Humans read emotional arousal in monkey vocalizations: Evidence for evolutionary continuities in communication

Jay W Schwartz [Corresp., 1, 2] , **Harold Gouzoules** [1]

[1] Department of Psychology, Emory University, Atlanta, GA, United States

[2] Psychological Sciences Department, Western Oregon University, Monmouth, OR, United States

Corresponding Author: Jay W Schwartz
Email address: schwartzj@wou.edu

Humans and other mammalian species communicate emotions in ways that reflect evolutionary conservation and continuity, an observation first made by Darwin. One approach to testing this hypothesis has been to assess the capacity to perceive the emotional content of the vocalizations of other species. Using a binary forced choice task, we tested perception of the emotional intensity represented in coos and screams of infant and juvenile female rhesus macaques (*Macaca mulatta*) by 113 human listeners without, and 12 listeners with, experience (as researchers or care technicians) with this species. Each stimulus pair contained one high- and one low-arousal vocalization, as measured at the time of recording by stress hormone levels for coos and the degree of intensity of aggression for screams. For coos as well as screams, both inexperienced and experienced participants accurately identified the high-arousal vocalization at significantly above-chance rates. Experience was associated with significantly greater accuracy with scream stimuli but not coo stimuli, and with a tendency to indicate screams as reflecting greater emotional intensity than coos. Neither measures of empathy, human emotion recognition, nor attitudes toward animal welfare showed any relationship with responses. Participants were sensitive to the fundamental frequency, noisiness, and duration of vocalizations; some of these tendencies likely facilitated accurate perceptions, perhaps due to evolutionary homologies in the physiology of arousal and vocal production between humans and macaques. Overall, our findings support a view of evolutionary continuity in emotional vocal communication. We discuss hypotheses about how distinctive dimensions of human nonverbal communication, like the expansion of scream usage across a range of contexts, might influence perceptions of other species' vocalizations.

1 **Humans read emotional arousal in monkey vocalizations: Evidence for evolutionary**
2 **continuities in communication**

5 Jay W. Schwartz[a,b] & Harold Gouzoules[a]

6 [a] Department of Psychology, Emory University, Atlanta, GA
7 [b] Psychological Sciences Department, Western Oregon University, Monmouth, OR

9 Corresponding Author:
10 Jay W. Schwartz
11
12 Psychological Sciences Department, Western Oregon University, 345 N. Monmouth Ave.,
13 Monmouth, OR 97361
14
15 Phone: (503) 838-9514
16 E-mail: schwartzj@wou.edu

32 **Abstract**

33 Humans and other mammalian species communicate emotions in ways that reflect evolutionary

34 conservation and continuity, an observation first made by Darwin. One approach to testing this

35 hypothesis has been to assess the capacity to perceive the emotional content of the vocalizations

36 of other species. Using a binary forced choice task, we tested perception of the emotional

37 intensity represented in coos and screams of infant and juvenile female rhesus macaques

38 (*Macaca mulatta*) by 113 human listeners without, and 12 listeners with, experience (as

39 researchers or care technicians) with this species. Each stimulus pair contained one high- and one

40 low-arousal vocalization, as measured at the time of recording by stress hormone levels for coos

41 and the degree of intensity of aggression for screams. For coos as well as screams, both

42 inexperienced and experienced participants accurately identified the high-arousal vocalization at

43 significantly above-chance rates. Experience was associated with significantly greater accuracy

44 with scream stimuli but not coo stimuli, and with a tendency to indicate screams as reflecting

45 greater emotional intensity than coos. Neither measures of empathy, human emotion recognition,

46 nor attitudes toward animal welfare showed any relationship with responses. Participants were

47 sensitive to the fundamental frequency, noisiness, and duration of vocalizations; some of these

48 tendencies likely facilitated accurate perceptions, perhaps due to evolutionary homologies in the

49 physiology of arousal and vocal production between humans and macaques. Overall, our

50 findings support a view of evolutionary continuity in emotional vocal communication. We

51 discuss hypotheses about how distinctive dimensions of human nonverbal communication, like

52 the expansion of scream usage across a range of contexts, might influence perceptions of other

53 species' vocalizations.

54 **Keywords:** Acoustics, arousal, communication, emotion, empathy, expression, *Macaca mulatta*,
55 nonverbal, primate, scream

56   **Introduction**

57   The hypothesis that the ways in which humans express emotions shared with other

58   species due to common descent dates back to Charles Darwin (1872). Nonverbal vocalizations

59   are a major means of communicating emotion, and the acoustic structures, functions, and

60   production mechanisms of human nonverbal vocalizations are largely evolutionarily conserved

61   and shared with other mammalian taxa (Sauter et al., 2010; Anikin, Bååth & Persson, 2018;

62   Pisanski et al., 2022). Also largely conserved across mammals are the acoustic correlates of

63   emotional states within vocalizations (Briefer, 2012; Zimmermann, Leliveld & Schehka, 2013).

64   For example, emotional arousal (activation of the sympathetic nervous system in association

65   with intense emotions; Mendl, Burman, & Paul, 2010; Russell, 2003) correlates positively with

66   the fundamental frequency (F0; perceived as pitch) of vocalizations across taxa, likely due to

67   tensing of the vocal fold muscles increasing the rate of oscillation (Scherer, 1986; Titze, 1994;

68   Riede, 2010; Briefer, 2012). Other acoustic correlates of emotional arousal include F0

69   modulation, vocal duration, and noisiness (Briefer, 2012). The effects of arousal on these

70   acoustic parameters likely reflect evolutionary homologies in the mechanisms of arousal and

71   vocal production.

72   The hypothesis of evolutionary homology in vocal emotion expression predicts that

73   listeners should be sensitive to the acoustic cues to the emotional states of conspecific and

74   heterospecific vocalizers alike, depending in part on the phylogenetic distance between the two

75   species and the evolutionary history of the particular vocalization in question (e.g., Fritz et al.,

76   2018; though see Filippi et al., 2017a). One approach to testing this prediction is to examine

77   human perceptions of emotion from heterospecific vocalizations (Linnankoski et al., 1994;

78   Nicastro & Owren, 2003; Pongrácz et al., 2005; Belin et al., 2008; McComb et al., 2009; Tallet

79   et al., 2010; Faragó et al., 2014; Scheumann et al., 2014; Faragó et al., 2017; Maruščáková et al.,

80   2015; Filippi et al., 2017a,b; Kelly et al., 2017; Fritz et al., 2018; Kamiloğlu et al., 2020;

81   Merkies, Crouchman & Belliveau, 2021; Massenet et al., 2022). These studies have generally

82   confirmed human sensitivity to emotional information available in heterospecific vocalizations,

83   and thus support Darwin's (1872) hypothesis that significant aspects of human emotional

84   expression are evolutionarily homologous with those of other species.

85   Listeners have tended to base perceptions of emotional arousal in heterospecific

86   vocalizations predominantly on the mean F0 (Nicastro & Owren, 2003; Pongrácz et al., 2005;

87   Faragó et al., 2014, 2017; Maruščáková et al., 2015; Filippi et al., 2017a; Kelly et al., 2017),

88   Other significant acoustic parameters have included vocal duration (Nicastro & Owren, 2003;

89   Faragó et al., 2014, 2017; Maruščáková et al., 2015) and noisiness (Faragó et al., 2014; Filippi et

90   al., 2017b; Massenet et al., 2022). Human listeners appear to show similar perceptual tendencies

91   with regard to the F0 and noisiness of conspecific nonverbal vocalizations (Juslin & Laukka,

92   2003; Sauter et al., 2010; Faragó et al., 2014; Schwartz & Gouzoules, 2019; Anikin et al., 2021).

93   Insofar as these acoustic parameters naturally vary according to vocalizer arousal (Briefer, 2012),

94   these perceptual tendencies are generally credited with facilitating accurate perceptions of

95   emotion from heterospecific vocalizations (for an exception, see Kelly et al., 2017).

96   A complicating factor is that many studies in this area have included multiple call types

97   as stimuli (for exceptions, see Nicastro & Owren, 2003; Pongrácz et al., 2005; Filippi et al.,

98   2017a,b; Faragó et al., 2017; Merkies, Crouchman & Belliveau, 2021; Massenet et al., 2022),

99   making it difficult to know whether listeners were reacting to acoustic variation *within* a single

100   call type, or *between* call types (for more on this distinction, see Fischer, Wadewitz &

101   Hammerschmidt, 2016; Schwartz, Engelberg & Gouzoules, 2020). Distinct call types typically

102 serve disparate social functions and potentially reflect differing cognitive processes, whereas

103 acoustic variation within a call type more reliably relates to the vocalizer's emotional state

104 (Schwartz, Engelberg & Gouzoules, 2020). It is therefore pertinent to investigate human

105 perception of emotion from acoustic variation within a call type of another species.

106   An additional issue is whether heterospecific emotion perception is best explained by

107 evolutionary homology, familiarity, or domestication (Maigrot, Hillmann & Briefer, 2022). Most

108 previous studies of human perception have used sounds from domestic animals such as cats,

109 dogs, and livestock (for exceptions, see Filippi et al., 2017a; Kelly et al., 2017; Fritz et al., 2018;

110 Kamiloğlu et al., 2020). These, of course, are species with which many humans are familiar—

111 and indeed, which might have undergone evolutionary changes in emotional expression to

112 improve communication with humans (Pongrácz, Molnár, & Miklósi, 2010). Some studies have

113 suggested accurate perception does not depend on familiarity (Pongrácz et al., 2005; Filippi et

114 al., 2017a; Merkies, Crouchman & Belliveau, 2021; Maigrot, Hillmann & Briefer, 2022), but

115 several studies have demonstrated an association between experience with a particular species

116 and increased accuracy in recognizing the emotional significance of its vocalizations (Nicastro &

117 Owren, 2003; McComb et al., 2009; Tallet et al., 2010; Scheumann et al., 2014; Faragó et al.,

118 2017; Parsons et al., 2019). This raises the question of the degree to which human accuracy in

119 heterospecific emotion recognition tasks reflects evolutionary homology, as opposed to a

120 capacity to learn to recognize any individual species' unique emotional expressions through

121 exposure.

