

Pupil dilation reflects the authenticity of received nonverbal vocalizations

Gonçalo Cosme¹, MSc, Pedro J. Rosa^{2,3}, PhD, César F. Lima³, PhD, Vânia Tavares^{1,4}, MSc, Sophie Scott⁵, PhD, Sinead Chen⁶, PhD, Thomas D.W. Wilcockson^{7,8}, PhD, Trevor J. Crawford⁷, PhD, Diana Prata^{*1,3,9} PhD

1. Instituto de Biofísica e Engenharia Biomédica, Faculdade de Ciências da Universidade de Lisboa, Campo Grande – 1749-016 Lisboa, Portugal
2. HEI-LAB: Human-Environment Interaction Lab/Universidade Lusófona de Humanidades e Tecnologias, Campo Grande, 376, 1749 - 024 Lisboa, Portugal
3. Instituto Universitário de Lisboa (Iscte-IUL), CIS-Iscte, Avenida das Forças Armadas, 1649-026 Lisboa, Portugal
4. Faculdade de Medicina da Universidade de Lisboa, Avenida Professor Egas Moniz MB, 1649-028 Lisboa, Portugal
5. Institute of Cognitive Neuroscience, University College London, Alexandra House, 17-19 Queen Square, Bloomsbury, London, UK
6. Risk Society and Policy Research Center, National Taiwan University, Roosevelt Rd., Daan Dist., Taipei 10617, Taiwan
7. Department of Psychology, Lancaster University, Lancaster, LA1 4YF, UK
8. School of Sport, Exercise, and Health Sciences, Loughborough University, Clyde Williams Building, Epinal Way, Loughborough, Loughborough LE11 3GE, UK
9. Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 16 De Crespigny Park, Camberwell, London SE5 8AF, UK

Corresponding author: Dr Diana Prata, Instituto de Biofísica e Engenharia Biomédica, Faculdade de Ciências da Universidade de Lisboa, Campo Grande 016, 1749-016 Lisboa, Portugal; [+351 21 750 0177](tel:+351217500177); diana.prata@kcl.ac.uk.

Abstract

The ability to infer the authenticity of other's emotional expressions is a social cognitive process taking place in all human interactions. Although the neurocognitive correlates of authenticity recognition have been probed, its potential recruitment of the peripheral autonomic nervous system is not known. In this work, we asked participants to rate the authenticity of authentic and acted laughs and cries, while simultaneously recording their pupil size, taken as proxy of cognitive effort and arousal. We report, for the first time, that acted laughs elicited higher pupil dilation than authentic ones and, reversely, authentic cries elicited higher pupil dilation than acted ones. We tentatively suggest the lack of authenticity in others' laughs elicits increased pupil dilation through demanding higher cognitive effort; and that, reversely, authenticity in cries increases pupil dilation, through eliciting higher emotional arousal. We also show authentic vocalizations and laughs (i.e. main effects of authenticity and emotion) to be perceived as more authentic, arousing and contagious than acted vocalizations and cries, respectively. In conclusion, we show new evidence that the recognition of emotional authenticity can be manifested at the level of the autonomic nervous system in humans. Notwithstanding, given its novelty, further independent research is warranted to ascertain its psychological meaning.

Keywords

Empathy, Theory of Mind, mentalization, emotion recognition, authenticity recognition, nonverbal vocalizations, pupillometry, autonomic nervous system

Significance Statement

For the first time, we probed authenticity recognition in human vocalizations for its effect on pupil dilation, a psychophysiological index for mental effort and arousal. We show that authentic cries and acted laughs elicited higher pupil dilation compared to acted cries and authentic laughs, respectively. These unprecedented findings suggest the socially complex process of authenticity recognition in nonverbal vocalizations can be reflected in a peripheral autonomic nervous system response, and that this effect depends on the emotion underlying the expression.

Introduction

We express emotions in social interactions to convey information about our affective states and intentions, which is essential for communication. In turn, we constantly evaluate the authenticity behind others' emotional expressions, even involuntarily. This complex process of cognitive empathy (also known as theory of mind or mentalization), uses perceptual and sensorimotor cues¹, and allows for an adequate social response, from more intuitive to more deliberate, such as decisions to trust or not to trust in the other individual, and thus whether to cooperate or to compete². Given that these decisions are vital for social bonding, defense from aggression, and ultimately social network structure, they have been of utmost importance for human survival³. As such they may be hard-wired in our nervous system. However, it is still unknown whether and how the evaluation of the authenticity of another's emotional expression engages the autonomic nervous system of the person perceiving it.

Emotions can be effectively expressed vocally without semantic content, such as in laughter and crying, unconstrained by linguistic structures⁴⁻⁶. Even in the absence of a situational context, nonverbal vocalizations provide relevant cues to infer emotional states⁷, and their recognition can transcend cultures⁸. Nonverbal vocalizations can vary in emotional category (e.g., amusement, sadness, anger, fear, surprise and disgust)⁹, valence⁴, arousal⁴, affiliative value (i.e. emotional contagion)¹⁰ and authenticity¹. Crying, for example, is an intense emotional expression of a negative state often accompanied by lachrimation, which, in a social context, is assumed to have the purpose of eliciting help from listeners¹¹, or, in an interpersonal context, is understood to function as relief and improve mood after shed tears¹². Conversely, and vastly more studied in nonverbal vocalization literature, laughter is an emotional expression of a positive state and has the role of promoting and maintaining social bonding¹³. The underlying emotion causing the expression of crying or laughter may vary such that acoustic differences have been found for different kinds of laughter (e.g. ticklish versus emotional)^{14,15}, for example.

Correctly recognizing the authenticity of others' laughter or crying is a social skill essential for avoiding deception^{2,16}. Authenticity recognition is the ability of discerning between an authentic (genuine) versus an

acted (deliberate) emotional expression, for which we use acoustic differences in the case of nonverbal vocalizations^{15,17}. We have shown that authentic laughter and cries are often more highly pitched, longer in duration and have different spectral characteristics compared to their acted variants¹⁸. Additionally, across various emotions, higher and more variable pitch, lower harmonicity, and less regular temporal structure are the best predictors of authenticity judgements¹⁹. Listeners recognize authenticity in laughter at roughly 70% accuracy (67%¹⁷, and, as we have shown, 72%¹⁵ and 63%¹⁹). Furthermore, authentic laughter is rated as more arousing and more positive than acted laughter¹⁵. Generally, genuine emotional expressions are produced reactively while deliberate expressions are intentional and controlled forms of communication¹³. Whereas authentic laughter is genuine and usually an immediate reaction to a positive and surprising stimulus, acted laughter is associated with polite agreement and (real or fake) appreciation^{13,17}. While authentic crying is also genuine and usually negative stimulus-driven, acted crying is associated with (manipulative) social deception¹⁶.

We have shown, during passive listening, the anterior medial prefrontal cortex (amPFC) and anterior cingulate cortex to be more strongly engaged for acted (than authentic) laughter¹. This is consistent with the view that interpreting non-authentic stimuli, and solving its ambiguity, is relatively more cognitively demanding²⁰, engaging cognitive empathy (i.e. mentalization) to a higher degree. In addition, a linear decrease in amPFC activation as perceived authenticity increases¹⁸, has been found. To our knowledge, the electrophysiological response to nonverbal vocalization authenticity has not yet been investigated, but visual stimuli's authenticity (e.g. genuine vs. ambiguous smiles) has been reflected in early event related potentials (ERPs) components' amplitude, e.g. P200, albeit only when the ambiguous smiles were blended with angry eyes, suggesting it is processed very early, and dependent on the salience of the expression²¹. More evidence that cognitive strategies are required to infer authenticity and that these demand a level of social maturity and experience, comes from the findings that adults', but not children's, cognitive and emotional empathy scores correlated with authenticity discrimination of happy faces²². In addition, high *emotional* trait empathy may also help authenticity recognition by facilitating the simulation of the emotion itself (through emotional contagion, i.e. the propensity to resonate with others' emotions), leading to a stronger emotional response to authentic expressions²². Indeed, we have

reported that emotional trait empathy, emotional contagion and authenticity recognition in laughter to be positively associated ¹⁰.

