nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-00963-z

Check for updates

Infants relax in response to unfamiliar foreign lullabies

Constance M. Bainbridge [1,8]✉, Mila Bertolo [1,8]✉, Julie Youngers [1,2], S. Atwood [1,3], Lidya Yurdum [1], Jan Simson [1], Kelsie Lopez [1,4], Feng Xing [1,5], Alia Martin [6] and Samuel A. Mehr [1,6,7]✉

Music is characterized by acoustic forms that are predictive of its behavioural functions. For example, adult listeners accurately identify unfamiliar lullabies as infant-directed on the basis of their musical features alone. This property could reflect a function of listeners' experiences, the basic design of the human mind, or both. Here, we show that US infants (N = 144) relax in response to eight unfamiliar foreign lullabies, relative to matched non-lullaby songs from other foreign societies, as indexed by heart rate, pupillometry and electrodermal activity. They do so consistently throughout the first year of life, suggesting that the response is not a function of their musical experiences, which are limited relative to those of adults. The infants' parents overwhelmingly chose lullabies as the songs that they would use to calm their fussy infant, despite their unfamiliarity. Together, these findings suggest that infants may be predisposed to respond to common features of lullabies found in different cultures.

Music is a human universal[1–4] that appears often in the lives of infants and their families[5–9]. Infants demonstrate a remarkable variety of responses to music as they develop: in the first few days of life, newborns remember melodies heard in the womb[10], distinguish consonant from dissonant intervals[11] and detect musical beats[12]. Older infants differentiate synchronous movement from asynchronous movement in response to music[13], become attuned to the rhythms of their native culture's music by their first birthday[14], garner social information from the songs they hear[15,16] and recall music in impressive detail[17,18] after long delays[15].

Why are infants so interested in music? One possibility centres on the dynamics of parent–offspring interactions. Relative to other animals, human infants are helpless; to survive, they rely on resources provided by parents and alloparents[19]. Such resources, whether material (such as food) or not (such as attention) constitute parental investment[20]. Human parental investment is routinely provided to infants in response to their elicitations, which often take the form of fussiness and crying[21].

Infant-directed songs may credibly signal parental attention to infants, conveying information to infants that an adult is nearby, attending to them and keeping them safe[22,23]. Singing indicates the location, proximity and orientation of the singer (even when the singer is not visible, as at night); and it is also costly, in that the singer could be expending their energy on some other activity. Because parental attention is a key resource for helpless infants, they probably are predisposed to attend to signals of it: infants should be particularly interested in and reassured by vocal music with features suggesting that it is directed toward them.

Studies of people with genomic imprinting disorders provide a unique test of this hypothesis because these disorders are characterized by divergent behaviours related to parental investment[24,25]. For example, infants with Prader–Willi syndrome elicit less parental investment than do typically developing infants: they have feeding difficulties, nursing less often; and they tend to be lethargic[26]. Children with Angelman syndrome show the opposite pattern: they elicit more parental investment, with frequent drooling and chewing, uncoordinated overfeeding and high degrees of social engagement[27].

Genomic imprinting disorders also alter the psychology of music, in a fashion consistent with the idea that infant-directed song signals parental investment. Compared with the relaxation response that typically developing people display during passive music listening, Prader–Willi syndrome is associated with an increased relaxation response[28], and Angelman syndrome is associated with a reduced relaxation response[29]. These effects are specific to music; they are not elicited by listening to pleasant speech, suggesting that singing is a particularly effective means of satisfying parental investment elicitations in Prader–Willi syndrome, and a particularly ineffective means of doing so in Angelman syndrome.

Credible signals have evolved repeatedly in many species, with similar patterns across senders and receivers[23,30]. The resulting innate links between the forms and functions of vocal signals[31–33] explain why, for example, hostile vocalizations across species—from growling tigers to shrieking eagles—are recognized as hostile by human listeners[34]. Because these signals are shaped by natural selection, they are expected to show consistency across members of a species.

Infant-directed vocalizations appear to fit this pattern. Infant-directed speech is acoustically distinct from adult-directed speech across cultures[35–40]. Lullabies, a common form of infant-directed song, are reliably distinguishable from other songs[41]; in a representative sample of music from small-scale societies, adult naïve listeners considered foreign lullabies likely to be 'used to soothe a baby', relative to dance, healing and love songs[42]. This result, which has also been supported by a massive conceptual replication (N = 29,357), is explained in large part by the striking musical consistency of

[1]Department of Psychology, Harvard University, Cambridge, MA, USA. [2]Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada. [3]Department of Psychology, University of Washington, Seattle, WA, USA. [4]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA. [5]Department of Education, Johns Hopkins University, Baltimore, MD, USA. [6]School of Psychology, Victoria University of Wellington, Wellington, New Zealand. [7]Data Science Initiative, Harvard University, Cambridge, MA, USA. [8]These authors contributed equally and are listed alphabetically: Constance M. Bainbridge and Mila Bertolo. ✉e-mail: cbainbridge@g.harvard.edu; mila_bertolo@g.harvard.edu; sam@wjh.harvard.edu

lullabies found across cultures: their slow tempos and smooth, minimally accented melodic contours[1]. Strikingly, these same musical features appear in infant-directed or low-arousal Western music[43–46].