122   The present study examined human perceptions of emotional arousal from acoustic

123 variation within call types of rhesus macaques (*Macaca mulatta*), a cercopithecid primate whose

124 lineage is estimated to have diverged from that of humans c. 30 million years ago (Steiper &

125 Young, 2008). Although rhesus macaques are common in research institutions and in urban areas

126 in South Asia, many humans have little to no direct experience observing or interacting with

127 them, presenting an opportunity to test the degree to which humans can perform fine-grained

128 discriminations within call types of an undomesticated, unfamiliar species. The rhesus macaque

129 vocal repertoire contains a variety of call types including coos—tonal vocalizations associated

130 with social functions ranging from signaling the presence of food to reuniting separated kin

131 (Bayart et al., 1990; Hauser, 1991; Hauser & Marler, 1993)—and screams—high-F0

132 vocalizations thought to be evolutionarily homologous with the screams of other primates

133 including humans, usually produced during agonistic interactions to recruit aid from matrilineal

134 kin and allies (Gouzoules, Gouzoules & Marler, 1984; Gouzoules, Gouzoules & Tomaszycki,

135 1998; Gouzoules, 2005). A recent study that compared stress hormone levels and coo

136 vocalizations produced in a formal behavioral test (Human Intruder Test; Kalin & Shelton,

137 1989), and also compared screams produced during naturally during social interactions that

138 varied in the degree of aggression received by the vocalizer, suggested that both call types

139 exhibit within-type acoustic variation that correlates with the emotional arousal of the vocalizer

140 (Schwartz, Sanchez & Gouzoules, 2022). The present study examined whether human listeners

141 are sensitive to the emotional significance of variation among coos and screams, and assessed the

142 role of experience, other participant characteristics, and acoustic properties.

143   In a previous study, human listeners did not accurately discriminate between positively

144 and negatively valenced rhesus macaque vocalizations including screams (Fritz et al., 2018),

145 although human listeners have shown sensitivity to other kinds of acoustic information in this

146 species' calls (Fugate, Gouzoules & Nygaard, 2008). Previous research has also shown

147 sensitivity to arousal cues in the vocalizations the closely related Barbary macaque, *Macaca*

148 *sylvanus* (Filippi et al., 2017a). We therefore hypothesized that participants would accurately

149 perceive differences in emotional arousal within pairs of coos and of screams. We also examined

150 perceived arousal differences between the two call types.

151   An additional objective was to test the hypothesis that familiarity affects sensitivity to the

152 emotional significance of another species' vocalizations (Nicastro & Owren, 2003; Tallet et al.,

153 2010; Parsons et al., 2019). We also examined the effects of empathy (Cambridge Behaviour

154 Scale; Baron-Cohen & Wheelwright, 2004), human emotion recognition (Cambridge Face-Voice

155 Battery; Golan, Baron-Cohen, & Hill, 2006; Golan et al., 2007), and attitudes toward animal

156 welfare (Animal Attitudes Scale; Herzog, Grayson, & McCord, 2015). The hypothesis that such

157 characteristics might correlate with the ability to recognize emotion from heterospecific

158 vocalizations is intriguing, but previous research to date has provided little support for this

159 notion (Maruščáková et al., 2015).

160   We predicted that mean F0 would play a role in participants' perceptions of arousal, in

161 line with previous research on human emotion perception, and with the positive relationship

162 between arousal and the overall F0 of rhesus macaque coos and screams (Schwartz, Sanchez &

163 Gouzoules, 2022). We also examined the role of noisiness and duration of vocalizations, since

164 these have been shown by the studies cited above to affect listeners' perceptions of the emotional

165 states of animals, and F0 range, which appears to correlate with emotional arousal in rhesus

166 macaque coos (Schwartz, Sanchez & Gouzoules, 2022).

167

168   **Materials & Methods**

169 **Ethical Statement**

170   All animal procedures were approved by the Emory University Institutional Animal Care

171   and Use Committee (YER-2003417- 032317A) in accordance with the Animal Welfare Act and

172   the U.S. Department of Health and Human Services "Guide for Care and Use of Laboratory

173   Animals." Procedures involved in sourcing vocalizations used as stimuli for this study took place

174   as part were part of longitudinal studies at the Emory National Primate Research Center

175   (ENPRC) funded by the U.S. National Institutes of Health NIH/NICHD R01 grant HD077623

176   (Principle Investigators: M. M. Sanchez & M. E. Wilson), independent of this study of vocal

177   perception. Human testing procedures were approved by and conducted in compliance with the

178   Emory University Institutional Review Board (IRB00102796).

179

180   **Participants**

181   One hundred thirteen undergraduate students at Emory University and volunteers from

182   the Atlanta area comprised our "inexperienced" sample; each indicated no experience working

183   with or observing rhesus macaques. Emory undergraduate participants were recruited via an

184   online portal, and received credit in an introductory psychology course; other participants

185   contacted the researchers after seeing fliers advertising the study posted around Emory's campus,

186   and were not compensated in any way. Twelve individuals employed by or affiliated with the

187   ENPRC comprised our "experienced" sample, each reporting a non-zero number of months

188   working with rhesus macaques (mean $\pm$ SD = 111 $\pm$ 76 months). These participants volunteered

189   after hearing about the study through an email announcement or word-of-mouth from other

190   ENPRC employees. All who were recruited using the above methods were invited to complete

191   the study; all those who did were sampled.

192   Written informed consent was obtained from all individual participants included in the

193   study. Each participant reported their gender (**inexperienced:** 73% female, 27% male, 0% other;

194   **experienced:** 67% female, 33% male, 0% other), age (**inexperienced** mean ± SD = 20.2 ± 4.2

195   years; **experienced** = 34.2 ± 8.2 years), and native language (**inexperienced:** 67% English

196   [including 6% natively bilingual], 19% Mandarin/Chinese, 4% Spanish, 3% Korean, 9% other

197   languages [< 2% each]; **experienced:** 100% English [including one natively bilingual

198   participant]). All but one of the experienced participants indicated professional work with other

199   taxa in addition to rhesus macaques, including chimpanzees and sooty mangabeys (housed at the

200   ENPRC), pigtailed macaques (formerly housed at the ENPRC), other great apes, lemurs, dogs,

201   and cats. Twenty-one inexperienced participants indicated professional work with animals,

202   including dogs, cats, and/or rodents. Participants were screened for hearing impairments, and

203   disclosure of an impairment that might interfere with perception of auditory stimuli was

204   considered an exclusionary criterion.

205

206   **Stimuli**

207   The stimuli used in this study comprised pairs of rhesus macaque vocalizations—either

208   two coos, two screams, or one coo and one scream (hereafter, "mixed")—separated by one

209   second of silence. These vocalizations were collected from infant and juvenile female rhesus

210   macaques living in naturalistic social groups within large outdoor enclosures at the ENPRC Field

211   Station in Lawrenceville, GA, using procedures described in detail in (Schwartz, Sanchez &

212   Gouzoules, 2022). Briefly, coos of 39 macaques were recorded during Human Intruder Tests

213   (Kalin & Shelton, 1989) consisting of three 10-minute stages—Alone (the macaque was alone in

214   a roomy cage within a testing room), Profile (a technician wearing a lifelike mask sat motionless

215   at a 90-degree angle to the cage), and Stare (the technician sat facing the cage and directing gaze

216   at the macaque but otherwise remained motionless). A Sony DCR-SR85 video camera rig

217   (Tokyo, Japan) (onboard Sony electret condenser microphone) captured video and audio

218   recordings of the tests, including any vocalizations uttered by the macaques. Audio was digitized

219   in a 48-kHz, 32-bit .WAV format. Plasma cortisol concentrations were measured from blood

220   sampled obtained immediately before and immediately after each test; the difference between

221   these two measurements served as an estimation of each macaque's physiological stress reaction

222   to the test. Screams of 18 macaques (a subset of the 39 from which coos were obtained) were

223   recorded during naturally occurring social interactions within the macaques' home enclosures,

224   using a Sennheiser ME66 directional microphone with MZW66 windshield (Wennebostel,

225   Germany) and a Marantz PMD671 solid state recorder (Tokyo, Japan). Recordings were

226   digitized as 44.1-kHz, 16- or 24-bit .WAV files. Each scream bout recording was accompanied

227   by a note of the vocalizer identity and whether the vocalizer received agonistic physical contact

228   prior to the scream(s).

229   Vocalizations were extracted from larger recordings and prepared for use as stimuli using

230   Adobe Audition CC (Adobe Systems, San Jose, CA). 48-kHz audio from the Human Intruder

231   Test coo recordings was down-sampled to 44.1 kHz and dithered from 32- to 16-bit depth, and

232   24-bit scream recordings were dithered to 16-bit depth. The Human Intruder Testing room

233   contained a noise generator producing brown noise at high volume, to create a controlled and

234   constant auditory background for testing; a noise reduction algorithm was used to prepare the

235   coos for presentation and acoustic analysis (described in detail in Schwartz, Sanchez, and

236   Gouzoules, 2022). The generally very high F0 of the screams (2-10 kHz), permitted removal of

237   background noise at lower frequencies; this was done by selecting noise at lower frequencies

238   using the marquee and lasso selection tools in Audition, and reducing the sound level of the

239   selection. After noise reduction/removal, all vocalizations underwent RMS amplitude

240   normalization (following best practices outlined by Owren & Bachorowski, 2007) in Audition,

241   using the equal loudness contour (which weights frequency amplitudes according to their

242   perceptual loudness). As a result, most of the vocalizations' sound envelopes occupied

243   approximately -12 dB, while the remaining background noise of the coos occupied a sound level

244   of approximately -42 dB. Final waveforms and spectrograms of individual vocalizations were

245   inspected to ensure that they contained no clipping or other distortions.