Pupil size is used as a proxy of both arousal ²³ and cognitive effort in emotion research ^{24,25} and it depends on autonomic peripheral nervous system activity, which may in turn be elicited by central nervous system input. Activation of the iris dilator muscle ²⁶ stems from a sympathetic response (known in the context of the ‘fight-or-flight’ mode) triggered by adrenaline release, and produces an enlargement of the pupil size (i.e. pupil dilation). Activation of the iris sphincter muscle ²⁶ stems from a parasympathetic response (typical of the ‘rest and digest’ mode) ^{23,27}, and produces a reduction of the pupil size (i.e. pupil constriction). The pupil dilates with higher arousal elicited by a stimulus ²³, thus, emotionally charged vocalizations evoke higher pupil dilation compared to neutral ones (with differences between positive and negative showing mixed results ²⁸). In addition, the pupil also dilates with cognitive effort ^{29,30}, and has been associated with amPFC activity, a key area for cognitive empathy as abovementioned ³¹. Furthermore, pupil mimicry (i.e. synchronization) during social interaction is proposed to be an emotional contagion mechanism and an implicit form of social communication ³². Indeed, pupil size changes are elicited by emotional contagion ³² which in turn facilitates authenticity recognition at least for laughter ¹⁰.

Examining how sensitive is, if at all, pupil size to the authenticity of perceived emotional expressions can reveal underlying processes of authenticity discrimination; in particular, whether the autonomic system is involved. Furthermore, by combining this examination with empathy traits and behavioural measurements, one could disentangle to what extent authenticity discrimination is a cognitively demanding and/or affective process. However, the direct association between pupil size and the recognition of authenticity has not yet been examined, to our knowledge. Although authenticity recognition is an essential cognitive empathy skill for an adaptive social behavior, and its central nervous system correlates have started to be unraveled, it is still unknown whether it engages the peripheric autonomic nervous system. Herein, we asked for the first time, whether the authenticity of an emotional expression induced an autonomic nervous system response during its perception. In the present study, we tested this by assessing the effect of nonverbal vocalizations’ authenticity

on the pupil size of the listener. We asked participants to rate the degree of authenticity in authentic and acted laughs and cries, during which we recorded changes in their pupil size and, posteriorly, the degree of emotional contagion and arousal of the same stimuli. Lastly, we measured the participants' cognitive and emotional empathy traits. Given the unprecedented examination into the autonomic nervous system's activity during authentic discrimination and the dual-proxy nature of pupil response, we had two possible (directional) predictions alternative to the null hypothesis, one assuming a preponderance of arousal, the other of cognitive effort: (1) authentic vocalizations would elicit *higher* pupil dilation compared to acted, because they have been found to be more arousing in general^{19,33}, and pupil dilation increases with arousal²³; or (2) by the contrary, authentic vocalizations would elicit *lower* pupil dilation, because authenticity discrimination, at least in laughter, has been found to decrease the engagement of prefrontal cognitive empathy-relevant brain areas^{1,18}, suggesting lower cognitive demand. As such, we aimed to disentangle whether it is arousal (supposedly higher in authentic) or cognitive load (supposedly higher in acted vocalization), during authenticity recognition, that engages autonomic nervous system the most. Additionally, by including an examination of the neural correlates of crying for the first time, we explored how our two predictions would depend on emotion valence (i.e. would differ between laughs and cries); which is warranted given that the social meaning of authenticity in each emotion can be quite different (i.e. a fake laugh can signify benign politeness or sarcasm, but a fake cry can mean costly deceit). Complementarily, (1) given that authenticity discrimination has been positively correlated with cognitive²² and emotional empathy^{10,22}, we test the association of these empathy traits with our behavioral and pupil size measures; and (2) we also ask whether previous positive associations between perceived authenticity and arousal¹⁸ and, for the first time, emotional contagion^{1,10}, in laughter, extend to crying.

Materials and Methods

Participants

All methods were carried out in accordance with relevant guidelines and regulations. Thirty-eight individuals were recruited to participate in the experiment via the laboratory's recruitment website. Eight participants were excluded from analysis due to technical problems in data acquisition (i.e. no eye data recorded across session or task trigger misconfiguration), and 2 due to inadvertently uncontrolled room luminosity, thus the final sample

consisted of 28 participants (13 male and 15 female) with an average age of 23.0 years ($SD = 1.38$, ranging from 21 to 26 years old). The inclusion criteria were right handedness (Edinburgh Handedness Inventory; due to EEG measures being collected in the same study)³⁴ and European Portuguese as first language. For female participants, an additional inclusion criterion was to be on the active weeks of the contraceptive pill, as time of the menstrual cycle has been shown to affect emotion recognition task performance³⁵. Participants provided written informed consent and were paid for their participation. To ascertain a normal distribution in terms of working memory, emotional state, and a mentally healthy sample, the Positive and Negative Affect Schedule (PANAS)^{36,37} (Positive Affect Score – $M = 43.41$, $SD = 8.52$; Negative Affect Score – $M = 29.95$, $SD = 9.80$); the Forward and Backward Digit Span Tests of Working Memory Index of the Weschler Adult Intelligence Scale – Third Edition (WAIS – III)³⁸ (WM Index = 19.71, $SD = 3.66$); and the Brief Symptom Inventory was administered (Global Severity Index (GSI) – $M = 0.66$, $SD = 0.49$), respectively.

Stimuli

The set of auditory stimuli comprised authentic and acted nonverbal vocalizations of amusement – laughter – and of sadness – crying – along with neutral vocalizations (e.g. the vowel ‘ah’ uttered with neutral intonation)³⁹. The neutral vocalizations were included only for comparison against cries and laughs; there were no acted-neutral or authentic-neutral stimuli, as neutral stimuli are, by nature, not affective and thus cannot be authentic or acted. Authentic vocalizations were elicited by the speakers while watching humorous videos (authentic laughter) or while recalling truly upsetting events (authentic crying), whereas acted laughter and crying were acted under full voluntary control. We used vocalizations we previously validated at the behavioral and neuroimaging level^{1,40}, as follows in short. Three male and 3 female speakers recorded the stimuli in an anechoic chamber. For authentic laughter, YouTube videoclips which were previously identified by the speakers as humorous, were shown to induce them to laugh out loud. For authentic crying, speakers were encouraged to recall personal upsetting events and/or start by posing crying in order to transition genuine crying. Lastly, the speakers were asked to simulate acted laughter and crying without feeling any genuine amusement or sadness, respectively. To avoid carry-over effects of genuine amusement or sadness, the recording of acted laughter or crying always preceded the recording of authentic laughter or crying. From each recording session, separate

audio files were created for laughter and crying, sampled at 44.1 kHz to mono.wav files with 16-bit resolution. To control for variability in the acoustic properties of the sounds, the audio was normalized for root-mean-square (RMS) amplitude using Praat software (www.praat.org)⁴⁰. In this study, for each condition (authentic laughter, acted laughter, authentic crying, acted crying), 18 vocalizations were used and presented twice. An additional 60 neutral vocalizations were presented once. In the end, the stimuli set consisted of 132 vocalizations, each with different durations (in milliseconds - authentic laughs: $M = 2399.94$, $SD = 460.73$, range = 1536.00, 3141.00; acted laughs: $M = 2248.89$, $SD = 400.15$, range = 1710.00, 2903.00; authentic cries: $M = 2684.55$, $SD = 289.36$, range = 2079.00, 2993.00; acted cries: $M = 2322.11$, $SD = 351.48$, range = 1959.00, 2990.00; neutrals: $M = 2498.74$, $SD = 292.08$, range = 2057.00, 2930.00). The acoustic properties of the stimuli (duration (ms), mean fundamental frequency – $F(0)$, mean intensity (dB)) were obtained using Praat software and reported in more detail in supplemental material (**Supplementary Table S1**).

To compare the acoustic properties between conditions, the main effects of authenticity and emotion were tested using non-parametric Kruskal-Wallis H Tests (due to the normality assumption not being fulfilled for their mean in each condition) and reporting ϵ^2 for effect size for duration, intensity and fundamental frequency. Further, pairwise comparisons were performed. There was an association of authenticity with pitch ($H(1) = 16.53$, $p < .001$, $\epsilon^2 = 0.97$) and duration ($H(1) = 9.98$, $p = .002$, $\epsilon^2 = 0.59$), whereby authentic vocalizations had higher pitch and were longer than acted ones; but not with intensity ($H(1) = 0.28$, $p = .596$, $\epsilon^2 = 0.02$). Additionally, there was an association of emotion with pitch ($H(1) = 63.81$, $p < .001$, $\epsilon^2 = 3.75$), whereby negative vocalizations had higher pitch than positive and neutral, and positive more than neutral; and intensity ($H(1) = 62.59$, $p < .001$, $\epsilon^2 = 3.68$) where positive vocalizations had higher intensity than negative and neutral; but not with duration ($H(1) = 2.57$, $p = .227$, $\epsilon^2 = 0.15$).

Task

Before the start of the task, which included concomitant pupil size recording (described below), the participants were informed that they would listen to sounds, and that they would be required to rate the sounds in terms of their perceived authenticity. For neutral sounds, the participants were instructed to just attend to the stimulus.