If infant-directed song indeed functions as a credible signal of parental attention, then the universal features of the signal should produce reliable relaxation effects in the receiver: singing should satisfy infants' fussy demands for parental investment, calming them. Common sense does suggest that infants are calmed by infant-directed song, but typically, this question has been tested in the context of songs that are known to the infant and/or are sung in a familiar language. This makes it difficult to measure the specific soothing effects of infant-directed song, independently of the soothing effects of familiar sounds, more generally. Adults' ratings of the familiarity and perceived relaxation of music are positively correlated[47], and parents produce music for their children often[5–9], so familiar music may produce mere-exposure effects[48] on infant relaxation.

Indeed, infant arousal, as indexed by electrodermal activity, decreased in response to maternal singing in a 'soothing' style, relative to a 'playful' style; but both styles were produced in familiar songs[49]. Listening to live or recorded lullabies reduced heart rate in pre-term infants, more so than a silent control, but the songs were well known and produced in a familiar language[50]. Singing reduced distress after a still-face procedure, as indexed by increased smiling and decreased ratings of negative affect, but the effects were driven by the familiarity of the songs[51]. Infants attended longer to singing than speech before becoming fussy, when both were produced in a foreign language[52], but whether this effect reflects increased attention to songs or increased relaxation as a result of listening to music is unknown. In sum, while there is some evidence that infant-directed songs produce relaxation effects in infants, the effects in prior studies may be attributable to infants' familiarity with the songs, rather than the songs' acoustic properties (as would be predicted by a credible signalling account[22,23]).

Here, we ask whether infants relax in response to infant-directed songs produced in unfamiliar languages from foreign societies. We studied 144 infants' responses to videos of animated characters who lip-synched to songs drawn from the Natural History of Song Discography[1], a collection of lullabies, dance songs, healing songs and love songs recorded in 86 world cultures. The songs were either infant-directed (the lullabies) or not (the other song types) (Fig. 1 and Table 1). We measured infants' heart rate, pupil dilation, electrodermal activity, frequency of blinking and gaze direction. Based on prior results in a similar listening paradigm[28,29], we pre-registered a hypothesis that infants would show decreased heart rate (that is, a relaxation response) during the lullabies, relative to the non-lullabies. We report a test of that hypothesis, a series of planned exploratory analyses of other measures of infants' responses and a measure of parents' intuitions about the songs.

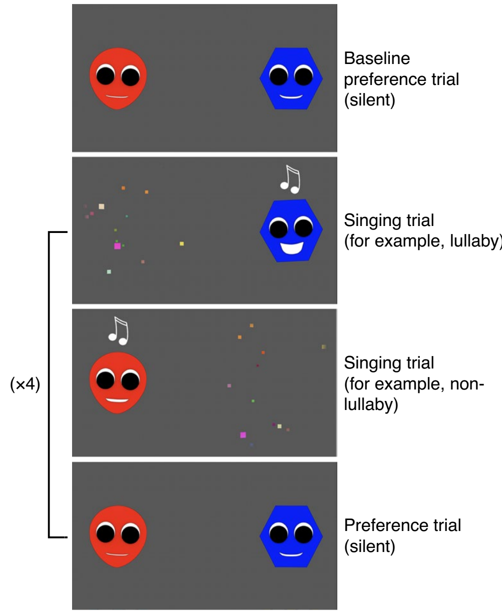

Fig. 1 | Structure of the experiment. Infants viewed videos of animated characters who either appeared in silence (during preference trials) or sang the songs one at a time, next to a distracting animation of slowly moving coloured boxes.

Results

Confirmatory analysis. We pre-registered the prediction that infants' heart rate would decrease more substantially as a result of listening to foreign lullabies than non-lullabies (the pre-registration is available at https://osf.io/f69mn). To this end, we normalized heart rate values during singing trials relative to the previous trial (where the previous trial was either a singing trial or a silent preference trial, in which the animated characters were silent but still visible; Fig. 1), such that z-scores are interpretable as immediate changes in heart rate, indexing moment-to-moment relaxation (note that this normalization procedure was also pre-registered): positive z-scores thus indicate an increase in heart rate from the previous trial, and negative scores a decrease. We obtained heart rate data by using a non-invasive physiology monitor (Empatica E4), which also measured electrodermal activity (see Exploratory analyses, below).