246   To construct the coo stimulus pairs, 6-month-old macaques were paired into dyads

247   consisting of one high-arousal macaque and one low-arousal macaque, defined by a difference in

248   cortisol reactivity (i.e., increase in plasma cortisol concentration from before until after the test)

249   of $\geq$ 1.5 SD. No dyad exhibited a difference in baseline cortisol concentrations of > 1 SD. Every

250   dyad meeting these criteria was used to create stimulus pairs, resulting in 28 dyads consisting of

251   13 individuals. Individual differences rather than arousal levels inevitably accounted for some

252   acoustic differences among coos within stimulus pairs; in principle, acoustic variation due to

253   individual characteristics should be random and should not be expected to have confounded the

254   results. For each individual, three coos were sampled from the Stare stage: the closest to the 2.5-

255   minute mark, the closest to the 5-minute mark, and the closest to the 7.5-minute mark. Then for

256   each macaque dyad, three coo stimulus pairs were constructed: one pairing the two 2.5-minute

257   coos, one pairing the two 5-minute coos, and one pairing the two 7.5-minute coos. This resulted

258   in a total of 84 coo stimulus pairs, each a combination of 39 individual coos. Each coo in each

259   stimulus pair was preceded and followed by a fade-in and fade-out of the background noise

260   lasting 0.3 s.

261   Each scream stimulus pair consisted of two screams from a single macaque during two

262   separate social interactions—one scream uttered after experiencing agonistic physical contact

263   (high arousal) and one with no contact (low-arousal). To construct these stimuli, tonal screams

264   (Gouzoules, Gouzoules & Marler, 1984) were sampled from the 8 macaques from whom tonal

265   screams were recorded during at least one interaction with contact and at least one without. We

266   selected one tonal scream from each bout (i.e., multiple screams emitted in response to a single

267   agonistic interaction) on the basis of recording quality (no clipping, uninterrupted by other

268   vocalizations or loud sounds, e.g., a monkey striking a fence or enrichment structure during a

269   display). All possible combinations of one high-arousal scream and one low-arousal scream from

270   the same individual were exploited to construct stimuli, resulting in 39 stimulus pairs

271   representing 33 screams. Numbers of screams per macaque ranged from 2-9, resulting in 1-18

272   scream stimulus pairs per macaque.

273   Each mixed stimulus pair included one scream and one coo; no individual vocalization

274   was repeated across these mixed stimuli. Each of the 33 screams was randomly paired with a

275   coo—either from a subset of the coos comprising the coo stimulus pairs, or randomly chosen

276   from a larger corpus—resulting in 33 mixed stimulus pairs. To account for any potential

277   confounding effects of the presence of background noise in the coo recordings versus absence

278   thereof in the scream recordings, we added pure noise, sampled from the Human Intruder Test

279   (coo) recordings, to the scream recordings to match the frequency profile and sound level of the

280   coo recordings, and included 0.3-s fade-ins and fade-outs before and after the screams to match

281   the coos.

282   The order of the two vocalizations comprising each stimulus pair was alternated so that

283   an equal number of stimulus pairs contained a high-arousal vocalization first and a low-arousal

284   vocalization second, and that each individual vocalization appeared first equally often as it

285   appeared second. Where applicable, the one second of silence between the two sound files began

286   at the offset of the noise fade-out of the first vocalization, and ended at the onset of the noise

287   fade-in of the second vocalization. The stimulus pairs were re-inspected by ear to ensure that the

288   RMS-amplitude normalization employed earlier had in fact resulted in equal perceptual loudness.

289   In a small number of cases where one vocalization sounded louder than the other to the

290   researchers, the amplitude of one was adjusted manually in Audition.

291

292   **Procedure**

293   Testing took place in the Bioacoustics Laboratory at Emory University (inexperienced

294   participants, and 3 of 12 experienced participants), or in an office in the research building at the

295   ENPRC Field Station (9 of 12 experienced participants). After providing their informed consent

296   and demographic information, inexperienced participants completed two pen-and-paper surveys:

297   the 40-item Cambridge Behavioral Scale (Baron-Cohen & Wheelwright, 2004), which yields a

298   score between 0 and 80 with higher scores indicating greater empathy, and the 10-item brief

299   Animal Attitudes Scale (Herzog, Grayson & McCord, 2015), which yields a score between 0 and

300   50 with higher scores indicating greater concern for animal welfare. Due to job-related time

301   constraints, experienced participants did not complete these surveys.

302   Subsequent procedures took place on a Dell Optiplex 755 desktop computer (Round

303   Rock, TX) running E-Prime 2.0 software (Psychology Software Tools, Inc., Pittsburgh, PA).

304   Participants were instructed that they would hear a series of pairs of vocalizations and that their

305   task was to choose, for each pair, the vocalization that they thought reflected a more intense

306   emotion. They were given the examples of calmness or boredom as reflecting low-intensity

307 emotions, and fear or stress as examples of high-intensity emotional situations. We used the

308 terminology of intense emotions rather than "arousal" in order to circumvent participants'

309 potential unfamiliarity with the formal concept of arousal. Arousal has been described as the

310 intensity of an emotional state (e.g., Briefer, 2012; though see Russell, 2003), and previous

311 research has shown a strong positive correlation between listener ratings of the 'intensity' and

312 'activation' (i.e., arousal) of vocal stimuli (Laukka, Juslin & Bresin, 2005).

313   Each participant listened to every stimulus pair through headphones (Beyerdynamic DT

314 770 Pro, Beyerdynamic GmbH & Co., Heilbronn, Germany, or Sony MDR-CD180, Tokyo,

315 Japan). On each trial, they indicated whether they perceived the first or second vocalization to be

316 more emotionally intense by pressing a '1' or '2' key on a peripheral serial response box (model

317 200a, Psychology Software Tools, Inc., Pittsburgh, PA). Each response triggered a two-second

318 pause followed by the onset of the subsequent stimulus pair. The task was broken into four

319 blocks of stimulus pairs: screams (N = 39 stimulus pairs), mixed (N = 33), one half of the coo

320 stimuli (N = 42), and the other half of the coo stimuli (N = 42). These four blocks were presented

321 in a random order, as were the stimulus pairs within each block. After each block, participants

322 were offered the option to take a break lasting as long as they needed, ending when the

323 participant pressed a key indicating they were ready to continue.

324   The subset of inexperienced participants reporting English as a first language then

325 completed the Cambridge Mindreading Face-Voice (CAM) battery (Golan, Baron-Cohen & Hill,

326 2006; Golan et al., 2007). This 100-item task was presented on the same computer; in it,

327 participants view or listen to human emotional facial or vocal expressions and, for each, choose

328 one of four emotion labels to best describe the expression. It yields a score between 0 and 100,

329 with higher scores indicating greater recognition of human emotional facial and vocal

330   expressions. The CAM battery requires a high level of English language proficiency, hence we

331   limited testing to participants who reported English as a first language. Experienced participants

332   did not complete the CAM battery, due to their job-related time constraints. Finally, after

333   completing all other procedures, experienced participants provided written answers to qualitative

334   questions about their knowledge of rhesus macaque vocal communication and their thought

335   process during the task.

336

337   **Acoustic Analyses**

338   Measurements of mean F0, min F0, max F0, mean harmonic-noise ratio (HNR, with low

339   values indicating a noisier call), and duration were obtained in Praat 6.0.29 (Boersma &

340   Weenink, 2013). Spectrograms were generated using fast-Fourier transform with a Gaussian

341   window shape and 2-ms DFT size. Each vocalization was highlighted manually, looking at the

342   spectrogram and waveform and listening to the vocalization, to obtain its duration while

343   excluding any reverberation. F0 measurements were then obtained from the highlighted portion

344   using the Quantify Source command in the GSU Praat Tools package Version 1.9 (Owren,

345   2008). The default settings for this command were used, with the exception that the pitch ceiling

346   was set to 3000 Hz for coos and 10000 Hz for screams to account for the F0 range of these

347   vocalizations. The command uses Praat's To Pitch autocorrelation function to estimate a F0

348   contour, which the user then inspects and can manually correct if necessary (e.g., octave

349   correction, removal of any unvoiced segments) (Owren, 2008). F0 range was calculated as the

350   difference between max and min F0. Mean ± SD acoustic parameter values were as follows:

351   **coos**: mean F0 = 995 ± 169 Hz, F0 range = 340 ± 214 Hz, mean HNR = 13.5 ± 4.32 dB, duration

352   = 0.580 ± 0.145 s; **screams**: mean F0 = 4879 ± 1447 Hz, F0 range = 3533 ± 2029 Hz, mean

353 HNR = 0.78 ± 3.31 Hz, duration = 0.504 ± 0.211 s. For each stimulus pair, the relative difference

354 in mean F0, F0 range, mean HNR, and duration between the two vocalizations was calculated as

355 the value for the second vocalization minus the value for the first vocalization, divided by the

356 value for the first vocalization; similarly, the relative difference between the high- and low-

357 arousal vocalization was calculated as the value for the high-arousal vocalization minus the value

358 for the low-arousal vocalization, divided by the latter. Three screams were too noisy to estimate

359 the F0 contour, resulting in four scream stimulus pairs missing relative differences on acoustic

360 parameters other than duration. Responses to these stimulus pairs were excluded from the

361 statistical analyses of the effects of acoustics.