Always showing a fixation cross on screen, a trial started with silence for 4000ms plus a jitter of 500ms, followed by the presentation of the sound stimuli and then a 3000ms interstimulus interval. After this, a 7-point Likert scale showed on screen for up to 5000ms for the participants to answer their perceived authenticity ranging from 1 (“Genuine” – authentic) to 7 (“Posed” – acted). For ease of interpretation and discussion, the scale was reversed for the statistical analysis. The task took 36 minutes to complete and had 204 trials in a pseudo-randomized and fixed sequence, balancing condition transitions trial-by-trial so that conditions transitioned equally between themselves as to minimize the effects of pupil habituation ⁴¹.

After the above pupillometry-recorded task round, participants (n=21 of the 28, due to unforeseen time limitation) were instructed to evaluate the perceived arousal and emotional contagion of the previously presented vocal stimuli in a 7-point Likert scale (Arousal: 1- Low arousal, 7- High arousal; Emotional contagion: 1- Not contagious at all; 7- Highly contagious), except for neutral sounds. Divided in two blocks, first they rated all 72 emotional sounds in terms of their arousal (72 sounds in total, 18 for each condition: authentic laughter, acted laughter, authentic crying, acted crying), and in the second block they rated the same sounds for their contagion. Herein, a trial consisted in 1500ms plus 500ms of jitter, then stimuli presentation followed by 1000ms of interstimulus interval, always with a fixation cross on the screen, after which the Likert scale of arousal or contagion (depending of the block) would be shown. This task lasted for 15 minutes and had 124 trials, also pseudo-randomized and fixed sequence with balanced condition transitions. Herein, each stimulus was presented once in each block.

Pupil size recording

The fixation cross and Likert scales were shown in a Lenovo 23.8-inch screen with 1920x1080 resolution and 60 Hz refresh rate. Gaze tracking and pupil measurements were recorded using the SR Research EyeLink 1000 Plus eye tracker. A chin rest was used to minimize head movement and keep a fixed distance to both the screen and camera, at approximately 56cm for all participants. Raw data was collected monocularly at 1000Hz with average accuracy of 0.15 visual angle.

After data collection, the pupil size was down sampled to 250Hz (to save on computational costs), blinks and datapoints 100ms before and after blinks were considered as missing data. A low-pass filter 4Hz cut-off frequency was applied to the signal. Pre-trial baseline was obtained for each trial as the median pupil size immediately before stimuli onset, in an interval that was 2% of the whole trial, which varied depending on the duration of the stimuli ($M = 204.14$, $SD = 4.71$ milliseconds). This median value was then subtracted across all datapoints of its trial as advised in the literature⁴². Finally, if the missing data did not exceed 600ms, as blinks longer than this are considered microsleeps⁴³⁻⁴⁵, the signal was linearly interpolated^{46,47}. These preprocessing steps were employed for all pupil dilation analyses. Four time windows of 1 second, after stimuli onset, were created to evaluate pupil size measures across time. As our stimuli had variable duration, a 4-second analysis period ensures the inclusion of peak dilation and consequent return to baseline, so that both peak and mean pupil diameter can be adequately measured. The segmentation into time windows allowed a more sensitive and thorough assessment of pupil response to authenticity and emotion, and its consequent constriction during and post stimuli presentation, and has been precedently employed^{48,49}. For each trial, maximum and mean pupil sizes were extracted in each individual time window. Lastly, as recommended in guidelines for pupillometry pre-processing and analysis⁵⁰, and common^{31,51,52}, we have excluded from the group analysis maximum and mean outlier pupil size datapoints at the level of the trial. Datapoints were considered outliers if their mean per condition was above or below 1.5 times the interquartile range (following standard criteria⁵³).

Procedure

Each participant underwent the experiment in one session lasting 2 hours and a half, sitting comfortably in a quiet room at the Centre for Clinical Research (Centro de Investigação Clínica) of the Medical Academic Centre of Lisbon (Centro Académico Médico de Lisboa), whose Ethical Committee approved all experimental protocols. During the task, the auditory stimuli was presented binaurally through a set of Senheiser CX 3.00 ear-canal phones at a comfortable listening level that was individually adjusted at the start of the experiment. The experiment was developed using MATLAB version 8.3.0 (R2014a) with Psychtoolbox 3⁵⁴. Participants were encouraged to respond as intuitively as possible. Buttons of the response pad were marked with the Likert

scale points to minimize memory demands. To facilitate the response, participants were asked to put three fingers of their left hand in response keys 1, 2 and 3 and four fingers in the remaining response keys. Three pauses of 30 seconds were distributed equally along the experiment to minimize fatigue. Concomitant electroencephalography recording also took place (data not yet analyzed). After the pupillometry-recorded task, participants rated their perceived arousal and emotional contagion for every sound, and finally responded to the Empathy Quotient (EQ) and the Reading the Mind in the Eyes Test (RMET), which assesses emotional and cognitive trait empathy.

Statistical Analysis

Correlation analyses (between trial-by-trial stimuli and pupillometry measures; between trait empathy scores and pupillometry measures, with authenticity as a categorical moderator; and between trait empathy scores and authenticity discrimination index – described below) were performed in R software 3.6⁵⁵, and when applicable, using the *rmcorr* package⁵⁶. To verify reliability of the rating scales, Cronbach's alpha was calculated for each one⁵⁷. To infer the main effects of authenticity and its interaction with emotion on each behavioral measure (authenticity, arousal and emotional contagion ratings), a repeated measure Analysis Of Variance (rpANOVA) model for each measure was conducted in SPSS (IBM Corp. Released 2017, IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.). For completeness, the association of authenticity, arousal and contagion ratings with the two pupil size measures was also estimated. For authenticity discrimination, ratings between 1 and 3 were converted to 'posed', and ratings between 5 and 7 were converted to 'authentic'. For completeness and to attempt a replication of our previous work (Neves et al., 2018), we also computed an index for authenticity detection ability, for each emotion and each subject, by subtracting the average authenticity ratings of acted stimuli from the average authenticity ratings of authentic ones; and we report, in supplemental material (**Supplementary Table S2**), the correlation between these indexes and individual trait empathy (EQ and RMET questionnaire) scores.

Pupillometry-wise, two separate rmMANOVA models were constructed in SPSS, each including Time window as a within-subject factor (0–1, 1–2, 2–3, 3–4 seconds): one to estimate the main effect of the within-subject

factor Authenticity (authentic, acted) and its interaction with the within-subject factor Emotion (laughter, crying); and another to estimate the main effect of emotion (laughter, neutral, crying) given the inexistence of authentic neutral and acted neutral sounds. The dependent variables of both rmMANOVAs were the two pupil size measures (maximum and mean pupil size) for which we report the two corresponding univariate rmANOVA results. All effects of interest were followed up by post hoc pairwise comparisons and reported after False Discovery Rate (FDR) correction in R⁵⁵ (and considered statistically significant when FDR corrected $p < .05$). Partial Eta-square (η_p^2) is reported as a measure of effect size. We report the main effect of authenticity and emotion as well as interactions between them and with time window. Little's Missing Completely At Random (MCAR)⁵⁸ tests were performed for all pupil size measures to test the randomness of missing values. Given our previous work suggesting the acoustic properties of the stimuli mediate authenticity recognition¹⁹ and since they naturally differ depending on their authenticity and/or emotion, complementary repeated measures correlation analyses between each stimuli (trial-by-trial) and pupil size measures were conducted to evaluate their direct effect on pupil dilation and hence their potentially mediating role⁵⁹ (see Results as supplementary material (**Supplementary Table S5**)).

Power

To our knowledge, only two studies^{60,61}, using affective auditory stimuli, have reported effects sizes of the effect of emotion on pupil size, whilst none have reported effects of authenticity. One⁶⁰ used 26 participants to report a main effect of emotion (positive, negative and neutral) on the mean gradient of pupil diameter, across 0-2 seconds after stimuli onset, with a size of $\eta_p^2 = .20$, while the other⁶¹ used 97 subjects to report a main effect of pleasantness (pleasant, unpleasant and neutral) on pupil diameter with a similar effect size of $\eta_p^2 = 0.22$ for the peak time window (2 - 4 seconds). We note these results were not independent as the latter study has a 33% stimuli overlap with the former study. An a priori power analysis, considering $\eta_p^2 = .20$, in GPower 3.1.9.4.⁶², pointed to a need of 22 subjects to achieve 80% power, at a 5% alpha in a repeated measures ANOVA, to detect effects of emotion (3 categories). We further aimed our sample at 38 participants (and used a final $N = 28$), following the literature's high standard for affective research on pupil size^{27,28,60,63-65}. A sensitivity analysis

indicates our sample (N = 28) could detect (80% power, 5% alpha): the main effect of emotion on pupil dilation at a minimum effect size of $\eta_p^2 = .16$.

