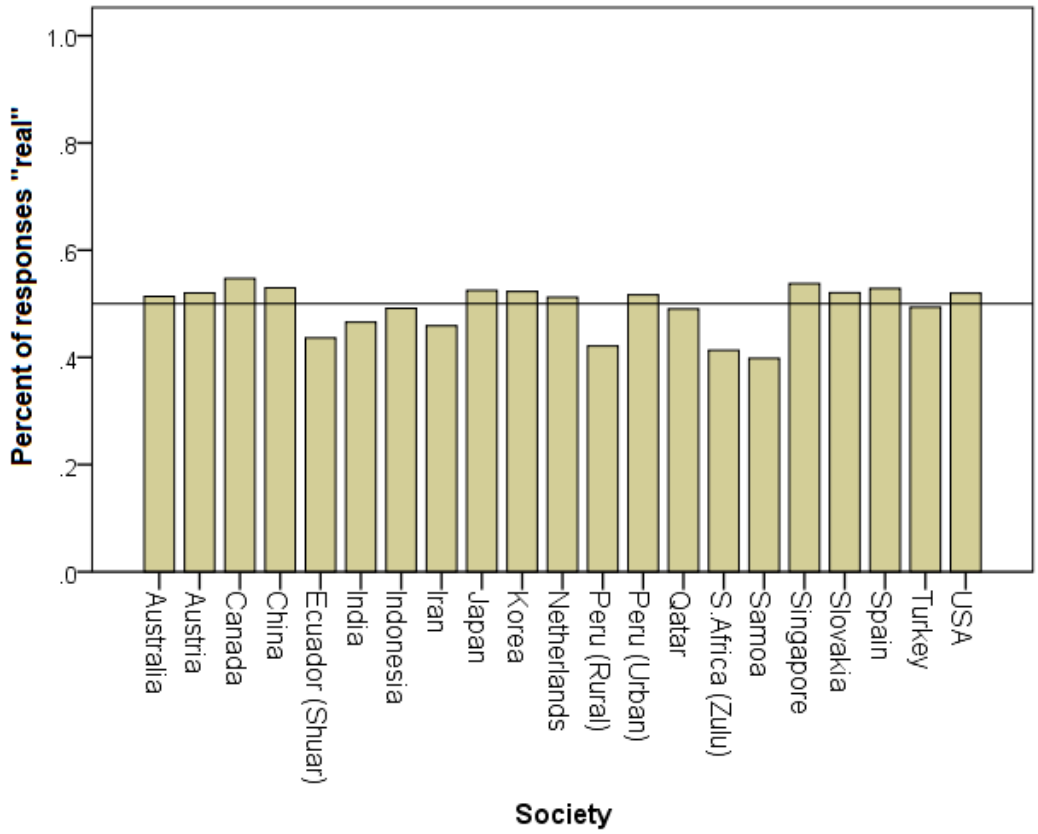
Table S2. Demographic information across 21 societies