362

363 **Statistical Analyses**

364 All statistical analyses were conducted in R Statistical Environment (R Core Team, 2018)

365 and used an alpha of 0.05. To assess the overall accuracy of participants' responses, each

366 participant was attributed one Accuracy Score per stimulus type (coos, screams, mixed), equal to

367 proportion of stimulus pairs to which the participant responded correctly. For mixed stimulus

368 pairs, selecting the scream was arbitrarily treated as "correct" for the purposes of analysis (but

369 not in interpretation, since arousal was measured using different methods for the two call types,

370 thus one call type cannot be considered objectively higher-arousal than the other). We used one-

371 sample t-tests (t.test command in R) to determine whether the two participant groups—

372 experienced and inexperienced—exhibited significantly greater Accuracy Scores than the 50%

373 accuracy predicted by chance, for each of the three stimulus types. These tests were one-tailed

374 for coo stimuli and scream stimuli (to determine whether participant accuracy was greater than

375   predicted by chance), and two-tailed for mixed stimuli (to determine whether participants

376   selected one call type significantly more often than the other).

377   To determine the effects of experience working professionally with rhesus macaques on

378   responses, we ran three binomial, logit-link generalized linear mixed models (GLMM)—one for

379   each stimulus type. The outcome variable was whether or not an individual response was correct,

380   and the binomial experience variable (whether the participant had professional work experience

381   with rhesus macaques) was entered as a fixed effect. To account for non-independence of

382   observations among responses from a single participant, or to a single stimulus pair, participant

383   and stimulus pair were entered as crossed random intercept effects. GLMM were likewise used

384   to determine the effects of other participant characteristics on responses; this analysis was

385   limited to the inexperienced participants' responses since the experienced participants did not

386   complete the surveys or the CAM battery. We again ran one GLMM per stimulus type. The

387   outcome variable for each model and random effects structure were the same; fixed effects

388   included Empathy Quotient, Animal Attitudes Score, CAM Score, and whether a participant had

389   worked with any animal species (other than rhesus macaques) in a professional capacity. Finally,

390   four more GLMM were used to assess the effects of acoustic variables on participant

391   responses—one per participant group for coo and scream stimuli (not mixed stimuli). The

392   random effects structure was the same as the other models; the outcome variable was whether the

393   participant pressed "1" or "2", irrespective of which response was correct. Relative differences in

394   mean F0, F0 range, mean HNR, and duration between the two vocalizations in a stimulus pair

395   were entered as fixed effects. To avoid overfitting, nonsignificant fixed effects were removed

396   from each model. All GLMM were fit by maximum likelihood (Laplace Approximation) using

397   the glmer function in the lme4 package in R (Bates et al., 2015). Expected correlations of fixed

398   effect coefficients were checked to avoid multicollinearity. Each nonsignificant fixed effect was

399   individually re-entered into the model to assess its level of non-significance (Barr et al., 2013)

400   and check against Type II errors.

401

402   **Results**

403   One-sample t-tests indicated that Accuracy Scores (percentage correct) of inexperienced

404   as well as experienced participants were significantly higher than predicted by chance (50%) for

405   coo stimuli (**inexperienced:** mean = 57.8, 95% CI > 56.7, t = 12.98, p < 0.0001; **experienced:**

406   mean = 54.5, 95% CI > 50.8, t = 2.19, p = 0.025); the same was true for scream stimuli

407   (**inexperienced:** mean = 52.2, 95% CI > 50.8, t = 2.53, p = 0.006; **experienced:** mean = 59.4,

408   95% CI > 55.4, t = 4.26, p < 0.001) (Fig. 1). GLMM indicated that experienced participants

409   responded correctly significantly more often than inexperienced participants on scream stimuli

410   (Table 1, Fig. 1). Overall, inexperienced participants did not show a significant tendency to

411   select coos over screams or vice versa in response to mixed stimuli (mean = 49.5, 95% CI[46.1,

412   52.9], t = -0.289, p = 0.773), in contrast to experienced participants, who showed a strong bias

413   towards selecting screams as more emotionally intense (mean = 81.8, 95% CI[68.3, 95.4], t =

414   5.16, p < 0.001) (Fig. 1). GLMM indicated that this contrast represented a significant difference

415   between inexperienced and experienced participants (Table 1). In a questionnaire administered

416   after the task, every experienced participant indicated knowledge of the names of the

417   vocalizations they had heard, and 6 reported knowledge of the contexts in which coos and

418   screams occur. 6 experienced participants reported an explicit strategy of selecting screams over

419   coos in mixed trials, and 4 of those 6 indicated that this strategy was because they perceived the

420   agonistic contexts associated with screams to be more emotionally intense than contexts

421 associated with coos. Finally, among inexperienced participants, those who had professional

422 work experience with a non-primate animal responded significantly less accurately to coo stimuli

423 than those who had not worked with any animals (Table 1); we did not find any other significant

424 effects of participant characteristics on responses.

425   On coo trials, inexperienced and experienced participants alike most often selected as

426 more emotionally intense the coo that was higher in mean F0, lower in HNR (i.e., noisier), or

427 longer in duration, while inexperienced participants also showed a slight tendency to select the

428 coo showing a wider F0 range (Table 2; Fig. 2). Congruent with this, the higher-arousal coo in a

429 pair was, on average, higher in mean F0 and F0 range, and lower in HNR, though some of these

430 average relative differences were minute and relative differences among all parameters ranged

431 from negative to positive (Table 2). Plotting the coo pairs by relative mean F0 difference and

432 average participant response (Fig. 2a) shows that, for most (though by no means all) of the pairs

433 in which the second coo represented greater arousal than the first (orange/light points), the

434 second coo was also higher in F0 (i.e., relative F0 difference > 0), and was more often selected

435 by participants (i.e., proportion of '2' responses > 0.5). A strong positive correlation is also

436 visible between relative duration difference and proportion of '2' responses, but this does not

437 appear to have translated into correct responses; if anything, it seems that in the majority of coo

438 pairs, the coo representing greater arousal was shorter in duration (i.e., more orange/light points

439 falling at relative duration difference < 0) (Fig. 2a), though the average duration difference

440 between high- and low-arousal coos was close to zero (Table 2).

441   On scream trials, inexperienced participants showed a slight tendency to select the

442 scream exhibiting a wider F0 range, while experienced participants tended to select the scream

443 exhibiting a higher mean F0 and longer duration (Table 2; Fig. 2b). Including mean F0 or

444   duration individually in the experienced participant GLMM showed significant effects of each

445   (Table 2), however, including both predictors in a single model caused one predictor to become

446   nonsignificant and yielded a coefficient of correlation of fixed effects = -0.746, indicating

447   collinearity. We therefore tested the correlation between relative differences in scream F0 and

448   duration across stimulus pairs, and found a significant positive correlation (scream pairs with a

449   larger difference in F0 tended to have a larger difference in duration; Pearson's r = 0.78, t = 7.18,

450   p < 0.001), suggesting that the effects of F0 and duration on scream perception here are best

451   interpreted together. As with responses to coos, each of these effects of acoustic differences on

452   responses to scream stimuli was in the direction consistent with the mean relative acoustic

453   difference between high- and low-arousal screams (Table 2). Examining Fig. 2b, it appears that

454   the tendency shown by experienced participants to select the higher-F0 and/or longer scream led

455   them to respond accurately more often than not (i.e., more green/dark points in the lower-left

456   corner and more orange/light points in the upper-right corner).

457

458   **Discussion**

459   The hypothesis of evolutionary homology in emotional vocal communication predicts

460   that humans should be able to accurately perceive the emotional significance of heterospecific

461   vocalizations—specifically, to discern this based on acoustic variation within (not only between)

462   call types. We tested this prediction by having participants listen to rhesus macaque coos and

463   screams, and choose the vocalization they perceived to reflect greater emotional intensity.

464   Inexperienced and experienced participants alike exhibited above-chance accuracy, relative to

465   differences in emotional arousal between coos, as measured by increases in stress hormone

466   levels, and between screams, as inferred from differences in the intensity of aggression received

467   by vocalizers. Overall, these findings support the hypothesis that essential aspects of emotional

468   vocal communication are shared between humans and rhesus macaques.

469

470   **Accuracy and Experience**

471   Overall accuracy was modest (mean accuracy less than 60%), though comparable with

472   the findings of similar studies (Filippi et al., 2017a,b). That said, it is important to note the fine-

473   grained nature of the discriminations we asked participants to perform. Whereas many similar

474   studies have sampled vocalizations from relatively disparate contexts, e.g., feeding vs.

475   aggression (Faragó et al., 2017) or isolation vs. reunion vs. nursing vs. veterinary procedures

476   (Maruščáková et al., 2015), our coos were all sampled from an identical controlled experimental

477   behavioral test; the only factors differentiating the contexts from which high- and low-arousal

478   coos were sampled were the characteristics and background of an individual macaque, shaping

479   its subjective experience of the test, as reflected in stress hormone levels. Insofar as separation

480   from the mother and the rest of the group (albeit for only 30 minutes) probably evoked a high-

481   arousal, negative emotional state in every six-month-old macaque, the emotional difference

482   between coos within a pair was probably relatively small. The same reasoning applies to

483   screams; almost every rhesus macaque scream occurs in a context associated with a relatively

484   high-arousal, negative emotional state. Thus, while we operationally labelled vocalizations

485   "high- and low-arousal," these labels are relative; it might in fact be more accurate to think of the

486   arousal levels represented among our stimuli as ranging from "moderately high" to "very high."

487   In light of that, our participants' accuracy was significantly above chance not only in a statistical

488   but also in a biological sense. One explanation for these findings is that participants performed

489   an anthropomorphic extrapolation from human emotional communication to that of rhesus

490   macaques, perceiving emotion from vocalizations as though they were produced by humans.

491   Such anthropomorphism could have resulted in accurate responses more often than the reverse,

492   thanks in part to any similarities in the acoustic indicators of arousal stemming from humans'

493   and macaques' shared evolutionary ancestry (discussed further below). Participants could also

494   have drawn on experience with the communication of familiar nonhuman species, though

495   surprisingly, having formally worked with non-primate animals was associated with significantly

496   decreased accuracy on coo stimuli. Participant characteristics other than experience—empathy,

497   human emotion recognition, and animal attitudes—failed to show any relationship with

498   responses, in line with one previous study (Maruščáková et al., 2015).