Results

Behavioral analysis

1. Authenticity Rating

Interparticipant reliability was high for the authenticity rating (Cronbach's $\alpha = 0.92$). Expectedly, authentic vocalizations (M = 3.39, SD = 0.23) were perceived as more authentic than acted (M = 4.46, SD = 0.20) ones [$F(1, 27) = 43.84, p < .001, \eta_p^2 = .62$]. The effect of emotion was smaller but also significant, whereby laughs (M = 3.51, SD = 0.20) were reported as more authentic than cries (M = 4.36, SD = 0.21) [$F(1, 27) = 16.61, p < .001, \eta_p^2 = .38$]. No significant interaction was found ($p = .670$).

Participants recognized the authenticity of laughs and cries at a level that statistically significantly exceeded chance [laughter: $\chi^2(1) = 145.25, p < .001$; cry: $\chi^2(1) = 152.21, p < .001$], with accuracies of 63.4% and 69.3% for laughter and crying, respectively. Authenticity discrimination indexes were tested for correlation with trait empathy scores, as reported in supplemental material (**Supplementary Table S2**), with no statistically significant association found.

2. Arousal Rating

Interparticipant reliability was high for the arousal rating ($\alpha = 0.88$). Authentic vocalizations (M = 4.10, SD = 0.96) were perceived as more arousing than acted (M = 3.03, SD = 0.80) ones [$F(1, 22) = 55.82, p < .001, \eta_p^2 = .72$]; and laughs (M = 4.13, SD = 0.90) were perceived as more arousing than cries (M = 3.00, SD = 0.95) [$F(1, 22) = 38.22, p < .001, \eta_p^2 = .64$]. There was also a significant authenticity by emotion interaction on arousal ratings [$F(1, 22) = 15.05, p = .001, \eta_p^2 = .43$] (whereby the difference within laughs [$t(22) = 6.56, p < .001, \eta_p^2 = .66$], was slightly larger than the difference between cries [$t(22) = 6.18, p < .001, \eta_p^2 = .63$]).

3. Emotional Contagion Rating

Interparticipant reliability was also high for emotional contagion rating ($\alpha = 0.86$). Authentic vocalizations ($M = 4.38$, $SD = 0.82$) were perceived as more contagious than acted ($M = 3.38$, $SD = 0.75$) ones [$F(1,25) = 70.29$, $p < .001$, $\eta_p^2 = .74$]; and laughs ($M = 4.21$, $SD = 0.82$) were perceived as more contagious than cries ($M = 3.50$, $SD = 0.86$) [$F(1,25) = 21.70$, $p < .001$, $\eta_p^2 = .47$]. No authenticity by emotion interaction effect was found ($p > .05$).

4. Correlations between ratings

Repeated measures correlation analysis indicated that the authenticity rating was positively associated with the arousal [$r_{rm}(1603) = .42$, $p < .001$], and the contagion ratings [$r_{rm}(1460) = .31$, $p < .001$]; as well as arousal and contagion ratings between them [$r_{rm}(1260) = .40$, $p < .001$].

Pupil size analysis

1. Overview

MCAR tests indicated that missing values were random for the maximum and mean pupil sizes for all time windows ($p > .999$ for all tests). **Figure 1** illustrates the pupil diameter for each condition averaged across all participants. As the main effect of authenticity on pupil dilation measures was not statistically significant ($p < .05$), and we verified a statistically significant cross-over authenticity by emotion interaction pattern, we will only report and discuss the latter and not the former.

2. Maximum pupil size

The effect of authenticity on maximum pupil size was dependent on emotion (i.e. reflected in a statistically significant 'authenticity x emotion' interaction) [$F(1, 27) = 10.60$, FDR corrected $p = .003$, $\eta_p^2 = .28$] and, to a lower degree, on time window [$F(2.32, 62.51) = 4.56$, FDR corrected $p = .010$, $\eta_p^2 = .15$]. Pairwise comparisons for the 'emotion x authenticity' interaction show that acted laughs ($M = 216.22$, $SD = 12.24$) elicited significantly (FDR corrected $p = .028$) higher pupil dilation than authentic ones ($M = 192.88$, $SD = 10.88$), and authentic cries ($M = 200.91$, $SD = 9.92$) elicited significantly (FDR corrected $p = .013$) higher pupil dilation than acted ones ($M = 175.87$, $SD = 11.79$), across time windows. This interaction did not significantly differ

between time windows, as indicated by a non-statistically significant 3-way interaction [$F(1.64, 44.12) = 1.59$, FDR corrected $p = .332$, $\eta_p^2 = .05$]. Nevertheless, for the purpose of aiding future studies in selecting time windows for pupil size analysis, we report the ‘authenticity x emotion’ interaction results in each time window – which were all statistically significant (FDR corrected $p < .05$) – in the **Table 1** and in **Figure 2**.

Tested in a separate ANOVA including neutral vocalizations, the main effect of emotion on maximum pupil size reached statistical significance [$F(1.56, 42.00) = 21.11$, FDR corrected $p < .001$, $\eta_p^2 = .44$] as did its interaction with time window [$F(2.45, 66.11) = 6.11$, FDR corrected $p = .001$, $\eta_p^2 = .19$]. Pairwise comparisons show laughs ($M = 210.77$, $SD = 11.62$) elicited significantly (FDR corrected $p = .015$) higher maximum pupil dilation than cries ($M = 191.98$, $SD = 13.69$), and each higher than neutral ($M = 149.26$, $SD = 13.90$) (FDR corrected $p < .001$ and $p = .002$, respectively). In each time window, these comparisons hold statistical significance except in 0 – 1 seconds, between cries and neutral, and in the 2 – 3 seconds, between laughs and cries, as shown in **Figure 2**.

3. Mean pupil size

The effect of authenticity on mean pupil size significantly depended on emotion [$F(1, 27) = 11.15$, FDR corrected $p = .003$, $\eta_p^2 = .29$] and on time window [$F(1.80, 48.54) = 7.12$, FDR corrected $p = .005$, $\eta_p^2 = .21$]. As on maximum pupil size, and across time windows, authentic cries ($M = 109.85$, $SD = 12.61$) elicited significantly (FDR corrected $p = .013$) higher mean pupil dilation than acted ones ($M = 85.13$, $SD = 9.95$), but unlike for maximum dilation, the difference between acted laughs ($M = 122.68$, $SD = 9.32$) and authentic ones ($M = 108.60$, $SD = 9.92$) did not reach significance (FDR corrected $p = .114$). As on maximum pupil size, this interaction did not differ between time windows, as suggested by a non-statistically significant 3-way interaction [$F(1.76, 47.60) = 1.11$, FDR corrected $p = .332$, $\eta_p^2 = .04$]. Nevertheless, we report the ‘authenticity x emotion’ interaction results in each time window – which only reached significance (FDR corrected $p < .05$) in time windows 1-2 seconds and 3-4 seconds - in **Table 1** and in **Figure 3**.

Tested in a separate ANOVA including neutral vocalizations, the main effect of emotion on mean pupil size reached statistical significance [$F(1.48, 40.02) = 22.63$, FDR corrected $p < .001$, $\eta_p^2 = .46$] as did its interaction

with time window [$F(2.45, 66.21) = 9.28$, FDR corrected $p < .001$, $\eta_p^2 = .26$]. Like for maximum pupil size, pairwise comparisons show laughs ($M = 117.56$, $SD = 9.53$) elicited significantly (FDR corrected $p = .006$) higher mean pupil dilation compared to cries ($M = 99.32$, $SD = 11.00$), and each compared to neutral ($M = 57.69$, $SD = 10.41$) (FDR corrected $p < .001$ and $p = .001$, respectively). Also as for maximum pupil size, in each time window, the comparisons hold statistical significance except in 0 – 1 seconds between cries and neutral, and in 2 - 3 seconds between laughs and cries, as shown in **Figure 3**.

4. Association of pupil size with trait empathy scores

There was a statistically significant moderate negative association between mean pupil size and the EQ Cognitive Empathy trait score in laughter [$r(26) = -.42$, $p = .028$]; see **Table 2**, but not cries (collapsing all time windows; with no statistical significance reached when considering separate windows). Authenticity was not found to significantly moderate the association between trait scores and pupil size measures in laughs and cries, irrespective of time-window (uncorrected $p > .05$) (Supplementary Table S3), nor within each time window (data not shown).

5. Behavioral and pupil size analysis

None of the behavioral measures (authenticity, arousal, emotional contagion ratings) were significantly correlated ($p > .05$) with pupil size measures (within all time windows collapsed) which is reported as supplemental material (**Supplementary Table S4**). [This is in replication of an independent analysis in an independent and equally sized UK sample (data not shown or published)].