Country Ethnic group	Participant's native language	Language in which experiment was conducted	Typical participant's English fluency	Mass media exposure	~% mass media in English	Typical participant's education	Community or city scale (number of people)	Economic mode(s) of participants
Australia	English	English	primary language	extensive	100	Some college	small towns (<5000) and large cities (>500000)	Industrial: low and highly skilled
Austria	German	German	moderate	extensive	50	Some college	small towns (<5000) and large cities (>500000)	Industrial: low and highly skilled
Canada	English	English	primary language	extensive	100	Some college	large cities	Industrial: low and highly skilled
China	Chinese	Chinese (Mandarin)	moderate	daily	50	Some college	large cities	Industrial: low and highly skilled
Ecuador/Shuar	Shuar	Spanish	none	minimal	<25	4-7 years	small villages (<200 people)	Hunting and gathering, small scale horticulture, agriculture, pastoralism, trade
India	Kannada	English	moderate	extensive	50	Some college	medium cities	Industrial: highly skilled
Indonesia	Jakartan	Formal Indonesian	moderate	extensive	75	8-12 years	large cities	Small-scale trade Industrial: low and highly skilled
Iran	Farsi	Farsi	minimal to moderate	extensive	25	College degree	large cities (>500000)	Skilled professional in office or institutional setting
Japan	Japanese	Japanese	minimal	daily	<25	8-12 years	large cities	Industrial: low and highly skilled
Korea	Korean	Korean	moderate	extensive	25	Some college	large cities	Small-scale: trade Industrial: low and highly skilled
Netherlands	Dutch	Dutch	fluent (as second language)	extensive	75	Some college	small and medium cities	Industrial: highly skilled
Peru (rural)	Spanish	Spanish	minimal	extensive	<25	8-12 years	small towns	Small-scale horticulture, agriculture, pastoralism Industrial: low skill
Peru (urban)	Spanish	Spanish	moderate	extensive	50	Some college	large cities	Industrial: highly skilled
Qatar	Arabic	Arabic	minimal	extensive	50	College degree	large cities	Industrial: low and highly skilled
Samoa	Samoan	Samoan	moderate	extensive	<50	8-12 years	large villages (200-1000) and small towns	Small-scale: trade Industrial: low and highly skilled
Singapore	English	English	primary language	extensive	75	Some college	large cities	Industrial: low and highly skilled
Slovakia	Slovak	Slovak	minimal	extensive	<25	College degree	large towns (5000-10000) and small cities	Industrial: highly skilled
South Africa/Zulu	isiZULU	isiZulu	minimal	occasional	<25	8-12 years	large villages (200-1000) and small towns	Small-scale agriculture, pastoralism, trade Industrial: low-skill
Spain	Spanish	Spanish	minimal	extensive	<25	Some college	small and medium cities	Industrial: low and highly skilled
Turkey	Turkish	Turkish	moderate	extensive	50	8-12 years	large cities	small scale trade industrial: low and highly skilled
USA	English,	English	primary language	extensive	100	some college	large cities	Industrial: low and highly skilled

Impact of demographic variables on response patterns

In any given trial, respondents decided whether the presented laugh was "real" or "fake." The percent responding "real" are identifying the token as a spontaneous or real laugh. Across our samples, participants in societies that can be characterized as small-scale and/or rural tended to respond with "real" less than 50% of the time (i.e., biasing their responses toward "fake"). See Figure S1.

Figure S1. Responses of "real laugh" across 21 societies



For each study site, the respective researcher provided descriptive demographic information, including average levels of English fluency, exposure to mass media (both in any language, and the proportion of media in English), education, community scale (small villages to large cities), and the economic mode of the society (small-scale, traditional skills to industrialized, professional skills). As an exploratory analysis examining the possible role of population characteristics affecting judgments of laughter (using the same analytical method described in the main text), we modeled overall response patterns as a function of the six different demographic variables. We started with the best fitting model from the main analysis, and created six new models, each one adding only one demographic variable. Based on variance estimates for each demographic variable, and AIC comparison across models, economic mode was most associated with the pattern of responses. See Table S3. Figure S2 shows the pattern of responses of "real" across each demographic variable. There is a clear resemblance in each of these variables—higher overall rates of judging a laugh as "real" when originating from societies with higher rates of industrialization and professional skill. Because variation within study sites has been eliminated, care should be taken in interpreting demographic data of this kind (Kievit,

Frankenhuis, Waldorp, & Borsboom, 2013). Nevertheless, these data do suggest that economic mode, community scale, and media exposure potentially play a role in shaping people's responses in our task, whereas familiarity with English appears less important.