499   While even inexperienced participants performed with above-chance accuracy, we did

500   find some effects of experience with rhesus macaques, supporting the hypothesis that familiarity

501   with a species can improve sensitivity to emotional communication by that species (Nicastro &

502   Owren, 2003; Tallet et al., 2010; Parsons et al., 2019). Experienced participants exhibited

503   slightly (but statistically significantly) greater accuracy with scream stimulus pairs, and far more

504   often chose the scream over the coo in mixed stimulus pairs, though experience had no effect on

505   accuracy with coo stimulus pairs. One possibility is that humans might typically have more

506   experience perceiving cues to emotion in vocalizations exhibiting the F0 of coos (generally 400-

507   3000 Hz) than that of rhesus macaque screams (generally 2000-10000 Hz), since the former is

508   closer to the human vocal range. Previous research suggests listeners of a given species may be

509   more sensitive to emotional vocalizations with a F0 falling within that species' typical range

510   (Root-Gutteridge et al., 2021). Emotion perception from screams might therefore have been less

511   intuitive for inexperienced human listeners, making experience a more significant factor. That

512   said, tonal screams comprise only one of several rhesus macaque scream classes, warranting

513   caution in extrapolating to screams more generally.

514   Interpreting the effect of experience on responses to mixed stimuli (containing one coo

515   and one scream)—in which experienced participants much more often indicated the scream as

516   reflecting a more intense emotion, while inexperienced participants did not—is more

517   straightforward: several participants reported prior knowledge of the socioecological contexts

518   associated with coos and screams, and indicated that they drew on this knowledge to judge

519   screams as more emotionally intense than coos. Inexperienced participants, in contrast,

520   presumably lacked such a basis of prior knowledge on which to form a consistent explicit

521   strategy, and therefore responded in a much less consistent manner. Similarly, human listeners in

522   a recent study did not provide significantly different ratings of emotional valence (positive vs.

523   negative states) to rhesus macaque feeding calls vs. screams (Fritz et al., 2018). The effect of

524   experience on responses to mixed stimuli was by far the strongest effect of any participant

525   characteristic; its effects on accuracy with scream pairs was modest by comparison.

526   Experienced participants' knowledge of the contexts in which macaque screams occur

527   might also have played a role in their greater accuracy on scream pairs (judging which of two

528   screams was associated with greater arousal). The acoustic structure of screams is generally

529   conserved across a phylogenetically diverse array of taxa, yet unlike nonhuman primates which

530   produce screams primarily during aggression, human screams occur in a wide variety of

531   emotional contexts, including positive ones, e.g., intense pleasure-joy-elation, sadness, pain,

532   anger-rage, and fear (Engelberg, Schwartz & Gouzoules, 2021; Frühholz et al., 2021). To

533   speculate, these species differences could have influenced participants' responses, if perception

534   of the macaque screams drew on processes similar to those involved in conspecific scream

535   perception. Experienced participants, most of whom reported knowing that macaque screams

536   occur primarily during aggression, might have avoided any resulting biases to a greater extent,

537   thus responding more accurately on scream trials. The question of how heterospecific scream

538   perception is influenced by humans' expanded scream usage seems an intriguing area for further

539   study.

540

541   **Acoustic Cues to Emotional Arousal**

542   Vocal F0 is a well-established correlate of emotional arousal across mammals including

543   humans, and humans have been shown to perceive higher-F0 vocalizations from a variety of

544   species (including conspecific vocalizations) as more emotionally intense (Scherer, 2003; Juslin

545   & Laukka, 2003; Nicastro & Owren, 2003; Pongrácz et al., 2005; Sauter et al., 2010; Briefer,

546   2012; Faragó et al., 2014, 2017; Maruščáková et al., 2015; Filippi et al., 2017a; Kelly et al.,

547   2017; Schwartz & Gouzoules, 2019). This trend is probably due in part to phylogenetically

548   widespread mechanisms of arousal and vocal production: arousal increases tension in the vocal

549   fold muscles, resulting in faster oscillation, i.e., higher F0 (Scherer, 1986; Titze, 1994; Riede,

550   2010; Briefer, 2012). In the present study, inexperienced and experienced participants alike

551   tended to indicate the higher-F0 coo in a pair as more emotionally intense. Experienced, but not

552   inexperienced, participants appeared to base scream selections partly on F0 as well, perhaps

553   contributing to their greater accuracy. These findings are consistent not only with the literature in

554   general but also with the findings of a recent study on rhesus macaque vocal production, which

555   demonstrated a positive correlation between F0 and arousal among coos as well as screams

556   (Schwartz, Sanchez & Gouzoules, 2022). Thus, vocal F0 appears to be a significant indicator of

557   emotional arousal in humans and rhesus macaques (and other mammalian species) alike,

558   probably due to evolutionary homologies in the physiology and anatomy of vocal production. In

559   the present study as in previous ones (Kelly et al., 2017), humans' sensitivity to F0 seems to have

560   led participants to respond correctly in some cases and incorrectly in other cases, but the former

561   appears to have been more common than the latter in the present study (Fig. 2).

562   Given the significant role of F0 in within-call-type discriminations, it is puzzling that

563   inexperienced participants did not consistently indicate high-F0 screams as reflecting a more

564   intense emotion than lower-F0 coos. Likewise, participants in a recent study did not perceive

565   similar call types by rhesus macaques as differing in emotional valence (Fritz et al., 2018). Just

566   as F0 and other acoustic parameters seem to correlate more strongly with arousal within a call

567   type than between call types from a vocal production standpoint (Schwartz, Engelberg &

568   Gouzoules, 2020), our participants appeared to rely on F0 as a cue to emotional intensity more

569   for within- than between-type discrimination. This observation is consistent with the

570   "identification-attribution model" of nonverbal vocal perception, which postulates that

571   classification or identification of call types is cognitively and perhaps neurally distinct from the

572   attribution of emotional states to vocalizations (Spunt & Lieberman, 2012; Sperduti et al., 2014;

573   Anikin, Bååth & Persson, 2018).

574   In addition to F0, inexperienced and experienced participants both tended to select a coo

575   when it was longer in duration or noisier (lower HNR) than the alternative, consistent with

576   previous studies human emotion perception (Faragó et al., 2014; Filippi et al., 2017b; Schwartz

577   & Gouzoules, 2019; Anikin et al., 2021; Massenet et al., 2022). A recent study suggested a

578   negative correlation between arousal and HNR in coos (Schwartz, Sanchez & Gouzoules, 2022),

579   and among coo stimulus pairs used in the present study, the higher-arousal coo was noisier on

580   average (Table 2). Thus, the effects of noisiness on participants' decisions might thus have led

581 them to respond correctly on some trials. It should be noted, however, that perception of

582 noisiness is not fully distinct from that of F0; chaos and other nonlinear phenomena can affect

583 perceived pitch (Anikin et al., 2021). Participants' tendency to select the longer coo likely played

584 no role in their overall accuracy, instead leading them to the correct and incorrect responses an

585 approximately equal proportion of the time (Fig. 2a). Experienced participants also tended to

586 select the longer scream, though the effects of scream F0 and duration on responses appeared to

587 be collinear, making them difficult to disambiguate. Finally, inexperienced participants showed a

588 preference for the coo or scream in a pair that showed a wider F0 range, though estimates of

589 these effects were small.

590   Overall, consistencies between our findings and those of other studies suggest that human

591 listeners might use similar acoustic cues to assess the emotional arousal represented by acoustic

592 variants within a call type, whether it be a human nonverbal vocalization (Sauter et al., 2010;

593 Schwartz & Gouzoules, 2019), a vocalization of a species with which the listener is familiar

594 (Nicastro & Owren, 2003; Faragó et al., 2014, 2017; Massenet et al., 2022), or a vocalization of

595 an unfamiliar species (Filippi et al., 2017a; Kelly et al., 2017). These perceptual rules of thumb

596 (e.g., a higher-F0 sound indicates higher arousal) seem to lead to accurate perceptions more often

597 than not, thanks in part to broad similarities in the acoustic indicators of arousal owing to shared

598 evolutionary ancestry. However, the utility of these rules of thumb in making accurate

599 judgments, and the role of familiarity and experience in determining the salience of different

600 acoustic cues, might vary across species and call types.

601

602 **Emotional Arousal and Valence**

603   A remaining issue is the distinction between emotional arousal and valence, i.e., the

604   spectrum from negative to positive states (Russell, 2003; Mendl, Burman & Paul, 2010; Briefer,

605   2012). Participants were not instructed on the distinction between emotional valence and arousal,

606   and thus it is possible that some of them might have interpreted "a more intense emotion" to

607   mean a state that is more negative as well as more aroused. However, the term "intensity" is

608   often used interchangeably with "arousal," and previous research has shown that listeners

609   provide similar ratings for both terms with respect to vocal stimuli (Laukka, Juslin & Bresin,

610   2005). In truth, the stimulus pairs might have represented differences in emotional valence in

611   addition to arousal: higher concentrations of stress hormones might reflect a more negative state

612   than lower concentrations, and contact aggression might evoke a more negative emotional

613   reaction than non-contact aggression. As in many other studies of animal emotional vocal

614   communication (reviewed in Briefer, 2012), the present one does not permit us to discern the

615   precise roles of emotional arousal and valence. With that said, the mechanisms by which

616   emotional arousal affects the voice are more straightforward than for valence (Briefer, 2012),

617   and F0—a well-established correlate of emotional arousal but not valence—strongly influenced

618   our participants' responses. For these reasons, it seems most parsimonious to conclude that

619   emotional arousal played a role in our findings.