Discussion

We asked whether the evaluation of the authenticity behind others emotional expressions might engage the autonomic nervous system. We did this by investigating whether the level of authenticity of an emotional expression affects the autonomic nervous system of the person during perception. In particular, we estimated the effect of nonverbal vocalizations' authenticity on the pupil size of the listener. Additionally, we tested whether pupil size response depends on the valence of the emotion being vocalized - which we found to be the

case. We show that a listener's pupil size is affected by a laughter's and cry's authenticity such that acted laughs induce more pupil dilation than authentic ones, and, reversely, authentic cries induce more pupil dilation than acted ones. These findings are not confounded by intrinsic natural differences in the acoustic properties of authentic and acted vocalizations. Additionally, we show that laughter was perceived as more authentic, arousing and emotionally contagious than crying.

In detail, the interaction between authenticity and emotion explained almost one third of the variance in pupil dilation (i.e. 28% and 29% for maximum and mean pupil dilation, respectively), left unexplained by the other modelled main effects and interactions, including time window. Broken down per emotion, laughter authenticity negatively explained a large portion (up to 28%) of the variance unexplained otherwise, with the peak effect at 0-1 seconds for maximum pupil dilation, and at 1-2 seconds for mean pupil dilation. In crying, authenticity positively explained (i.e. in the opposite direction to laughs), an equally large portion (up to 28%) of the otherwise unexplained variance with the peak effect at 1-2 seconds for maximum pupil dilation, and at 3-4 seconds for mean pupil dilation. Even though this 'authenticity x emotion' interaction was present in all time windows (at least for maximum pupil size), it was most statistically significant in the 1-2 seconds time window (for both pupil measures, FDR corrected $p = .002 - .005$). In this time window, moreover, the simple authenticity effects peaked for *both* laughter and crying (and in both pupil measurements) and were all statistically significant (**Table 1**).

Aiming to disentangle whether it is cognitive effort (supposedly higher in acted vocalizations) or arousal (supposedly higher in authentic vocalizations) that engages the autonomic nervous system the most during authenticity recognition, since both have been associated with pupil dilation^{27,66}, our results seem to support the former prediction for laughter and the latter for crying. The cognitive effort interpretation of the laughter finding is consistent with: (1) acted (vs. authentic) laughs (and their subjective discrimination) having been found to increase engagement of prefrontal cognitive empathy-relevant brain areas^{1,18}, suggesting higher cognitive demand; (2) in the present study, only in laughter, pupil size was negatively correlated with the cognitive – and not the emotional - empathy score, and (3) authenticity discrimination in facial stimuli increasing with cognitive

empathy trait scores ²² (albeit, in vocal stimuli, we have only found it to increase with emotional empathy, previously ¹⁰). The arousal interpretation of the crying finding is supported by: (1) authentic vocalizations having been found to be more arousing ^{18,19}, which we also replicated in the current work albeit also for laughter; (2) authentic cries being rated as much more arousing than acted ones in the current study, albeit also for laughter; and (3) crying eliciting higher amygdala activation than laughter, a highly replicated finding ^{67,68}, and amygdala activation being robustly and positively associated with arousal ^{69,70}. Thus, a possible explanation to the large and opposite effect of authenticity in laughs and cries may be that: the heightened sympathetic autonomic nervous system response we detected for acted laughs (vs. authentic) and authentic cries (vs. acted) may be due to them eliciting higher cognitive load and arousal, respectively.

Indeed solving ambiguity is cognitively demanding ⁷¹. As indexed by pupil dilation, there is evidence that more effort is required to solve the cognitive conflict caused by auditory incongruent stimuli, compared to congruent ⁷². Thus, the inherently awkward and incongruent acted vocalizations are suggestively more cognitively demanding than authentic. Moreover, previous studies show that higher cognitive demand and low confidence in emotion recognition lowers the perceiver's discrimination ability and leads to increased pupil response ³⁰. We posit that the reason why we found an opposite effect for laughs and cries, may be that the discriminating authenticity in laughs depends relatively more on cognitive effort (than on emotional arousal), whilst in cries, the discrimination of authenticity may depend more on the level of emotional arousal they elicit. The hypothesis that discriminating authenticity in laughs is more cognitively demanding than in cries is consistent with it recruiting higher-order prefrontal cortical brain areas (as we have shown ^{1,18}), whilst a lower-order activation centered in the amygdala is typically the brain response to crying ^{67,68} (although neuroimaging inspection of authenticity recognition in cries has not yet been reported). This line of thought is also consistent with the degree of 'malignancy' of a fake laugh and of a fake cry in social interactions. While a fake laugh is considered a more recent cultural tool ¹⁷ to communicate polite appreciation or sarcasm, fake cries are thought to have a manipulative role. In fact, pretending criers are deemed more manipulative, less reliable, warm and competent ⁷³. Believing in fake cries (and then spending resources altruistically) can be costly to the person being deceived. The costlier the expenditure, the more hard-wired (and evolutionarily older) the relevant stimulus processing

may be in sub-cortical brain structures; and as such the recognition of authenticity in a cry may more plausibly depend on an more lower-order amygdala-mediated emotional arousal response, more than a higher-order prefrontal-cognitive one. Indeed, crying is a biological siren and is extremely arousing for listeners as it is one of the most primitive and early behaviors we have ⁷⁴. Unlike laughter, it is the first newborn's form of communication and arguably the most essential for their survival. This early biological underpin to cries is perhaps sufficient to facilitate our faster and more immediate authenticity discrimination in them (compared to laughter) – we may be ‘programmed’ to act urgently upon cries which are authentic, especially of newborns and children, and simply feel desensitized/unconvinced by acted cries, by perceiving them as much less arousing. Indeed, cries are arguably harder to fake and thus acted cries may be easier to spot (compared to laughs; 63% vs. 69% discrimination index accuracy in the present data), making them less cognitive demanding. Last but not least, in the present data, the acted crying was the condition rated as the least arousing to the subjects, and eliciting the lowest pupil dilation (**Figure 1**), en par with neutral vocalizations which had obligatorily a minimal contribution from cognitive load due to the authenticity rating not being asked in those trials. In sum, pupil dilation displays a cumulative contribution from both arousal and cognitive effort ³⁰ and herein, the pupil dilated more in laughs than it did in cries (**Figure 1**), suggesting that this additive effect is indeed less predominant in the latter, showcasing the presumed lower need of cognitive effort in cries' recognition.

Regarding empathy traits, cognitive empathy is usually referred to as the ability to take the perspective of another and to understand another's feelings or internal state, while not necessarily having an emotional response ⁷⁵. Cognitive empathy is related to social awareness and demands more cognitive effort than the more authentic emotional empathic responses, in line with its higher recruitment of the amPFC ⁷⁶. Such higher-order mentalizing computation is suggestively performed in order to distinguish between authentic and acted expressions ^{1,18}. Thus, in our study and, noticeably only in the laugh condition, mean pupil size was negatively correlated with the cognitive empathy score, and not correlated with emotional empathy (**Table 2**). This is consistent with the above deduction that non-authentic laughs triggered higher pupil dilation (vs. authentic) due to them being more cognitively difficult to decipher. If pupil size depends on cognitive load (and recruits prefrontal and cognitive empathic-relevant brain regions ^{31,32}, it follows that participants with superior cognitive

empathy would recognize authenticity more easily, and thus show lower pupil dilation, compared to those with lower cognitive empathy scores. Finally, our reported main effect of emotion valence on pupil dilation, whereby the pupil dilated more for laughs compared to cries and each more compared to neutral, adds a piece to the puzzle to a literature context where it is currently unclear if positive or negative sounds elicit higher pupil dilation, as recently reviewed ²⁸. The inconsistency thus far is possibly due to the heterogeneity of the stimuli libraries being used, which could vary in volume and acoustic properties.

Regarding our behavioral findings, as expected, and consistent with previous validation of the stimuli ^{10,77}, participants perceived authentic vocalizations as being more authentic than acted ones, and such discrimination was statistically significant above chance level, at an 66% average in line with previous studies (ranging from 65-72% ^{15,17,19}). This indicates that our participants could correctly perceive the authenticity of the stimuli and that they were engaged in the task. Regarding the authenticity discrimination index association with emotional empathy trait scores, likely due to low statistical power, as we found the same effect direction, we could not corroborate our previous findings (with a larger sample of 119) ¹⁰.

In respect to the effect of authenticity in all three ratings (perceived authenticity, arousal and emotional contagion ratings), it was statistically significant and these were all positively correlated with each other. As authentic vocalizations may be produced in a more spontaneous fashion (vs. acted ones), free of intention and voluntary control over the voice, they provoke higher arousal perception ^{15,19}. The positive correlation between ratings of authenticity and arousal, suggest that arousal perception might be involved in authenticity discrimination ¹⁹, however conflicting associations have been reported ^{15,77}. Participants also rated authentic vocalizations as more emotionally contagious than acted ones, replicating our previous results ¹⁰, and asserting the plausibility of mimicry/synchronization behaviors (e.g. body gestures, facial expressions), which are associated with emotional contagion ³¹, to occur preferably when the receiver perceives emotions as authentic.