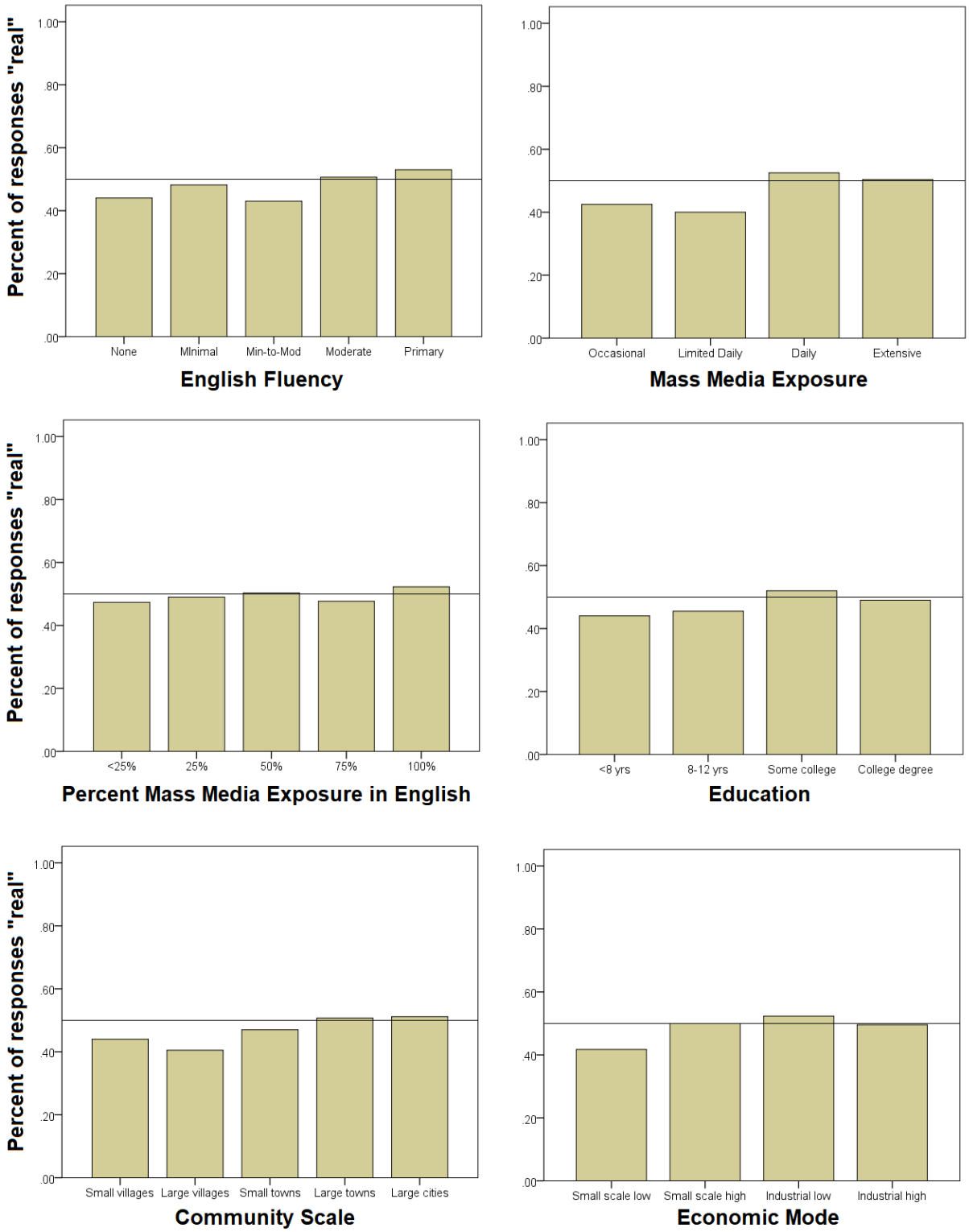
Table S3. Best-fit model of response patterns in judgments of "real" and "fake."