620

621   **Conclusions**

622   The present study demonstrated that listeners with and without job-related experience

623   with rhesus macaques have the capacity to discern fine-grained distinctions in the emotional

624   arousal level represented in the coos and screams of that species. Experience was associated with

625   limited improvement in this ability, but the most obvious effect of experience was seen in

626 between-vocalization-type discriminations, where semantic knowledge about the socioecology of

627 coos and screams appeared to influence responses. Thus our findings are consistent with the

628 hypothesis that humans' capacity for accurate emotional perception of rhesus macaque

629 vocalizations stems in part from the evolutionary ancestry shared by the two species.

630 Specifically, it appears that listeners utilized acoustic cues that reliably indicate emotional

631 arousal in humans and rhesus macaques alike due to homology in the physiology and anatomy of

632 arousal and vocal production. Some nonhuman calls are directly evolutionarily homologous with

633 human nonverbal vocalizations, screams being one example. A remaining question is how

634 humans' usage of such vocalizations, as in the contextual expansion seen in human screams,

635 influences perceptions of heterospecific calls. Future research examining emotion expression

636 within a variety of call types of a phylogenetically diverse array of taxa, and human perception

637 thereof, will further improve our understanding of the evolutionary history of emotional

638 communication.

639

640   **Acknowledgments**

641 We thank Drs. Mar M. Sanchez and Mark E. Wilson for the design and supervision of the human

642 intruder and cortisol testing used to generate the coo stimuli. Thanks also to Natalie Brutto, Kelly

643 Bailey, Manuel Bautista, Patrice Rando, Jalani Paul, Jonathan Engelberg, Jodi Godfrey, and

644 Desiree De Leon for their assistance in collection of those human intruder and cortisol data and

645 coo recordings, and Emma Satty for her assistance collecting participant responses. Finally, we

646 thank Dr. Andrey Anikin and one anonymous reviewer for their feedback and suggestions for

647 revisions to this manuscript.

648

Funding

650 This study was supported by the U.S. National Science Foundation Graduate Research

651 Fellowship under Grant No. DGE – 1343012, by the U.S. National Institutes of Health (NIH)

652 grant number 1R01HD077623 and the NIH Office of the Director, Office of Research

653 Infrastructure Programs, P51OD011132 (Emory National Primate Research Center -ENPRC-

654 base grant). The ENPRC is fully accredited by AAALAC, International.

655

656 **References**

657 Anikin A, Bååth R, Persson T. 2018. Human non-linguistic vocal repertoire: Call types and their

658   meaning. *Journal of Nonverbal Behavior* 49:53–80. DOI: 10.1007/s10919-017-0267-y.

659 Anikin A, Pisanski K, Massenet M, Reby D. 2021. Harsh is large: Nonlinear vocal phenomena

660   lower voice pitch and exaggerate body size. *Proceedings of the Royal Society B*

661   288:20210872. DOI: 10.1098/RSPB.2021.0872.

662 Baron-Cohen S, Wheelwright S. 2004. The empathy quotient: An investigation of adults with

663   asperger syndrome or high functioning autism, and normal sex differences. *Journal of*

664   *Autism and Developmental Disorders* 34:163–175. DOI:

665   10.1023/B:JADD.0000022607.19833.00.

666 Barr DJ, Levy R, Scheepers C, Tily HJ. 2013. Random effects structure for confirmatory

667   hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278. DOI:

668   10.1016/j.jml.2012.11.001.

669 Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4.

670   *Journal of Statistical Software* 67:1–48. DOI: 10.18637/jss.v067.i01.

671 Bayart F, Hayashi KT, Faull KF, Barchas JD, Levine S. 1990. Influence of maternal proximity

672   on behavioral and physiological responses to separation in infant rhesus monkeys (*Macaca*

673   *mulatta*). *Behavioral Neuroscience* 104:98–107. DOI: 10.1037/0735-7044.104.1.98.

674   Belin P, Fecteau S, Charest I, Nicastro N, Hauser MD, Armony JL. 2008. Human cerebral

675   response to animal affective vocalizations. *Proceedings of the Royal Society B: Biological*

676   *Sciences* 275:473–481. DOI: 10.1098/rspb.2007.1460.

677   Boersma P, Weenink D. 2013.Praat: Doing phonetics by computer (Version 5.3.51). *Available at*

678   *http://www.praat.org/* (accessed January 9, 2018).

679   Briefer EF. 2012. Vocal expression of emotions in mammals: Mechanisms of production and

680   evidence. *Journal of Zoology* 7:1–20. DOI: 10.1098/rsbl.2011.0832.

681   Darwin C. 1872. *The expression of the emotions in man and animals*. London, UK: John Murray.

682   DOI: 10.1037/h0076058.

683   Engelberg JWM, Schwartz JW, Gouzoules H. 2021. The emotional canvas of human screams:

684   Patterns and acoustic cues in the perceptual categorization of a basic call type. *PeerJ*

685   9:e10990. DOI: 10.7717/PEERJ.10990/SUPP-2.

686   Faragó T, Andics A, Devecseri V, Kis A, Gacsi M, Miklosi A. 2014. Humans rely on the same

687   rules to assess emotional valence and intensity in conspecific and dog vocalizations.

688   *Biology Letters* 10:20130926. DOI: 10.1098/rsbl.2013.0926.

689   Faragó T, Takács N, Miklósi Á, Pongrácz P. 2017. Dog growls express various contextual and

690   affective content for human listeners. *Royal Society Open Science* 4:170134. DOI:

691   10.1098/rsos.170134.

692   Filippi P, Congdon J V., Hoang J, Bowling DL, Reber SA, Pašukonis A, Hoeschele M,

693   Ocklenburg S, de Boer B, Sturdy CB, Newen A, Güntürkün O. 2017a. Humans recognize

694   emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for

695   acoustic universals. *Proceedings of the Royal Society B: Biological Sciences* 284:20170990.

696   DOI: 10.1098/rspb.2017.0990.

697 Filippi P, Gogoleva SS, Volodina E V., Volodin IA, de Boer B. 2017b. Humans identify negative

698   (but not positive) arousal in silver fox vocalizations: Implications for the adaptive value of

699   interspecific eavesdropping. *Current Zoology* 63:445–456. DOI: 10.1093/cz/zox035.

700 Fischer J, Wadewitz P, Hammerschmidt K. 2016. Structural variability and communicative

701   complexity in acoustic communication. *Animal Behaviour* 134:229–237. DOI:

702   10.1016/j.anbehav.2016.06.012.

703 Fritz T, Mueller K, Guha A, Gouws A, Levita L, Andrews TJ, Slocombe KE. 2018. Human

704   behavioural discrimination of human, chimpanzee and macaque affective vocalisations is

705   reflected by the neural response in the superior temporal sulcus. *Neuropsychologia*

706   111:145–150. DOI: 10.1016/j.neuropsychologia.2018.01.026.

707 Frühholz S, Dietziker J, Staib M, Trost W. 2021. Neurocognitive processing efficiency for

708   discriminating human non-alarm rather than alarm scream calls. *PLOS Biology*

709   19:e3000751. DOI: 10.1371/JOURNAL.PBIO.3000751.

710 Fugate JMB, Gouzoules H, Nygaard LC. 2008. Recognition of rhesus macaque (*Macaca*

711   *mulatta*) noisy screams: Evidence from conspecifics and human listeners. *American Journal*

712   *of Primatology* 70:594–604. DOI: 10.1002/ajp.20533.

713 Golan O, Baron-Cohen S, Hill J. 2006. The Cambridge Mindreading (CAM) Face-Voice Battery:

714   Testing complex emotion recognition in adults with and without Asperger Syndrome.

715   *Journal of Autism and Developmental Disorders* 36:169–183. DOI: 10.1007/s10803-005-

716   0057-y.

717 Golan O, Baron-Cohen S, Hill JJ, Rutherford MD. 2007. The "Reading the Mind in the Voice"

718   test-revised: A study of complex emotion recognition in adults with and without autism

719   spectrum conditions. *Journal of Autism and Developmental Disorders* 37:1096–1106. DOI:

720   10.1007/s10803-006-0252-5.

721 Gouzoules H. 2005. Monkeying around with symbolism: Are vocalizations simple symbols...or

722   more like cymbals? In: Namy L ed. *Symbol use and symbolic representation:*

723   *Developmental and comparative perspectives.* Mahwah, NJ: Lawrence Erlbaum Associates

724   Publishers, 245–265.

725 Gouzoules S, Gouzoules H, Marler P. 1984. Rhesus monkey (*Macaca mulatta*) screams:

726   Representational signalling in the recruitment of agonistic aid. *Animal Behaviour* 32:182–

727   193. DOI: 10.1016/S0003-3472(84)80336-X.

728 Gouzoules H, Gouzoules S, Tomaszycki M. 1998. Agonistic screams and the classification of

729   dominance relationships: Are monkeys fuzzy logicians? *Animal Behaviour* 55:51–60. DOI:

730   10.1006/anbe.1997.0583.

731 Hauser MD. 1991. Sources of acoustic variation in rhesus macaque (*Macaca mulatta*)

732   vocalizations. *Ethology* 89:29–46. DOI: 10.1111/j.1439-0310.1991.tb00291.x.

733 Hauser MD, Marler P. 1993. Food-associated calls in rhesus macaques (*Macaca mulatta*): I.

734   Socioecological factors. *Behavioral Ecology* 4:194–205. DOI: 10.1093/beheco/4.3.194.

735 Herzog H, Grayson S, McCord D. 2015. Brief measures of the animal attitude scale. *Anthrozoos*

736   28:145–152. DOI: 10.2752/089279315X14129350721894.

737 Juslin PN, Laukka P. 2003. Communication of emotions in vocal expression and music

738   performance: Different channels, same code? *Psychological Bulletin* 129:770–814. DOI:

739   10.1037/0033-2909.129.5.770.

740 Kalin NH, Shelton SE. 1989. Defensive behaviors in infant rhesus monkeys: Environmental cues

741   and neurochemical regulation. *Science* 243:1718–21. DOI: 10.1126/science.2564702.