In the main analyses, we analysed trial-wise mean z-scores for each infant, split by song type. As in previous work[28,29], we trimmed (a) all values on trials for which there were fewer than five heart rate observations during the normalization period (the previous trial), as this would produce uninterpretable standard deviation values with which to compute z-scores, and (b) extreme values, defined as $|z| > 5$. These trimming rules dropped 2.19% and 0.31% of the heart rate observations, respectively, and 2 of the 144 participants. These decisions did not substantively affect any of the results. For this and all following analyses, assumptions for the relevant statistical tests were met.

Mean normalized heart rate during lullabies (Fig. 2a) differed significantly from 0, indicating a decrease in heart rate relative to the previous trial (in z-scores, mean (M) = −0.15, s.d. = 0.43, 95% CI [−0.23, −0.08]; $t(140) = -4.28$, $P < 0.001$, $d = 0.36$, one-sample two-tailed t-test). In contrast, heart rate during non-lullabies was comparable to 0, indicating no change in heart rate relative to the previous trial (M = −0.01, s.d. = 0.4, 95% CI [−0.07, 0.06]; $t(139) = -0.21$, $P = 0.83$, one-sample two-tailed t-test).

The within-subjects difference between mean heart rates (that is, the main pre-registered analysis) showed a clear difference between song types, such that lullabies decreased heart rates significantly more than non-lullabies (Fig. 2a; $t(138) = -2.75$, $P = 0.007$, paired two-tailed t-test). These findings confirm the pre-registered prediction of reduced heart rate in response to unfamiliar foreign lullabies.

We conducted three planned follow-up analyses. First, to determine what drove the mean difference in heart rate across lullabies and non-lullabies, we visualized the trajectory of heart rate within singing trials in a time-series analysis (Fig. 2b). While heart rates dropped almost immediately following the onset of singing, regardless of song type, this drop was more pronounced during lullabies. Because time-wise heart rate trends were nonlinear, and in the absence of any a priori predictions about those trends, we elected not to model them directly.

Second, we tested whether the heart rate effects were driven by any particular age range of infants. They were not: a regression

Table 1 | The songs infants heard. Using the Natural History of Song Discography[1], we chose eight lullabies and paired them with non-lullabies drawn from the other three song types in the corpus (dance, love or healing), matching the perceived sex of the singer. All songs were produced by solo voices without instrumental accompaniment

	Lullaby			Paired non-lullaby			
Sex	**Society**	**Region**	**Language**	**Type**	**Society**	**Region**	**Language**
Female	Saami	Scandinavia	Luk Saami	Love	Nenets	North Asia	Tundra Nenets
	Nahua	Maya Area	Western Nahuatl	Love	Serbs	Southeastern Europe	Serbian Standard
	Igulik Inuit	Arctic and Subarctic	Western Canadian Inuktitut	Dance	Chachi	Northwestern South America	Cha'palaa
	Kuna	Central America	Border Kuna	Love	Highland Scots	British Isles	Scottish Gaelic
Male	Iroquois	Eastern Woodlands	Cherokee	Love	Kurds	Middle East	Central Kurdish
	Hopi	Southwest and Basin	Hopi	Healing	Hawaiians	Polynesia	Hawaiian
	Ona	Southern South America	Selk'nam	Love	Chuuk	Micronesia	Chuukese
	Highland Scots	British Isles	Scottish Gaelic	Healing	Seri	Northern Mexico	Seri