Random effects			**Fixed effects**				
Factor	***Variance***	***STD***	***Factor***	***Estimate***	***SE***	***z***	***Pr(>|z|)***
Subject	0.26071	0.5106					
Laugh Trial	1.95308	1.3975					
Society × Condition	0.04556	0.2135					
			(Intercept)	-0.7256	0.2738	-2.650	0.0080
			Condition	0.4161	0.2223	1.872	0.0612
			Econ mode2	0.4927	0.1393	3.538	0.0004
			Econ mode3	0.6172	0.1097	5.627	1.84e-08
			Econ mode4	0.4509	0.1218	3.702	0.0002

Participants from small-scale societies were more accurate assessing volitional laughs than spontaneous laughs, a pattern reversed in most of the other populations. This could reflect either greater skepticism regarding laughers' emotional engagement, or greater accuracy differentiating between spontaneous and volitional laughs. We labeled laughs produced in natural conversation between friends as spontaneous; however, some may actually have been volitional. In contrast, laughs labeled volitional were produced on command; a naïve undergraduate would be unlikely to produce a spontaneous laugh in such a context. Hence, it may be that, rather than being more skeptical, participants from small-scale societies were slightly more accurate. This pattern might reflect the greater importance in small-scale societies of deep and complex social relationships. In such a context, accurately judging another's degree of emotional engagement is critical in predicting behavior, hence listeners may be acutely attuned to indications of affective state. In contrast, large-scale societies have more relatively anonymous interactions in which the parties' respective behaviors are shaped primarily by roles in the social structure; such encounters are smoothed by the polite exchange of superficial tokens independent of genuine sentiment—potentially leading listeners to attend less closely to indices thereof.

Our stimuli suffer the limitations that i) some of our spontaneous laughs may actually be volitional, and ii) our volitional laughs were not produced in a social context, and thus may differ from more naturalistic volitional laughs. Taken together, these limitations indicate that our results may underestimate perceivers' discriminative ability because of added noise. Moreover, acoustic analyses identifying features of laughs associated with these eliciting conditions might not capture additional variation introduced by other possible eliciting conditions, such as non-interactive spontaneous laughs. Finally, it is important to note that the overall categorization of laughs as spontaneous, and spontaneous laughers as more aroused, are indirect as we did not directly collect measures from the laugh producers. These are important issues that future research should address.

Figure S2. Percent of responses “real” as a function of six demographic variables.



II. Acoustic Analysis and Signal Detection Analysis

Measures

Having demonstrated that participants could accurately judge whether laughs are spontaneous or volitional, we then measured a wide range of acoustic features of the laughter to identify which features would best explain the variance in participants' judgments. We examined the 36 laughs used in the experiment, 18 spontaneous and 18 volitional.

For each individual laugh within a given audio clip we measured the intervoicing intervals. We first calculated bout duration for each laugh from the onset of visible acoustic energy as viewed in a spectrogram (FFT method, window length: 0.005 s., time steps: 1000, frequency steps: 250, Gaussian window shape, dynamic range: 50 dB) to the offset of energy in the final call, or bout-final inspiratory element. Calls were counted based on audible and visible separated voiced energy. Mean intervoicing interval (IVI) was calculated as the summed lengths of all unvoiced intervals between calls (i.e., voiced call offset to voice call onset) divided by call number minus one. Unvoiced portions were determined by a lack of formant structure as viewed through a spectrogram with settings described above, and lack of periodicity with standard pitch range values. Finally, rate of IVI was calculated using the following formula:

$$\frac{\left(\frac{(\sum x_i)}{(c-1)}\right)}{\left(\frac{d}{c}\right)}$$

where x_i are the inter-voicing interval values, c is the total call number, and d is the bout duration of the series. This measure captures the averaged rate of unvoiced segments per call across a laugh bout.

Using Covarep (Degottex, 2014), we extracted fundamental frequency (F_0) (frequency range = 70-400 Hz), intensity, and harmonics-to-noise ratio of the laughs every 10 ms. F_0 values were converted to a logarithmic scale to approximate perceptual pitch. Per each of these measures, we calculated traditional descriptive statistics and (except for harmonics-to-noise ratio) temporal dynamics measures.

Descriptive statistics. We calculated: a) the total, voiced and unvoiced duration of each laugh, as well as the rate of intervoicing interval (IVI), b) the mean, standard deviation, median, interquartile range, coefficient of variation (standard deviation divided by the mean) and mean absolute deviation of pitch, intensity, and harmonics-to-noise ratio.