742   Kamiloğlu RG, Slocombe KE, Haun DBM, Sauter DA. 2020. Human listeners' perception of

743   behavioural context and core affect dimensions in chimpanzee vocalizations. *Proceedings*

744   *of the Royal Society B* 287. DOI: 10.1098/RSPB.2020.1148.

745   Kelly T, Reby D, Levréro F, Keenan S, Gustafsson E, Koutseff A, Mathevon N. 2017. Adult

746   human perception of distress in the cries of bonobo, chimpanzee, and human infants.

747   *Biological Journal of the Linnean Society* 120:919–930. DOI: 10.1093/biolinnean/blw016.

748   Laukka P, Juslin PN, Bresin R. 2005. A dimensional approach to vocal expression of emotion.

749   *Cognition and Emotion* 19:633–653. DOI: 10.1080/02699930441000445.

750   Linnankoski I, Laakso M, Aulanko R, Leinonen L. 1994. Recognition of emotions in macaque

751   vocalizations by children and adults. *Language and Communication* 14:183–192. DOI:

752   10.1016/0271-5309(94)90012-4.

753   Maigrot AL, Hillmann E, Briefer EF. 2022. Cross-species discrimination of vocal expression of

754   emotional valence by Equidae and Suidae. *BMC Biology* 20:1–14. DOI: 10.1186/S12915-

755   022-01311-5/TABLES/5.

756   Maruščáková IL, Linhart P, Ratcliffe VF, Tallet C, Reby D, Špinka M. 2015. Humans (*Homo*

757   *sapiens*) judge the emotional content of piglet (*Sus scrofa domestica*) calls based on simple

758   acoustic parameters, not personality, empathy, nor attitude toward animals. *Journal of*

759   *Comparative Psychology* 129:121–131. DOI: 10.1037/a0038870.

760   Massenet M, Anikin A, Pisanski K, Reynaud K, Mathevon N, Reby D. 2022. Nonlinear vocal

761   phenomena affect human perceptions of distress, size and dominance in puppy whines.

762   *Proceedings of the Royal Society B* 289. DOI: 10.1098/RSPB.2022.0429.

763   McComb K, Taylor AM, Wilson C, Charlton BD. 2009. The cry embedded within the purr.

764   *Current Biology* 19:507–508. DOI: 10.1016/j.cub.2009.05.033.

765   Mendl M, Burman OHP, Paul ES. 2010. An integrative and functional framework for the study

766   of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences*

767   277:2895–2904. DOI: 10.1098/rspb.2010.0303.

768   Merkies K, Crouchman E, Belliveau H. 2021. Human ability to determine affective states in

769   domestic horse whinnies. *https://doi.org/10.1080/08927936.2021.1999605* 35:483–494.

770   DOI: 10.1080/08927936.2021.1999605.

771   Nicastro N, Owren MJ. 2003. Classification of domestic cat (*Felis catus*) vocalizations by naive

772   and experienced human listeners. *Journal of Comparative Psychology* 117:44–52. DOI:

773   10.1037/0735-7036.117.1.44.

774   Owren MJ. 2008. GSU Praat tools: Scripts for modifying and analyzing sounds using Praat

775   acoustics software. *Behavior Research Methods* 40:822–829. DOI: 10.3758/BRM.40.3.822.

776   Owren MJ, Bachorowski J-A. 2007. Measuring emotion-related vocal acoustics. In: Coan J,

777   Allen J eds. *The handbook of emotion elicitation and assessment*. Oxford: Oxford

778   University Press, 239–266.

779   Parsons CE, LeBeau RT, Kringelbach ML, Young KS. 2019. Pawsitively sad: Pet-owners are

780   more sensitive to negative emotion in animal distress vocalizations. *Royal Society Open*

781   *Science* 6:181555. DOI: 10.1098/rsos.181555.

782   Pisanski K, Bryant GA, Cornec C, Anikin A, Reby D. 2022. Form follows function in human

783   nonverbal vocalisations. *Ethology Ecology & Evolution* 34:303–321. DOI:

784   10.1080/03949370.2022.2026482.

785   Pongrácz P, Molnár C, Miklósi Á. 2010. Barking in family dogs: An ethological approach. *The*

786   *Veterinary Journal* 183:141–147. DOI: 10.1016/J.TVJL.2008.12.010.

787 Pongrácz P, Molnár C, Miklósi Á, Csányi V. 2005. Human listeners are able to classify dog

788   (*Canis familiaris*) barks recorded in different situations. *Journal of Comparative*

789   *Psychology* 119:136–144. DOI: 10.1037/0735-7036.119.2.136.

790 R Core Team. 2018.R: A language and environment for statistical computing. R Foundation for

791   Statistical Computing, Vienna, Austria. *Available at http://www.r-project.org/*

792 Riede T. 2010. Elasticity and stress relaxation of rhesus monkey (*Macaca mulatta*) vocal folds.

793   *Journal of Experimental Biology* 213:2924–2932. DOI: 10.1242/jeb.044404.

794 Root-Gutteridge H, Ratcliffe VF, Neumann J, Timarchi L, Yeung C, Korzeniowska AT,

795   Mathevon N, Reby D. 2021. Effect of pitch range on dogs' response to conspecific vs.

796   heterospecific distress cries. *Scientific Reports 2021 11:1* 11:1–10. DOI: 10.1038/s41598-

797   021-98967-w.

798 Russell JA. 2003. Core affect annd the psychological construction of emotion. *Psychological*

799   *Review* 110:145–172. DOI: 10.1037/0033-295X.110.1.145.

800 Sauter DA, Eisner F, Calder AJ, Scott SK. 2010. Perceptual cues in nonverbal vocal expressions

801   of emotion. *Quarterly Journal of Experimental Psychology* 63:2251–2272. DOI:

802   10.1080/17470211003721642.

803 Scherer KR. 1986. Vocal affect expression: A review and a model for future research.

804   *Psychological Bulletin* 99:143–165. DOI: 10.1037/0033-2909.99.2.143.

805 Scherer KR. 2003. Vocal communication of emotion: A review of research paradigms. *Speech*

806   *Communication* 40:227–256. DOI: 10.1016/S0167-6393(02)00084-5.

807 Scheumann M, Hasting AS, Kotz SA, Zimmermann E. 2014. The voice of emotion across

808   species: How do human listeners recognize animals' affective states? *PLoS ONE* 9:e91192.

809   DOI: 10.1371/journal.pone.0091192.

810 Schwartz JW, Engelberg JWM, Gouzoules H. 2020. Evolving views on cognition in animal

811   vocal communication: Contributions from scream research. *Animal Behavior and Cognition*

812   7:192–213. DOI: 10.26451/abc.07.02.12.2020.

813 Schwartz JW, Gouzoules H. 2019. Decoding human screams: Perception of emotional arousal

814   from pitch and duration. *Behaviour* 156:1283–1307. DOI: 10.1163/1568539X-00003566.

815 Schwartz JW, Sanchez MM, Gouzoules H. 2022. Vocal expression of emotional arousal across

816   two call types in young rhesus macaques. *Animal Behaviour* 190:125–138. DOI:

817   10.1016/j.anbehav.2022.05.017.

818 Sperduti M, Guionnet S, Fossati P, Nadel J. 2014. Mirror Neuron System and Mentalizing

819   System connect during online social interaction. *Cognitive Processing* 15:307–316. DOI:

820   10.1007/s10339-014-0600-x.

821 Spunt RP, Lieberman MD. 2012. An integrative model of the neural systems supporting the

822   comprehension of observed emotional behavior. *NeuroImage* 59:3050–3059. DOI:

823   10.1016/j.neuroimage.2011.10.005.

824 Steiper ME, Young NM. 2008. Timing primate evolution: Lessons from the discordance between

825   molecular and paleontological estimates. *Evolutionary Anthropology: Issues, News, and*

826   *Reviews* 17:179–188. DOI: 10.1002/evan.20177.

827 Tallet C, Špinka M, Maruščáková I, Šimeček P. 2010. Human perception of vocalizations of

828   domestic piglets and modulation by experience with domestic pigs (*Sus scrofa*). *Journal of*

829   *Comparative Psychology* 124:81–91. DOI: 10.1037/a0017354.

830 Titze IR. 1994. *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

831 Zimmermann E, Leliveld LMC, Schehka S. 2013. Toward the evolutionary roots of affective

832   prosody in human acoustic communication: A comparative approach to mammalian voices.

833   In: Altenmüller E, Schmidt S, Zimmermann E eds. *Evolution of emotional communication:*

834   *From sounds in nonhuman mammals to speech and music in men*. Oxford: Oxford

835   University Press, 117–132. DOI: 10.1093/acprof:oso/9780199583560.001.0001.

836

Figure 1

Accuracy Scores.

Calculated for each participant to reflect percent accuracy selecting the vocalization reflecting a more intense emotion, from between two coos, two screams, or one of each. For mixed stimuli, selecting the scream was considered "accurate" for the purposes of analysis and visualization. Error bars within boxes show standard error of the mean . * p < 0.05, ** p < 0.01, *** p < 0.001.