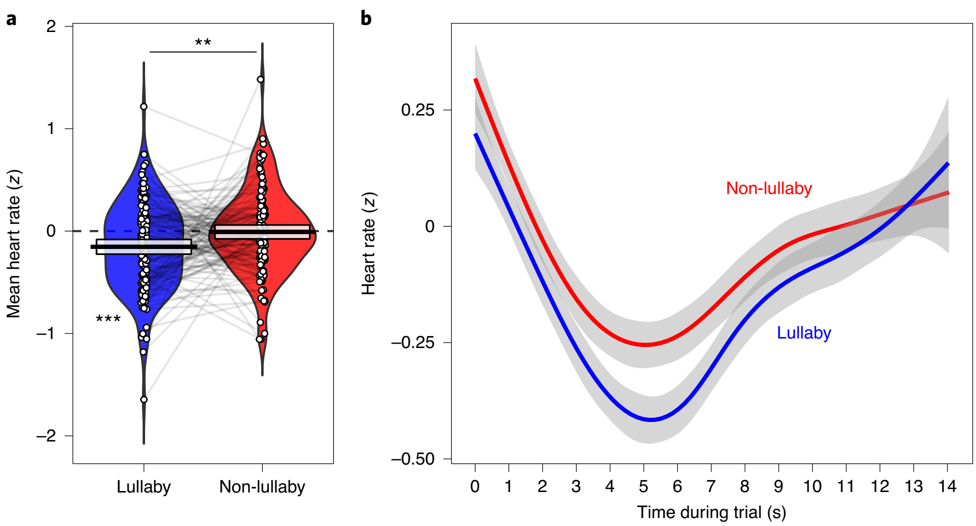

Fig. 2 | Lullabies reduce infant heart rate. a, The points depict mean trial-wise heart rates, normalized to the previous 14 s trial (regardless of its type), for each infant ($N = 142$), with the gray lines indicating the pairs of points that represent the same infants; the violin plots (coloured areas) are kernel density estimations; the horizontal black lines indicate the means across all participants, and the shaded white boxes indicate the 95% confidence intervals of the means. The points are jittered to improve clarity. Heart rates were reduced during lullabies (the mean z-score was negative and significantly different from 0 (in z-scores, $M = -0.15$, s.d. = 0.43, 95% CI [−0.23, −0.08]; $t(140) = -4.28$, $P < .001$, $d = 0.36$, one-sample two-tailed t-test), denoted by the horizontal dotted line), relative to the previous trial, but no such effect was found for non-lullabies ($M = -0.01$, s.d. = 0.4, 95% CI [−0.07, 0.06]; $t(139) = -0.21$, $P = 0.83$, one-sample two-tailed t-test). Within-infants, heart rate during lullabies was significantly lower than during non-lullabies ($t(138) = -2.75$, $P = 0.007$, paired two-tailed t-test). **b**, An analysis of heart rate over time, averaged across all trials, shows that while heart rate dropped initially in all singing trials, the drop was more pronounced in lullabies, driving the overall effect. The lines and confidence bands are from a generalized additive model that does not account for nesting. ***$P < 0.001$; **$P < 0.01$.

of the within-subjects difference between mean heart rate during lullabies versus non-lullabies on infant age found no significant effect (Supplementary Fig. 3; $F(1, 137) = 1.22$, $P = 0.27$, $R^2 = 0.01$, omnibus test).

Third, we tested whether a match between the sex of the infant's primary caregiver (as specified by the parent who attended the experiment with the infant) and the perceived sex of the singers predicted any difference in within-subjects main effects, because, for instance, when hearing male-sounding lullabies, those infants who have male primary caregivers may be likely to relax more than those infants with female primary caregivers, since male singers may sound more familiar to them. We found no evidence for such an effect: the within-subjects main effect was of comparable size across infants (main effect when sex of singer was matched to primary caregiver: $M = -0.16$, s.d. = 0.68, 95% CI [−0.32, 0.01]; main effect when sex of singer was not matched to primary caregiver: $M = -0.14$, s.d. = 0.57, 95% CI [−0.27, 0]; $t(131.28) = 0.18$, $P = 0.85$, independent-samples two-tailed t-test).

Exploratory analyses. We conducted a series of exploratory analyses to test for convergent evidence supporting the pre-registered result reported above, and to examine an alternative interpretation of the heart rate findings suggested by an anonymous reviewer: that rather than relaxing infants, the lullabies simply captured their attention more so than the other songs. Indeed, in some contexts, heart rate decreases can indicate increased attention to a stimulus[53],

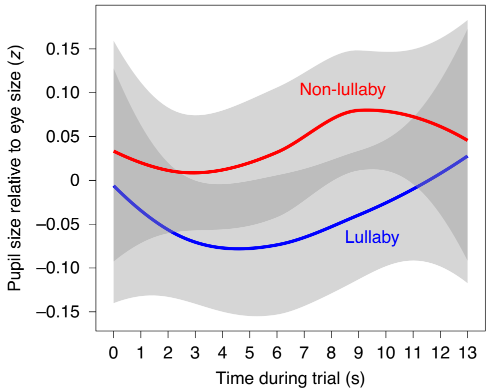

Fig. 3 | Pupil dilation is reduced during lullabies. Collapsing across all singing trials, pupil size was lower during lullabies than non-lullabies ($t(3,086) = 2.507$, $P = 0.012$, $\beta = 0.089$), in the subset of the participants studied ($N = 30$). The blue and red lines and confidence bands are from a LOESS regression that does not account for nesting.