Although negative emotions have been used in authenticity perception studies ¹⁹, the main effect of emotion (cries and laughs) on perceived authenticity, arousal and emotional contagion ratings is reported here for the

first time. We report that laughs were judged significantly more authentic, arousing and contagious than cries. These differences in the perception of laughs and cries may be due to the social use of both, at least nowadays. Laughter is usually used as a “social glue” to foster agreement and cooperation and is a common form of communication between people ¹³, whilst cries are usually expressed to much smaller, rarer and intimate audiences. As such, laughs may have been found more acceptable and less awkward when listened repeatedly and without context, as in such a controlled environment setting, whereas cries may have sounded more unpleasant, stranger, and thus less contagious, authentic or arousing. Our results are in line with our previous one of laughter showing to be more arousing than crying (although not statistically tested then) ⁷⁷, albeit another study found no significant difference ²⁷. Finally, the highly contagious effect of laughter which we found is extensively reported in the literature ^{10,13,78,79}.

Potential limitations

The MCAR test ⁵⁸ was used to validate pupil size recordings ⁸⁰, and in our study, the results indicate that missing datapoints due to blinks and other recording artifacts were completely random. The analysis of the effects of authenticity and emotion on pupil size measures was divided in time windows to better characterize the effects observed, however, we do not discuss the latency of such effects as pupil size has a variable response latency ⁴² and the stimuli used are continuous.

As for our complementary analyses, the fact that we did not find associations between pupil size measures and the ratings of authenticity, arousal and emotional contagion to be statistically significant is possibly because: (1) except authenticity, these were performed post-hoc after the pupillometry recording, (2) arousal was not modelled as a task condition (like all previous studies finding its association with pupil size ^{23,27,28}, and (3) the sample may have been insufficiently sized to examine individual behavioural differences such as arousal and contagion ratings (as it was designed to be powered to detect the pupil dilation response to the task ^{23,27,28,30,81}).

We acknowledge there are two additional tests that would have been useful to further support (or not) our suggestion that acted laughs may entail higher cognitive effort than authentic ones which we posed as an

explanation of our present findings. First, we could have tested whether the pupil dilates more during incorrect versus correct trials for laughter, which, if so, would have supported our latter suggestion. However, the ratio of correct/incorrect trials laughs was quite unbalanced between authentic and acted laughs, which prevented a reliable statistical inference. [Specifically, for authentic laughs, participants had the double of correct trials ($M = 18.61$, $SD = 8.62$) vs. incorrect trials ($M = 9.71$, $SD = 7.77$). For acted laughs it is the reverse – on average less correct trials ($M = 12.52$, $SD = 7.59$) than incorrect ($M = 15.43$, $SD = .30$). Thus, considering incorrectly rated authentic laughs, multiple participants have less than 5 trials, whereas in correct authentic laughs multiple have close to 36 (the maximum).] This could however be achieved with an adapted paradigm design suitable for such question, in a future study. Second, given the nature of the authenticity discrimination index which is a score of the participant's authenticity detection abilities for a specific emotion (calculated as described in Methods), it was also not possible to separately calculate and then compare the discrimination index for acted laughs and for authentic laughs. In a future study, an alternative index that would serve to compare the performance of authentic laughs vs acted laughs might be useful to further validate the suggested increase in cognitive effort stemming from an acted laugh.

We are aware that the exposure times to the auditory stimuli were rather short, and there was no visual information. Consequently, it was not possible to detect nonverbal emotional leakage, which made our task of distinguishing between authentic and acted nonverbal vocalizations more difficult than it is in real life. Furthermore, as pupil is a sensitive autonomic index, the recording of other autonomic indexes such as cardiac activity (e.g., Heart Rate Variability) could contribute to disentangle the influence of cognitive and affective processes in pupillary activity⁸².

The stimuli set used has different acoustic properties across conditions as detailed in supplemental material (**Supplementary Table S1**). For example, in the crying condition, authentic stimuli are longer than acted, and it is arguable that participants' ratings of authenticity, arousal and contagion were influenced by their difference. When building the stimuli set, we selected excerpts from actors' recordings (that became the stimuli used here) such that their acoustic properties were balanced as much as possible, but ultimately, we did not to control for

them because we needed to preserve the natural characteristics that may make up the authenticity of a vocalization. However, we report in supplemental material that the effects of authenticity and emotion on pupil size measures we report are not attributed to differences in acoustic properties across conditions.

Conclusions

In this work we asked if the process of authenticity recognition in nonverbal emotional cues induces an autonomic nervous system response in the listener. To do so we measured the pupil dilation of participants while exposing them to authentic and acted laughs and cries, in a task that required them to rate the authenticity of the stimuli. We report that acted laughs elicited higher pupil dilation than authentic, putatively through demanding higher cognitive effort; and that authentic cries elicited higher pupil dilation than acted ones, putatively through eliciting higher emotional arousal – in what is the first demonstration of a reflection of authenticity recognition in the autonomic sympathetic system. We also observed that authentic sounds were rated as more authentic, arousing, and contagious than acted ones, and that authenticity discrimination increases with cognitive trait empathy. Together, these findings seem consistent with available neuroimaging, psychological, cultural, and sociological features of laughter and crying. However, given their novelty, further independent examinations of the effect of others' non-verbal vocalizations authenticity on pupil size response is warranted to validate our interpretations.

Declaration of Conflicting Interests

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

Acknowledgements

DP was supported, during this work, by the European Commission Seventh Framework Programme Marie Curie Career Integration Grant FP7-PEOPLE-2013-CIG-631952, the 2016 Bial Foundation Psychophysiology Grant

- Ref. 292/16, the Fundação para a Ciência e Tecnologia FCT IF/00787/2014, LISBOA-01-0145-FEDER-030907 and DSAIPA/DS/0065/2018 grants, and the iMM Lisboa Director's Fund Breakthrough Idea Grant 2016; and is co-founder and shareholder of the neuroimaging research services company NeuroPsyAI, Ltd. VT was supported by an FCT PhD fellowship (PD/BD/114460/2016). GC was supported by an FCT PhD fellowship (SFRH/BD/148088/2019). TJC and TW were support by the UK EPSRC grant EP/M006255/1. During the preparation of this manuscript, CL was supported by an FCT Investigator Grant (IF/00172/2015). The work was also supported by SynaNet Twinning Project which has received funding from the European Commission's Horizon 2020 research and innovation programme under grant agreement No 692340. None of the authors declare a conflict of interest.

Contributions

GC has set up the eye-tracking experimental procedure, performed data processing and analysis, and drafted all the manuscript components. PR co-supervised the eye-tracking data processing and analysis. CL, SS and SC designed and built the task paradigm. VT, TW and TC contributed to the eye-tracking experimental procedure set up. DP supervised all stages of the work and majorly revised the manuscript. All authors contributed to the manuscript's revision and agree with its final version.

Data Availability

The datasets generated and/or analysed during the current study are available from the corresponding author on request.

References

1. McGettigan, C. *et al.* Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cereb. Cortex* **25**, 246–257 (2015).
2. Gervais, M. & Wilson, D. S. The Evolution and Functions of Laughter and Humor: A Synthetic Approach. *Q. Rev. Biol.* **80**, 395–430 (2005).
3. Aktipis, A. *et al.* Understanding cooperation through fitness interdependence. *Nat. Hum. Behav.* **2**, 429–431 (2018).
4. Sauter, D. A., Eisner, F., Calder, A. J. & Scott, S. K. Perceptual Cues in Nonverbal Vocal Expressions of Emotion. *Q. J. Exp. Psychol.* **63**, 2251–2272 (2010).
5. Pell, M. D. *et al.* Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biol. Psychol.* **111**, 14–25 (2015).
6. Scherer, K. R., Banse, R. & Wallbott, H. G. Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross. Cult. Psychol.* **32**, 76–92 (2001).
7. Planalp, S. Varieties of Cues to Emotion in Naturally Occurring Situations. *Cogn. Emot.* **10**, 137–154 (1996).
8. Elfenbein, H. A. & Ambady, N. On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychol. Bull.* **128**, 203–235 (2002).
9. Schröder, M. Experimental study of affect bursts. *Speech Commun.* **40**, 99–116 (2003).
10. Neves, L., Cordeiro, C., Scott, S. K., Castro, S. L. & Lima, C. F. High emotional contagion and empathy are associated with enhanced detection of emotional authenticity in laughter. *Q. J. Exp. Psychol.* **71**, 2355–2363 (2018).
11. Stadel, M., Daniels, J. K., Warrens, M. J. & Jeronimus, B. F. The gender-specific impact of