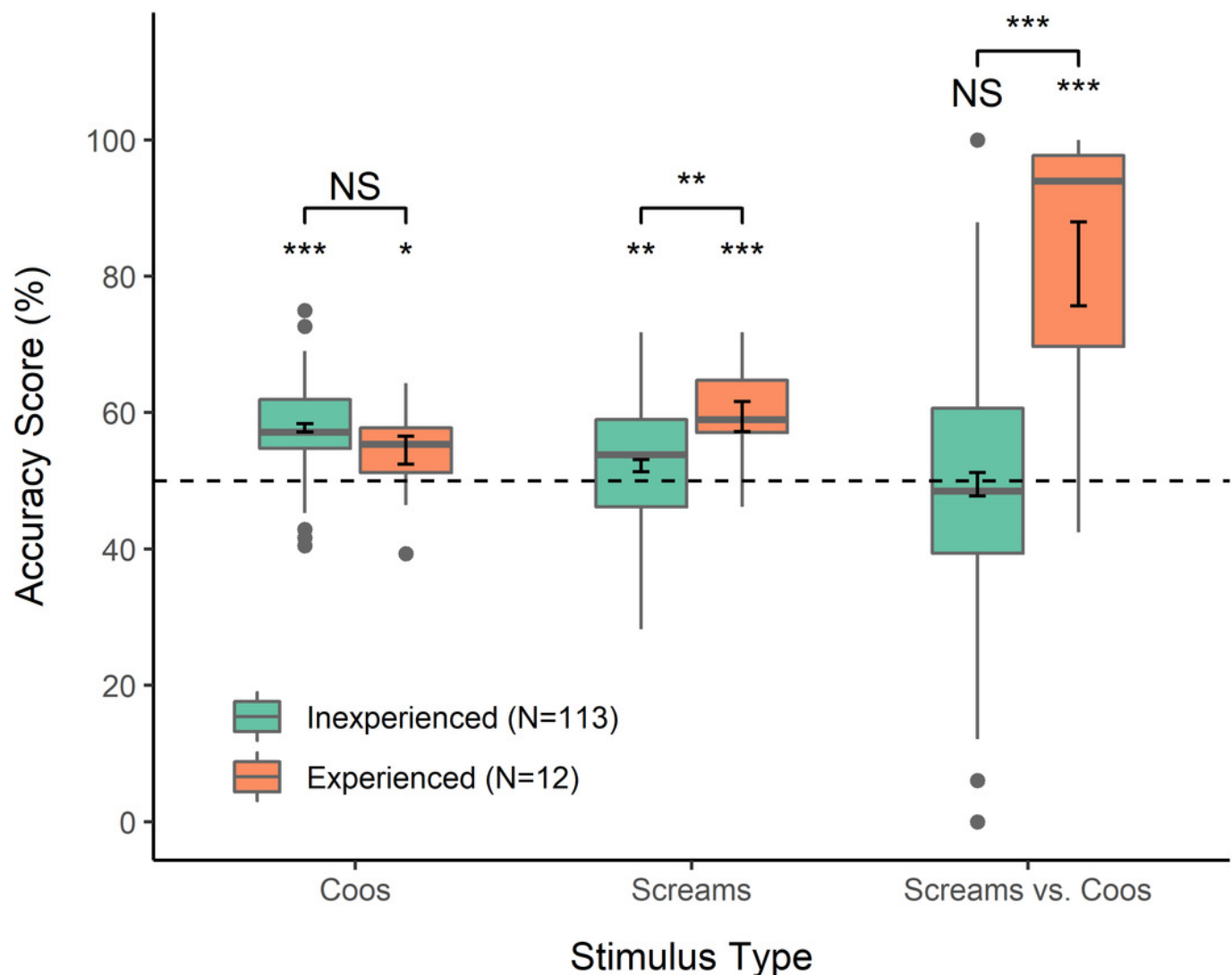
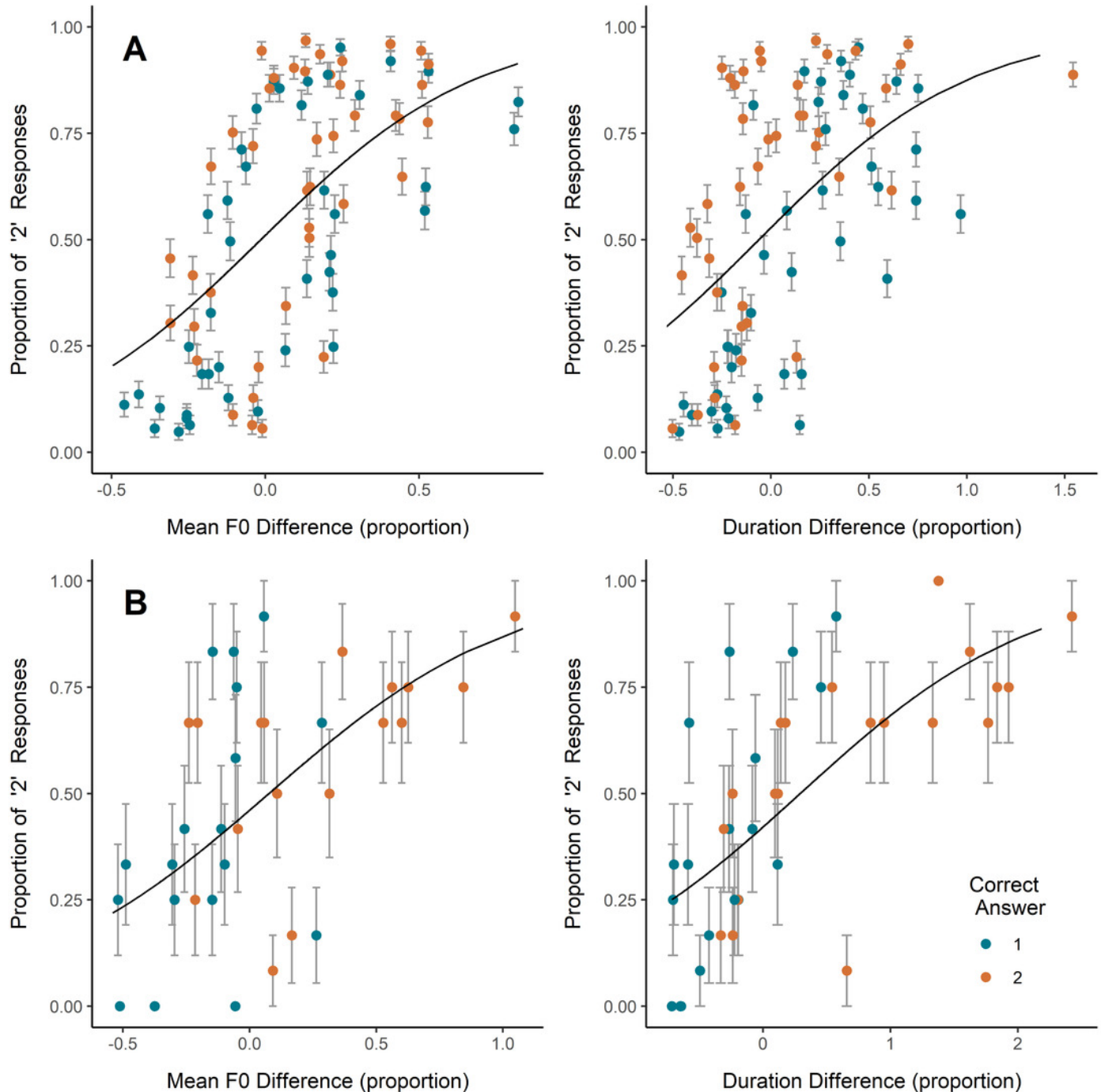



Figure 2

Effects of F0 and duration on perceptions.

Relative increase or decrease from the first to the second vocalization in (A) coo trials (all participants) and (B) scream trials (experienced participants only since no significant effect was observed in inexperienced participants), plotted against the proportion of participants who indicated the second vocalization as reflecting a more intense emotion. Each point represents one stimulus pair; color represents whether the first or second vocalization in the pair represented greater arousal. Error bars represent SE. Trendlines reflect model predictions.

Effects of participant characteristics.

Model estimates of effects on odds of correct responses or, for mixed stimuli, odds of selecting a scream over a coo. Characteristics other than experience with rhesus macaques were tested only among participants inexperienced with primates.

Coo Stimuli	Est. coeff.	SE	p value
Experience with rhesus macaques	-0.213	0.125	0.089
Empathy Quotient	0.001	0.093	0.813
Animal Attitudes Score	0.007	0.008	0.325
CAM Score	-0.001	0.004	0.881
Experience with non-primates	-0.226	0.093	0.015
Scream Stimuli			
Experience with rhesus macaques	0.368	0.141	0.009
Empathy Quotient	0.003	0.004	0.503
Animal Attitudes Score	0.002	0.008	0.798
CAM Score	< 0.001	0.004	0.940
Experience with non-primates	0.004	0.110	0.968
Mixed Stimuli			
Experience with rhesus macaques	2.428	0.379	< 0.001
Empathy Quotient	-0.008	0.009	0.409
Animal Attitudes Score	-0.012	0.018	0.514
CAM Score	0.001	0.010	0.937
Experience with non-primates	-0.380	0.241	0.116

Descriptives for and effects of acoustic parameters.

Left column shows relative differences in acoustic parameters between high- and low-arousal vocalizations within stimulus pairs; positive indicates higher-arousal calls had higher mean values. Center and right columns show model estimates of effects of relative change in the acoustic parameter from the first to the second vocalization on the odds of selecting the second vocalization as reflecting a more intense emotion; positive indicates the call with the higher value was more likely to be selected (regardless of whether doing so was correct).

		Inexperienced			Experienced		
Coo Stimuli	Mean ± SD	Est. coeff.	SE	p value	Est. coeff.	SE	p value
Mean F0	0.078 ± 0.274	2.818	0.480	< 0.001	3.316	0.534	< 0.001
F0 Range	0.866 ± 1.728	0.149	0.074	0.043	0.126	0.094	0.184
Mean HNR	-0.152 ± 0.373	-0.933	0.175	< 0.001	-1.033	0.209	< 0.001
Duration	0.011 ± 0.374	1.936	0.310	< 0.001	2.917	0.420	< 0.001
Scream Stimuli							
Mean F0	0.282 ± 0.385	-0.272	0.429	0.525	2.061	0.552	< 0.001 [a]
F0 Range	0.811 ± 1.438	0.315	0.106	0.003	0.056	0.372	0.710
Mean HNR	-3.013 ± 7.821	0.002	0.004	0.685	-0.004	0.007	0.587
Duration	0.779 ± 0.917	0.088	0.178	0.619	1.088	0.223	< 0.001 [a]

2   [a] Effects estimated individually due to collinearity.