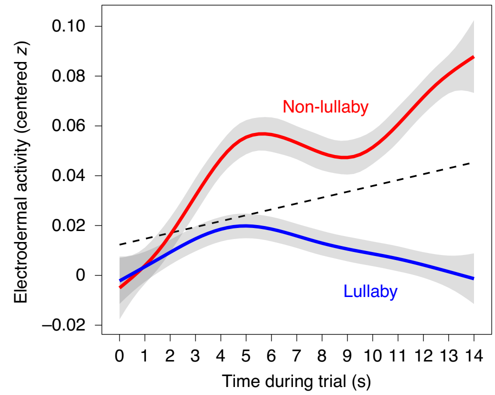

Fig. 4 | Lullabies attenuate increases in arousal. The black dotted line denotes the expected rise in electrodermal activity during a trial, from a linear model ($N = 25,938$ observations from 144 infants, $t(25,819) = 38.5$, $P < 0.001$, $\beta = 0.002$). This rise is attenuated during lullaby trials but not during non-lullaby trials, such that the expected level of electrodermal activity by the end of a lullaby trial is reduced (expected difference at time = 14 s; $\beta = 0.075$, 95% CI [−0.098, −0.052], $\chi^2 = 289.2$, $P < 0.001$, $d = 0.25$).

and music is known to attract infants' attention[52]. Additional measures can arbitrate between these interpretations.

First, we analysed infants' pupil dilation, an indicator of both attention to a stimulus[54] and emotional arousal in response to it[55], including during music listening[56,57]. If the lullabies relaxed infants, then pupil size should decrease during lullabies, relative to non-lullabies—contrasting sharply with an attention account for the heart rate findings, which would predict increases in pupil size.

Second, we analysed infants' electrodermal activity, an indicator of arousal used in prior studies of relaxation responses to music[49,51]. If the lullabies relaxed infants, then electrodermal activity should decrease during lullabies, relative to non-lullabies. Increased attention, however, does not imply a directional effect on electrodermal activity.

Third, we analysed infants' gaze toward the animated characters, and rate of blinking, as measures of interest in the songs. These measures do not bear on the relaxation hypothesis, but rather, they test the degree to which infants' attention to the animated characters varied as a function of whether they were singing lullabies or non-lullabies.

Last, in two additional analyses (unrelated to the relaxation and attention accounts described above), we explored the degree to which the perceived infant-directedness of the songs was predictive of infants' heart rates; and the degree to which *parents* made inferences about the different song types.

Relaxation response as indexed by pupillometry. Using videos of infants' faces during the experiment, we developed a procedure for manual annotation of pupil size and obtained pupil size annotations for the singing trials (Methods and Supplementary Figs. 1 and 2). We normalized these annotations across all available data from each infant, after binning observations by second to reduce noise. We analysed changes in pupil dilation over the course of a singing trial, collapsing across all trials; and tested for differences between lullabies and non-lullabies.

Consistent with a relaxation account, and in contrast to an attention account, pupils were smaller during lullabies than during non-lullabies (Fig. 3). We fitted a random-effects linear model to the z-scored observations, predicted from the time course of each trial, with a random effect of trial ($N = 3,096$ binned relative pupil size observations from 30 infants, mean = 103.2 observations per infant; likelihood ratio $\chi^2 = 7.682$, $P = 0.021$). The model showed that pupil size was smaller during lullabies than non-lullabies, on average ($t(3,086) = 2.507$, $P = 0.012$, $\beta = 0.089$). We found no time-by-trial-type interaction; this is likely because pupil size appeared to regress to the mean by the end of each trial (Fig. 3).

Relaxation response as indexed by electrodermal activity. We used the same normalization approach as the pupillometry analysis, because normalizing to the previous trial, as in the heart rate analyses, produced a distribution with unacceptably long tails (zs > 100). This is likely because the short trial length (14 s) affords only minimal variability in electrodermal activity, which generally changes much more slowly than does heart rate, inflating z-scored values. Normalization to the full experiment period produced a more acceptably narrow range of z scores, such that applying the same trimming criterion as we used for heart rate ($|z| > 5$) resulted in the removal of only 4 observations of nearly 100,000.

First, we noted an overall positive trend in electrodermal activity throughout the study, irrespective of the songs the infant was listening to. We fitted a random-effects linear model to all z-scored observations ($N = 25,938$ from 144 infants, mean = 180 observations per infant), which showed that electrodermal activity steadily increased throughout the experiment, on average ($t(25,819) = 38.5$, $P < 0.001$, $\beta = 0.002$).