Temporal dynamics measures. Traditional descriptive statistics do not capture other crucial aspects of time-series properties such as their regularity over time and the temporal dependence between successive data points. These properties express the stability and complexity of voice production and have proven particularly useful to assess vocal behavior in a wide variety of contexts (e.g. Cummins et al., 2015; Fusaroli et al., 2016; Washington et al., 2012). To assess these temporal dynamics we employed two non-linear methods: a) Recurrence Quantification Analysis (RQA) of both voiced/unvoiced sequences and pitch (Marwan et al., 2007); and b)

Teager–Kaiser energy operator of pitch (TKEO) (Tsanas et al., 2012). RQA is a general non-linear time-series analysis tool that quantifies multiple aspects of temporal stability of a time series, such as how repetitive, noisy, or stationary it is.

Relying on the time series in each laugh (e.g., a sequence of estimated pitch regularly sampled over time), RQA reconstructs the phase space of possible combinations of states and quantifies the structure of recurrence; that is, the number of instances in which the time series displays repeated dynamics, and the characteristics of these repetitions. To apply RQA, two steps are necessary: 1) reconstructing the phase space underlying the time series, and 2) production of a recurrence plot. The phase space of a time series is an n-dimensional space in which all possible states of a system are represented, so that it is possible to portray the trajectories of the system's behavior, be it periodic (repeatedly crossing the same regions at regular intervals), random, or chaotic. To reconstruct the phase space, we applied a time-delay method to each time series. After reconstructing the phase space, we constructed recurrence plots for each time series. Black dots on the plots represent every occasion at which a phase space trajectory goes through approximately the same region in the phase space. In mathematical terms, if we represent the trajectory of a system as

$$\{\vec{x_i}\}_{i=1}^{N}$$

the corresponding recurrence plot is based on the following recurrence matrix:

$$R_{i,j} = \begin{cases} 1: \vec{x_i} \approx \vec{x_j}, \\ 0: \vec{x_i} \not\approx \vec{x_j}, \end{cases} i,j = 1, \dots, N$$

where N is the number of considered states of the system and $\vec{x_i} \approx \vec{x_j}$ indicates that the two states are equal up to an error (or distance) ε. Note that this ε is essential in the case of continuous variables (as in F_0) as systems often do not recur exactly, but only approximately revisit states. To statistically analyze differences in laughs, we performed RQA on the recurrence plots. This makes it possible to statistically compare different dynamic systems (e.g., different dyads) in terms of such dynamics as the stability, structure, and complexity in the behavior of the system. Specifically, we analyzed:

Amount of repetition: The percentage of values that recur (are repeated) in the time series independently of the lag (recurrence rate, RR).

$$RR(\varepsilon) = \frac{1}{N^2} \Sigma_{i,j=1}^{N} R_{i,j}(\varepsilon)$$

Stability of repetition: articulated in:

Average length of sequences repeated (L)

$$L = \frac{\sum_{l=lmin}^{N} l\, P(l)}{\sum_{l=lmin}^{N} P(l)}$$

Length of longest repeated sequence (LMAX)

$$LMAX = \max(\{l_i\}_{i=1}^{N_l})$$

For more details about these indexes see Marwan et al. (2007).

TKEO has been widely employed to quantify jitter and shimmer; that is, perturbations in the regular cycles of pitch and intensity, respectively, which often characterize situations of stress and arousal, and are impacted by the ability to control the speech production system. TKEO is calculated as

$$\psi(x_n) = x_n^2 - x_{n+1} \cdot x_{n-1}$$

where the subscript n denotes the nth entry of the vector x (in our case, the time series of pitch). We computed the mean, standard deviation and 5th, 25th, and 95th percentile values of TKEO.

Overall, this resulted in 39 features for each laugh.

Table S4. 39 extracted features in acoustic analysis.