- emotional tears. *Motiv. Emot.* **43**, 696–704 (2019).
12. Simons, G., Bruder, M., van der Löwe, I. & Parkinson, B. Why Try (Not) to Cry: Intra- and Inter-Personal Motives for Crying Regulation. *Front. Psychol.* **3**, 597 (2013).
 13. Scott, S. K., Lavan, N., Chen, S. & McGettigan, C. The social life of laughter. *Trends Cogn. Sci.* **18**, 618–620 (2014).
 14. Szameitat, D. P. *et al.* Acoustic profiles of distinct emotional expressions in laughter. *J. Acoust. Soc. Am.* **126**, 354–366 (2009).
 15. Lavan, N., Scott, S. K. & McGettigan, C. Laugh Like You Mean It: Authenticity Modulates Acoustic, Physiological and Perceptual Properties of Laughter. *J. Nonverbal Behav.* **40**, 133–149 (2016).
 16. ten Brinke, L. & Porter, S. Cry me a river: Identifying the behavioral consequences of extremely high-stakes interpersonal deception. *Law and Human Behavior* vol. 36 469–477 (2012).
 17. Bryant, G. A. & Aktipis, C. A. The animal nature of spontaneous human laughter. *Evol. Hum. Behav.* **35**, 327–335 (2014).
 18. Lavan, N., Rankin, G., Lorking, N., Scott, S. & McGettigan, C. Neural correlates of the affective properties of spontaneous and volitional laughter types. *Neuropsychologia* **95**, 30–39 (2017).
 19. Anikin, A. & Lima, C. F. Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *Q. J. Exp. Psychol.* **71**, 1–21 (2017).
 20. Szameitat, D. P. *et al.* It is not always tickling: Distinct cerebral responses during perception of different laughter types. *Neuroimage* **53**, 1264–1271 (2010).

21. Calvo, M. G., Marrero, H. & Beltrán, D. When does the brain distinguish between genuine and ambiguous smiles? An ERP study. *Brain Cogn.* **81**, 237–246 (2013).
22. Dawel, A., Palermo, R., O’Kearney, R. & McKone, E. Children can discriminate the authenticity of happy but not sad or fearful facial expressions, and use an immature intensity-only strategy. *Frontiers in Psychology* vol. 6 462 (2015).
23. Wang, C.-A. *et al.* Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task. *Front. Neurol.* **9**, 1029 (2018).
24. van der Wel, P. & van Steenbergen, H. Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon. Bull. Rev.* **25**, 2005–2015 (2018).
25. Kinner, V. L. *et al.* What our eyes tell us about feelings: Tracking pupillary responses during emotion regulation processes. *Psychophysiology* **54**, 508–518 (2017).
26. Mathôt, S. Pupillometry: Psychology, Physiology, and Function. *J. Cogn.* **1**, (2018).
27. Partala, T. & Surakka, V. Pupil size variation as an indication of affective processing. *Int. J. Hum. Comput. Stud.* **59**, 185–198 (2003).
28. Zekveld, A. A., Koelewijn, T. & Kramer, S. E. The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends Hear.* **22**, 2331216518777174 (2018).
29. Kahneman, D. & Beatty, J. Pupil Diameter and Load on Memory. *Science (80-.).* **154**, 1583 LP – 1585 (1966).
30. Oliva, M. & Anikin, A. Pupil dilation reflects the time course of emotion recognition in human vocalizations. *Sci. Rep.* **8**, 4871 (2018).
31. Prochazkova, E. *et al.* Pupil mimicry promotes trust through the theory-of-mind network. *Proc. Natl. Acad. Sci.* **115**, E7265 LP-E7274 (2018).

32. Prochazkova, E. & Kret, M. E. Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neurosci. Biobehav. Rev.* **80**, 99–114 (2017).
33. Lima, C. F., Castro, S. L. & Scott, S. K. When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behav. Res. Methods* **45**, 1234–1245 (2013).
34. Oldfield, R. C. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
35. Derntl, B. *et al.* Facial emotion recognition and amygdala activation are associated with menstrual cycle phase. *Psychoneuroendocrinology* **33**, 1031–1040 (2008).
36. Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* vol. 54 1063–1070 (1988).
37. Galinha, I. C. & Pais-Ribeiro, L. Contribuição para o estudo da versão portuguesa da Positive and Negative Affect Schedule (PANAS): I - Abordagem teórica ao conceito de afecto. *Análise Psicológica* **23**, 209–218 (2005).
38. Wechsler, D. *Wechsler Adult Intelligence Scale (WAIS)*. (2008).
39. Warren, J. E. *et al.* Positive Emotions Preferentially Engage an Auditory–Motor “Mirror” System. *J. Neurosci.* **26**, 13067 LP – 13075 (2006).
40. Lavan, N., Lima, C. F., Harvey, H., Scott, S. K. & McGettigan, C. I thought that I heard you laughing: Contextual facial expressions modulate the perception of authentic laughter and crying. *Cogn. Emot.* **29**, 935–944 (2015).
41. Fletcher, P. D. *et al.* Dementias show differential physiological responses to salient sounds. *Front. Behav. Neurosci.* **9**, 73 (2015).

42. Mathôt, S., Fabius, J., Van Heusden, E. & Van der Stigchel, S. Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods* **50**, 94–106 (2018).
43. Wang, Y., Toor, S. S., Gautam, R. & Henson, D. B. Blink Frequency and Duration during Perimetry and Their Relationship to Test–Retest Threshold Variability. *Invest. Ophthalmol. Vis. Sci.* **52**, 4546–4550 (2011).
44. Schleicher, R., Galley, N., Briest, S. & Galley, L. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics* **51**, 982–1010 (2008).
45. Caffier, P. P., Erdmann, U. & Ullsperger, P. Experimental evaluation of eye-blink parameters as a drowsiness measure. *Eur. J. Appl. Physiol.* **89**, 319–325 (2003).
46. Foroughi, C. K., Sibley, C. & Coyne, J. T. Pupil size as a measure of within-task learning. *Psychophysiology* **54**, 1436–1443 (2017).
47. Urai, A. E., Braun, A. & Donner, T. H. Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nat. Commun.* **8**, 14637 (2017).
48. Widmann, A., Schröger, E. & Wetzels, N. Emotion lies in the eye of the listener : Emotional arousal to novel sounds is reflected in the sympathetic contribution to the pupil dilation response and the P3. *Biol. Psychol.* **133**, 10–17 (2018).
49. Finke, J. B., Deuter, C. E., Hengesch, X. & Schächinger, H. The time course of pupil dilation evoked by visual sexual stimuli: Exploring the underlying ANS mechanisms. *Psychophysiology* **54**, 1444–1458 (2017).
50. Winn, M. B., Wendt, D., Koelewijn, T. & Kuchinsky, S. E. Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends Hear.* **22**, 2331216518800869 (2018).

51. Wahn, B., Ferris, D. P., Hairston, W. D. & König, P. Pupil Sizes Scale with Attentional Load and Task Experience in a Multiple Object Tracking Task. *PLoS One* **11**, e0168087 (2016).
52. Katidioti, I., Borst, J. P. & Taatgen, N. A. What happens when we switch tasks: Pupil dilation in multitasking. *Journal of Experimental Psychology: Applied* vol. 20 380–396 (2014).
53. Tukey, J. W. *Exploratory data analysis*. (Addison-Wesley Pub. Co., 1977).
54. Kleiner, M. *et al.* What's new in psychtoolbox-3. *Perception* **36**, 1–16 (2007).
55. R Core Team. R: A Language and Environment for Statistical Computing. (2014).
56. Bakdash, J. Z. & Marusich, L. R. Repeated Measures Correlation. *Frontiers in Psychology* vol. 8 456 (2017).
57. Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297–334 (1951).
58. Little, R. J. A. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *J. Am. Stat. Assoc.* **83**, 1198–1202 (1988).
59. Jager, K. J., Zoccali, C., MacLeod, A. & Dekker, F. W. Confounding: What it is and how to deal with it. *Kidney Int.* **73**, 256–260 (2008).
60. Nakakoga, S., Higashi, H., Muramatsu, J., Nakauchi, S. & Minami, T. Asymmetrical characteristics of emotional responses to pictures and sounds: Evidence from pupillometry. *PLoS One* **15**, 1–16 (2020).
61. Burley, D. T., Gray, N. S. & Snowden, R. J. As Far as the Eye Can See: Relationship between Psychopathic Traits and Pupil Response to Affective Stimuli. *PLoS One* **12**, 1–22 (2017).
62. Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* **41**, 1149–1160 (2009).