Note that this result contrasts sharply with infants' responses during a distress induction procedure, as in previous research on the calming effects of singing[51]. In that type of study, arousal and fussiness increase during a negative interaction (for example, a still-face procedure), and subsequently decrease during a positive 'recovery phase'. This is unsurprising, however, given the structure of this experiment: infants often become bored and fussy during repetitive experiments, increasing arousal.

As such, we measured the rate of increase in electrodermal activity, and analysed changes in electrodermal activity as a function of lullaby or non-lullaby listening relative to this increase. This required centring the z-scores infant- and trial-wise. The key question is thus whether listening to a lullaby yields lower electrodermal activity than the predicted overall trial-wise increase, all else equal.

The results supported the relaxation account (Fig. 4). We fitted a random-effects linear model of electrodermal activity change scores over time, trial-wise, so as to test for a time by song type interaction. The model fit was acceptable (likelihood ratio $\chi^2 = 443$, $P < 0.001$), the interaction term was significant ($t(61,174) = -10.3$, $P < 0.001$, $\beta = -0.006$) and a general linear hypothesis test showed an expected difference in electrodermal activity between lullabies and non-lullabies at the end of the trial (time = 14 s; $\beta = -0.075$, 95% CI [−0.098, −0.052], $\chi^2 = 289.2$, $P < 0.001$, $d = -0.25$). These results indicate that lullabies attenuated increases in electrodermal activity.

Visual attention to singers. We ran two sets of exploratory analyses concerning infants' visual attention to the animated characters, using manual annotations of infants' duration of gaze toward them (see Methods). In previous research, infants demonstrated social preferences for a person who had previously sung a song familiar to the infant[15,16]; as such, we explored whether such a preference could be elicited purely on the basis of a difference in the types of songs a singer produced.

We found no evidence for such an effect. Infants looked for comparable durations to the two characters during singing trials (Supplementary Fig. 4; in seconds, lullabies: $M = 8.1$, s.d. = 2.57, 95% CI [7.68, 8.53]; non-lullabies: $M = 7.92$, s.d. = 2.77, 95% CI [7.46, 8.37]; $t(144) = 0.68$, $P = 0.5$, two-tailed paired t-test). The two one-sided test procedure for equivalence testing[58] confirmed that these rates of attention were statistically equivalent ($\Delta = 1$ s; Δ_L: $t(144) = 4.335$, $P < 0.001$; Δ_U: $t(144) = -2.972$, $P = 0.002$).

The same pattern was observed during the preference trials: attention to the two characters in silence, and after they had each sung a lullaby or non-lullaby, did not differ (Supplementary Fig. 4; attention in seconds to lullaby singer: $M = 4.42$, s.d. = 2.45, 95% CI [4.02, 4.82]; non-lullabies: $M = 4.62$, s.d. = 2.66, 95% CI [4.18, 5.05]; $t(145) = -0.53$, $P = 0.6$, two-tailed paired t-test). These rates were statistically equivalent ($\Delta = 1$ s; Δ_L: $t(145) = 2.202$, $P = 0.015$; Δ_U: $t(145) = -3.264$, $P < 0.001$). Note that these analyses include a few more infants than the heart rate analyses do; this is because some infants completed the study and were subsequently excluded from the heart rate analyses due to a poor physiology monitor signal, but had usable gaze data.

As an additional exploratory measure, we counted the number of eye blinks during the singing trials, as blinks may index perceived stimulus salience[59]. Infants blinked slightly less during lullabies (number of blinks per trial: median = 1, interquartile range: 0.5–1.5) than non-lullabies (median = 1, interquartile range: 0.5–2), suggesting that they were more interested in the singers during lullabies than during non-lullabies ($z = -2.25$, $P = 0.02$, Wilcoxon two-tailed signed-rank test). But blinking was rare, so this exploratory result should be interpreted with caution, as it may be an artefact of restricted range.

Relation between songs' infant-directedness and relaxation effects. The lullabies we studied differ acoustically from non-lullabies in a number of ways: they tend to be less accented and slower in tempo, and have smaller pitch ranges and more variable macro-meters than the other songs[1]. These features are reflected in naïve listeners' ratings: the lullabies are perceived to have lower melodic and rhythmic complexity, slower tempo, less steady beat, lower arousal, lower valence and lower pleasantness[42]. Together, these features predict the degree to which listeners perceive a song as infant-directed[35,42].