Voiced / Unvoiced segments	**Pitch**	**Intensity**	**Harmonics to Noise Ratio (HNR)**
Total duration of the laugh	Pitch Mean	Intensity Mean	HNR Mean
Voiced duration	Pitch SD	Intensity SD	HNR SD
Unvoiced duration			
Mean intervoicing interval (IVI)	Pitch Median	Intensity Median	HNR Median
	Pitch IQR	Intensity IQR	HNR IQR
	Pitch CV	Intensity CV	HNR CV
	Pitch Mean Absolute Deviation	Intensity Mean Absolute Deviation	HNR Mean Absolute Deviation
	Recurrence Rate (RR)	Recurrence Rate (RR)	
	Mean length of recurrent sequence (L)	Mean length of recurrent sequence (L)	
	Maximum length of recurrent sequence (LMAX)	Maximum length of recurrent sequence (LMAX)	
	Mean TKEO	Mean TKEO	
	SD of TKEO	SD of TKEO	
	5th percentile of TKEO	5th percentile of TKEO	
	25th percentile of TKEO	25th percentile of TKEO	
	95th percentile of TKEO	95th percentile of TKEO	

This dataset was used to assess which acoustic features i) best discriminate between spontaneous and volitional laughs, and ii) might be employed by listeners when judging whether a laugh was spontaneous ("real") or volitional ("fake"). We call this measure the *Spontaneity Ratio* (SR), defined as the overall likelihood of each laugh being judged as spontaneous.

To examine cross-cultural reliability, we then employed the selected features to predict within-cultures SR and assessed the amount of variance explained through Adjusted R^2. All acoustic features were linearly transformed on a scale from 0 to 1 for better performance in the feature selection process.

Analysis and machine learning process

Feature selection. The process described above produces a large set of features, exemplifying what is commonly termed the curse of dimensionality. In other words, the presence of a large number of features makes the statistical models both difficult to interpret and at risk of overfitting, producing results that are not generalizable. To address this, we used a common algorithm to select a parsimonious subset of features, the Elastic Net extension of the LASSO (Zou & Hastie, 2005). In principle, this step could reduce overall accuracy, but it increases the interpretability and generalizability of the results; that is, the ability to accurately describe new laughs with characteristics similar to the laughs in the current study.

Statistical models. To assess the overall model relying on the selected features, we used a 5-fold cross-validated logistic regression to discriminate spontaneous versus volitional laughs, and a 5-fold cross-validated linear regression model to reconstruct participants' likelihood of judging a given laugh to be spontaneous (SR). The dataset was divided into 5 subsets (or folds) each containing a non-overlapping sixth of the laughs. A combination of 4 folds was used for feature selection and model fitting. The model was then assessed on the remaining fold. This procedure was repeated for all four possible combinations of folds, hence the accuracy of the model was assessed only on data on which it had not been trained. We repeated the cross-validation process a total of 100 times, randomly permuting the data before splitting into training and testing subsets to ensure stability of the results across different random splits in 5 folds. Post-hoc testing was applied to estimate Betas and standard errors of the predictors in the regression models.

Samoan judgments and acoustic-based model prediction. There is a notable divergence between the use of acoustic features by our Samoan participants and the other study populations. The acoustic-based model explained none of the variance in their responses, and alternative models using other acoustic features also failed to explain variance. Importantly, the inability of the model to explain Samoan's judgments of laughs being spontaneous does not suggest they use these features in some opposing way, but rather that they were likely relying on some other feature(s) not captured by our acoustic analysis. Across all societies, Samoan overall accuracy was the lowest (56%), and their rate of judging laughs as volitional was highest (60%). The finding deserves specific follow-up in future research.

Table S5. Variance (measured as r squared) in the judgments of laughs being spontaneous (i.e., "real") for each sample of participants explained by the acoustic-based model.