63. Babiker, A., Faye, I., Prehn, K. & Malik, A. Machine learning to differentiate between positive and negative emotions using pupil diameter. *Front. Psychol.* **6**, 1–10 (2015).
64. Jin, A. B., Steding, L. H. & Webb, A. K. Reduced emotional and cardiovascular reactivity to emotionally evocative stimuli in major depressive disorder. *Int. J. Psychophysiol.* **97**, 66–74 (2015).
65. Gingras, B., Marin, M. M., Puig-Waldmüller, E. & Fitch, W. T. The Eye is Listening: Music-Induced Arousal and Individual Differences Predict Pupillary Responses. *Front. Hum. Neurosci.* **9**, 619 (2015).
66. Kret, M. E., Fischer, A. H. & De Dreu, C. K. W. Pupil Mimicry Correlates With Trust in In-Group Partners With Dilating Pupils. *Psychol. Sci.* **26**, 1401–1410 (2015).
67. Seifritz, E. *et al.* Differential sex-independent amygdala response to infant crying and laughing in parents versus nonparents. *Biol. Psychiatry* **54**, 1367–1375 (2003).
68. Newman, J. D. Neural circuits underlying crying and cry responding in mammals. *Behav. Brain Res.* **182**, 155–165 (2007).
69. Williams, L. M. *et al.* Arousal Dissociates Amygdala and Hippocampal Fear Responses: Evidence from Simultaneous fMRI and Skin Conductance Recording. *Neuroimage* **14**, 1070–1079 (2001).
70. Madan, C. R., Fujiwara, E., Caplan, J. B. & Sommer, T. Emotional arousal impairs association-memory: Roles of amygdala and hippocampus. *Neuroimage* **156**, 14–28 (2017).
71. Hirsch, C. R., Meeten, F., Krahe, C. & Reeder, C. Resolving Ambiguity in Emotional Disorders: The Nature and Role of Interpretation Biases. *Annu. Rev. Clin. Psychol.* **12**, 281–305 (2016).
72. Peysakhovich, V., Dehais, F. & Causse, M. Pupil Diameter as a Measure of Cognitive Load

- during Auditory-visual Interference in a Simple Piloting Task. *Procedia Manuf.* **3**, 5199–5205 (2015).
73. van Roeyen, I., Riem, M. M. E., Tonicic, M. & Vingerhoets, A. J. J. M. The Damaging Effects of Perceived Crocodile Tears for a Crier's Image. *Front. Psychol.* **11**, 172 (2020).
74. Zeskind, P. S. Infant crying and the synchrony of arousal. in *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man* (eds. Altenmüller, E., Schmidt, S. & Zimmermann, E.) 155–172 (Oxford Scholarship Online, 2013). doi:10.1093/acprof:oso/9780199583560.001.0001.
75. Walter, H. Social Cognitive Neuroscience of Empathy: Concepts, Circuits, and Genes. *Emot. Rev.* **4**, 9–17 (2012).
76. Nummenmaa, L., Hirvonen, J., Parkkola, R. & Hietanen, J. K. Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage* **43**, 571–580 (2008).
77. Lima, C. F., Castro, S. L. & Scott, S. K. When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behav. Res. Methods* **45**, 1234–1245 (2013).
78. Provine, R. R. Laughter as a scientific problem: An adventure in sidewalk neuroscience. *J. Comp. Neurol.* **524**, 1532–1539 (2016).
79. Provine, R. R. Contagious laughter: Laughter is a sufficient stimulus for laughs and smiles. *Bull. Psychon. Soc.* **30**, 1–4 (1992).
80. Rosa, P. J., Esteves, F. & Arriaga, P. Beyond Traditional Clinical Measurements for Screening Fears and Phobias. *IEEE Trans. Instrum. Meas.* **64**, 3396–3404 (2015).
81. Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15**, 1040–1046 (2012).

82. Kreibig, S. D., Wilhelm, F. H., Roth, W. T. & Gross, J. J. Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology* **44**, 787–806 (2007).

Figures Legends

Figure 1 – Pupil diameter change relative to baseline (in arbitrary units (AU)) for each condition, averaged across all participants. Dashed vertical lines delineate the time windows used for analysis.

Figure 2 – Box plots of maximum pupil diameter as a function of authenticity per emotion, showcasing the authenticity by emotion interaction (above), and of maximum pupil dilation as a function of emotion showcasing the main effect of emotion (below) – both for each time window (seconds). Statistically significant ($p < .05$, after FDR correction) pairwise comparisons are highlighted with an *. Error bars represent ± 1.5 SD; AU = arbitrary units.

Figure 3 – Box plots of mean pupil diameter as a function of authenticity per emotion, showcasing the authenticity by emotion interaction (above) and of mean pupil dilation as a function of emotion showcasing the main effect of emotion (below) – both for each time window (seconds). Statistically significant ($p < .05$, after FDR correction) pairwise comparisons are highlighted with an *. Error bars represent ± 1.5 SD; AU = Arbitrary Units

Tables

Table 1 - F-statistic (numerator and denominator degrees of freedom), p -value after multiple comparison correction with False Discovery Rate (FDR) and partial eta squared (η_p^2) for the authenticity by emotion interaction on maximum and mean pupil size, in each time window, are presented. Mean difference (M_{diff}), standard error difference (SD_{diff}), t-test (and degree of freedom), p -value after FDR correction and η_p^2 are also presented for the follow-up post-hoc pairwise comparisons whereby authentic versus acted vocalizations were contrasted for each emotion (whilst means and SD per condition are provided in the Results' section text).

		Authenticity x Emotion interaction			Pairwise Comparisons (Authentic vs. Acted)						
Time window (seconds)	Pupil size measure	F-statistic (1, 27)	FDR corrected p -value	η_p^2	Emotion	M_{diff}	SD_{diff}	t-test (27)	FDR corrected p -value	η_p^2	Significant Contrasts
0-1	Maximum	5.77	.047*	.18	Laughter	-25.10	41.42	-3.23	.006*	.28	Acted > Authentic
					Crying	0.47	31.38	0.08	.937	<.01	-
	Mean	4.12	.052	.13	Laughter	-12.45	29.11	-2.26	.064	.16	-
					Crying	6.33	28.42	1.18	.249	.05	-
1-2	Maximum	9.36	.005*	.26	Laughter	-27.03	67.86	-2.11	.044*	.14	Acted > Authentic
					Crying	24.98	45.42	2.91	.014*	.24	Authentic > Acted
	Mean	13.82	.002*	.34	Laughter	-33.36	55.36	-3.19	.007*	.27	Acted > Authentic
					Crying	18.59	46.18	2.13	.042*	.14	Authentic > Acted
2-3	Maximum	6.61	.032*	.20	Laughter	-24.33	75.27	-1.71	.099	.10	-
					Crying	25.33	56.78	2.36	.051	.17	-
	Mean	2.75	.109	.09	Laughter	-14.91	85.07	-0.93	.362	.03	-
					Crying	21.69	56.28	2.04	.103	.13	-
3-4	Maximum	6.79	.019*	.20	Laughter	-16.90	80.61	-1.11	.277	.04	-
					Crying	49.36	85.42	3.06	.010*	.26	Authentic > Acted
	Mean	6.23	.019*	.19	Laughter	4.42	75.92	0.31	.760	<.01	-
					Crying	52.26	86.27	3.21	.007*	.28	Authentic > Acted

Table 2 – Pearson’s *r* and *p*-values for the correlations between pupil size measures (collapsed across all time windows) and trait empathy scores (Empathy Quotient, and its subscales, and reading the Mind in the Eyes Test scores), in each emotion separately.

Empathy Trait Scores	Pupil size measures			
	Maximum		Mean	
	Laughter			
	Pearson’s <i>r</i>	<i>p</i> -value	Pearson’s <i>r</i>	<i>p</i> -value
Empathy Quotient (EQ; total)	.28	.151	.12	.543
EQ Cognitive Empathy	-.37	.050	-.42	.028*
EQ Social Skills	-.14	.467	-.05	.791
EQ Emotional Reactivity	-.17	.379	.04	.835
EQ Empathic Difficulty	.29	.130	.11	.590
Reading the Mind in the Eyes Test (RMET)	-.11	.567	-.30	.125
	Crying			
Empathy Quotient (EQ; total)	.22	.258	.08	.677
EQ Cognitive Empathy	-.27	.157	-.20	.297
EQ Social Skills	-.08	.655	.18	.371
EQ Emotional Reactivity	-.15	.450	.21	.290
EQ Empathic Difficulty	.27	.170	.05	.784
Reading the Mind in the Eyes Test (RMET)	-.07	.718	-.22	.265

Statistical significance level: * $p < .05$