Infants' relaxation responses to lullabies should be explicable by their responses to these acoustic features. To test this, we asked whether we could predict variability in infant physiological responses as a function of the degree of infant-directedness of each song, using the adult ratings from prior work[42]. Modelling trials and participants as random effects in a linear regression, we predicted infant heart rate from songs' perceived infant-directedness. We found a significant negative relationship, of modest size, such that the more infant-directed a song was, the larger the expected reduction in infant heart rate ($t(15,239) = -6$, $P < 0.001$, $\beta = -0.052$). This result confirms that the acoustic features of the songs drove the relaxation effects on infants.

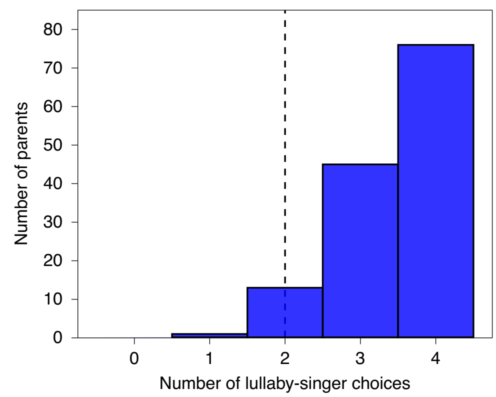

Fig. 5 | Parents prefer foreign lullabies to non-lullabies for soothing their own infants. The histogram displays the distribution of parents' choices of whose song (lullaby singer or non-lullaby singer) they would prefer to sing to their own infant (if the infant were fussy and if they knew how to sing both songs; $N = 135$ parents of our 144 infants participated in this follow-up survey). Parents made this choice four times, so the maximum number of lullaby singer choices was four. The dashed line indicates the chance level of two choices. Parents almost always chose the lullaby singer (interquartile range: 3–4; $z = 9.89$, $P < 0.001$, Wilcoxon two-tailed signed-rank test).

Parent intuitions about foreign lullabies. After each infant had participated in the experiment, we showed a parent the same animated characters that the infant had seen. The characters sang the eight songs that were not presented during the experiment, such that the songs were unfamiliar to the parent.

We asked the parent, for each pair of songs, to choose the character whose song they would prefer to sing to soothe their infant, if the infant were fussy and the parent knew how to sing the songs. Given previous findings that adults are sensitive to the soothing functions of foreign lullabies[1,42], we expected parents to choose the lullaby singers more often than the non-lullaby singers.

They did (Fig. 5). For the four pairs of songs, the median number of choices for the lullaby singer was four (all of them), a rate higher than the chance level of two choices (interquartile range: 3–4; $z = -9.89$, $P < 0.001$, Wilcoxon two-tailed signed-rank test).

Discussion

We found that infants relax in response to unfamiliar foreign lullabies. Relative to non-lullabies, infants' heart rates slowed while listening to lullabies; this effect did not merely reflect attention-related heart rate deceleration, as it was accompanied by decreased pupil dilation and attenuated electrodermal activity. Moreover, the size of the heart rate effect remained steady across all ages of infants in the age range 2–14 months, suggesting that it was not altered by infants' rapidly growing experience with music. And the effect was predictable as a function of the degree of infant-directedness of the songs, suggesting that a core set of acoustic features associated with infant-directedness across cultures produced the psychophysiological effects in the infants.

Infants were also highly attentive to the simple animated characters who produced the songs: they reliably attended to the characters for the majority of each singing trial, rarely blinking, and blinking modestly less during lullabies than non-lullabies. Moreover, parents uniformly chose lullabies over non-lullabies as the songs that they themselves would prefer to use to calm a fussy infant.

Infants and parents demonstrated all these behaviours without having previously learned anything about the music in question: they were given no cues as to the original behavioural contexts of the songs, as all the music was produced by solo voices without accompanying instruments; and they were unfamiliar with the songs they heard, unfamiliar with the languages they were sung in, and unfamiliar with the musical styles of the societies that originally produced the songs.

These findings support a hypothesized role for infant-directed song in the ecosystem of parental investment[22,23]—including the proclivity of parents to sing to their infants[6–8], the acoustic features that characterize infant-directed songs worldwide[1,35,41–46], infants' ability to perceive them and motivation to engage with them[60,61] and their calming effects[49–52]—that is both universal and innately specified.

Note, however, that we only studied infants and parents from a single Western, educated, industrialized, rich and democratic (WEIRD) society[62], who heard music from many other societies during this experiment. We expect that the findings will repeat with infants and parents in any society and are eager to find out whether they do.