Group	**Region**	R^2
Australia	Oceania	0.62
Samoa	Oceania	0
South Africa/Zulu	Africa	0.48
Singapore	Asia	0.60
Korea	Asia	0.63
Japan	Asia	0.63
India	Asia	0.56
Iran	Asia	0.70
China	Asia	0.59
Indonesia	Asia	0.64
Qatar	Asia	0.70
Netherlands	Europe	0.64
Slovakia	Europe	0.66
Turkey	Europe	0.67
Spain	Europe	0.65
Austria	Europe	0.60
Canada	N. America	0.59
USA	N. America	0.64
Ecuador (Shuar)	S. America	0.24
Peru (urban)	S. America	0.66
Peru (rural)	S. America	0.54

Signal detection analysis

We performed a signal detection analysis in the form of a multilevel probit regression (DeCarlo, 1998). The binomial response (judgment of "real" versus "fake") was predicted by intercept (equivalent to criterion) and condition (spontaneous versus volitional laugh; equivalent to sensitivity). Both parameters were also modelled as random effects; that is, varying by society and participant, and as potentially correlated. The intercept indicates a bias in the responses—in particular, lower negative bias values indicate a greater tendency to respond "fake." Note that multilevel models perform partial pooling of information, so estimates of each society are influenced by the data available for all countries. This might reduce differences between societies, but it also provides more conservative estimates, and has been shown to improve generalizability of the models (Gelman & Hill, 2007). See Table S6 for all values.

To draw ROC curves by society, we estimated the predictions of the model above and employed them to assess the effects of varying decision thresholds on the sensitivity and specificity of the model binomial predictions.

Table S6. Signal detection values across 21 societies.

Society	Hit	False alarm	Correct rejection	Sensitivity	Bias	AUC
Australia	0.67	0.36	0.64	0.79	-0.37	0.68
Austria	0.69	0.35	0.65	0.87	-0.38	0.70
Canada	0.65	0.44	0.56	0.65	-0.27	0.65
China	0.67	0.39	0.61	0.77	-0.34	0.68
Ecuador (Shuar)	0.50	0.37	0.63	0.41	-0.38	0.64
India	0.56	0.37	0.63	0.53	-0.36	0.63
Indonesia	0.63	0.35	0.65	0.73	-0.39	0.66
Iran	0.59	0.32	0.68	0.68	-0.42	0.67
Japan	0.71	0.34	0.66	0.95	-0.40	0.71
Korea	0.71	0.34	0.66	0.94	-0.40	0.72
Netherlands	0.69	0.33	0.67	0.90	-0.40	0.71
Peru (Rural)	0.53	0.31	0.69	0.56	-0.44	0.65
Peru (Urban)	0.67	0.36	0.64	0.80	-0.37	0.68
Qatar	0.63	0.35	0.65	0.71	-0.38	0.68
S. Africa (Zulu)	0.51	0.31	0.69	0.52	-0.47	0.63
Samoa	0.46	0.36	0.66	0.36	-0.42	0.58
Singapore	0.66	0.42	0.58	0.68	-0.29	0.65
Slovakia	0.66	0.38	0.62	0.74	-0.33	0.67
Spain	0.68	0.38	0.63	0.81	-0.35	0.68
Turkey	0.66	0.32	0.68	0.85	-0.42	0.70
USA	0.69	0.35	0.65	0.89	-0.39	0.70

III. GLMM comparisons

We used a model comparison approach, assessing model fit using the Akaike Information Criterion (AIC). This approach allows researchers to assess which combination of variables best fit the pattern of data without comparison to a null model. Model 4 below (bolded) had the best fit, and is reported in the main text. The fit was almost identical to Model 6, the only difference being that Model 6 includes a non-significant sex difference in performance.

Table S7. Model comparisons for accuracy in the judgment task.

Model	**Fixed factors**	**Random factors**	**Estimate**	**SE**	**z**	**Variance**	**SD**	**AIC**
M1	(Intercept)		0.6001	0.2270	2.643			33814.2
	Condition		0.2298	0.1946	1.181			
		Subject				0.06618	0.2573	
		Laugh Trial				1.52421	1.2346	
M2	(Intercept)		0.5992	0.2326	2.576			33693.1
	Condition		0.2333	0.1965	1.187			
		Subject				0.02764	0.1662	
		Society				0.03930	0.1982	
		Laugh Trial				1.52717	1.2358	
M3	(Intercept)		0.57416	0.22870	2.511			33814.2
	Condition		0.23002	0.19498	1.180			
	Sex		0.04579	0.03216	1.424			
		Subject				0.06567	0.2563	
		Laugh Trial				1.52426	1.2346	
M4	**(Intercept)**		**0.6252**	**0.2389**	**2.617**			**33416.6**
	Condition		**0.1908**	**0.2188**	**0.872**			
		Subject				**0.03005**	**0.1733**	
		Society x Condition				**0.08939**	**0.2990**	
		Laugh Trial				**1.53619**	**1.2394**	
M5	(Intercept)		0.57607	0.23341	2.468			33693.3
	Condition		0.23340	0.19663	1.187			
	Sex		0.04057	0.03036	1.336			
		Subject				0.02728	0.1652	
		Society				0.03918	0.1979	
		Laugh Trial				1.52720	1.2358	
M6	(Intercept)		0.60190	0.23909	2.517			33416.9
	Condition		0.19096	0.21849	0.874			
	Sex		0.04092	0.03072	1.332			
		Subject				0.02968	0.1723	
		Society x Condition				0.08929	0.2988	
		Laugh Trial				1.53620	1.2394	
M7	(Intercept)		0.52005	0.23473	2.216			33686.9
	Condition		0.33952	0.20000	1.698			
	Sex		0.12550	0.04207	2.983			
	Condition x Sex		-0.15839	0.05434	-2.915			
		Subject				0.02735	0.1654	
		Society				0.03921	0.1980	
		Laugh Trial				1.53234	1.2379	

IV. Full text of experimental instructions

The following text was used in the experiment. If the participants did not speak English, these instructions were translated into the language to be used at the study site (usually the native language of the participants). If participants were unable to read, the instructions were read aloud to the participant and their answers entered into the computer by the experimenter.

Full text of instructions

Welcome to the fake-or-real laugh study. In this experiment you will listen to recordings of women laughing. In some of the recordings, the women were asked to laugh, but were not given any other reason for laughing (we call these fake laughs). Other recordings are of women laughing naturally while talking to a friend (we call these real laughs). For each recording, your job is to decide whether it is fake laugh or a real laugh. Each recording is of a different woman.

Before we begin with the actual study, you will be able to practice with one recording so that you will be familiar with the procedure.

When you are ready, press the space bar to hear the practice recording.

Do you think this laugh is a fake laugh or a real laugh?

Press 0 if you think that the laugh is fake. Press 1 if you think that the laugh is real.

If you have any questions, please ask the experimenter. If not, press the Enter key and the experiment will begin.

When you are ready, press the space bar to hear the recording.

You have now listened to all of the recordings. Thank you for your participation. Please tell the experimenter that you are finished.

V. Laughter samples

1. Spontaneous laugh 1 (spontaneous1.wav)
2. Sponatenous laugh 2 (spontaneous2.wav)
3. Volitional laugh 1 (volitional1.wav)
4. Volitional laugh 2 (volitional2.wav)

References

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication, 71*, 10-49.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods, 3*(2), 186-205.

Degottex, G. (2014). COVAREP: *A Cooperative Voice Analysis Repository for Speech Technologies.*

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research, 10*(3), 384-407.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevelhierarchical models (Vol. 1)*. New York, NY, USA: Cambridge University Press.

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology, 4*, 513. http://doi.org/10.3389/fpsyg.2013.00513

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*(5), 237-329.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B, 58*(1), 267-288.

Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Transactions on Biomedical Engineering, 59*(5), 1264-1271.

Washington, C. D. A., Assis, F. M., Neto, B. G. A., Costa, S. C., & Vieira, V. J. D. (2012, January). *Pathological voice assessment by recurrence quantification analysis*. In Biosignals and Biorobotics Conference (BRC), 2012 ISSNIP (pp. 1-6). IEEE.

Zou, H., & T. Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B, 67,* 301-320.