The findings reported here may also be compatible with an alternative account. Over the course of early infancy, infants probably learn associations between soothing, sleep-inducing contexts and lullabies produced by their caregivers. Perhaps the infants we studied listened to the unfamiliar, foreign lullabies, found that they sounded somewhat similar to the lullabies that their caregivers produce and subsequently relaxed. Such an account would not explain cross-cultural consistency in lullaby features, but these could have arisen via mechanisms other than innate specifications of infant responses (for example, convergent cultural evolution[63]). The consistency of the relaxation effects across a full year of infancy (see Supplementary Fig. 3), a time when infant music perception is actively shaped by musical experience[14], weighs against an experience-dependent interpretation, but we cannot rule it out.

Whether the relaxation effects reflect infants' predispositions, early learning, or both, two aspects of infants' responses to music are surprising. First, whereas prior work has demonstrated effects of music on infant arousal[50,51], they were likely bolstered by mere-familiarity effects[48], as infants have robust preferences for familiar, positive experiences. Here, because infants were unfamiliar with all aspects of the lullabies (including the languages in which they were sung and the societies in which they were recorded), the results imply a specific soothing effect of music, over and above any potential effects of familiarity. Second, despite the fact that lullabies are characterized by a universal set of acoustic features, there is nevertheless a great deal of variability in the lullabies infants heard (see stimuli at https://osf.io/2t6cy). This implies that infants' responses are robust to a degree of musical variability, providing further support for the idea that infant-directed song induces relaxation in infants.

We leave open at least three series of questions about how and why infant-directed music works the way it does.

First, what is it about lullabies that makes them relaxing for infants? The acoustic features of lullabies differ from those of other songs in systematic ways, universally[1]: which of these features drive the relaxing effect of lullabies, and how? Do those features reflect evolved predispositions that are specific to music[22,23]? Or might they reflect general form–function principles of animal vocal signals, such as those that lead alarm signals to be consistently loud and harsh across species[31,32,64]? Future studies could test these questions by comparing infants' physiological responses across different song types and across different acoustic signals, or by systematically manipulating the acoustic features present in songs to measure their relaxation effects on listeners. Prior research has outlined some musical features that parents in Western cultures exaggerate while singing lullabies[43,65], which correspond with acoustic differences between infant- and adult-directed song across many societies[35]; these are good candidates for possible features that may drive relaxation effects.

Experiments in adults might also inform whether soothing effects of particular acoustic features in music could underlie later responses to music in adulthood. While adults no longer seek out parental investment, perhaps the musical features that soothe fussy infants (for example, slower tempos and fewer rhythmic accents) correspond with musical features that shape adults' emotional responses to music[66,67], which could influence physiological responses to music listening.

Second, while infants in our studies listened to songs produced by simple animated characters in isolation, their real-world musical experiences are far richer, obviously. Parents sing to their infants in a multitude of environments (before a nap, in the car, during a bath) and as part of complex multimodal experiences including other actions (rocking, bouncing) and other content (stories, instructions). The relaxing properties of lullabies demonstrated here likely interact with all of these other features. Experiments that manipulate them—for example, by comparing the relaxation effects of music listening in isolation with those of music listening while being rocked—would more fully lay out the feature space of the infant's musical experience, specifying what it is about lullabies that infants find satisfying in natural contexts of parental investment. Moreover, lullabies are but one of many forms of infant-directed songs; songs directed toward infants in the context of play[68], which also appear universally[1], likely have their own unique, contrasting features, and corresponding effects on infants.

Third, at present we know very little about what inferences infants make about the songs they hear or the people who sing them. In previous work, infants preferred the singer of a song that infants had previously learned in the context of live social interactions[15,16]. Here, although infants relaxed more when listening to lullabies, they showed no preference for the lullaby singer over the other singer during the silent preference trials or during singing. This suggests that infants' physiological responses to music may be dissociated from their musical and/or social preferences. One possibility is that the lullabies are more relaxing than the other songs for infants, but that they do not necessarily prefer hearing the lullabies. Another is that infants do prefer listening to the lullabies over the other songs, but that this does not translate to a social preference for their singers. Infants' social preferences for singers may rely instead on the context in which they learned the songs they know, such that infants consider singers of songs which they previously learned in a social interaction to be particularly good social partners or sources of parental investment[15,16].

Taking this possibility a step further, might infants expect caregiving characters to produce infant-directed music for other distressed babies, just as they expect different adults who soothe the same baby to affiliate with one another[69]? Such results are plausible, given the known links between musical experience and infant social cognition[15,16,70,71], but have not yet been tested.

However these lines of research play out, the present findings immediately suggest that singing is an effective means by which caregivers can relax infants, and raise the possibility of cumulative positive effects of music on infant and parent well-being. Live and recorded music have shown promise in improving a variety of clinical outcomes[72,73], including in parents and infants[74–78]. Music may also play an everyday role in improving health in infants—a role it has taken on across cultures and across human history[1].

Methods

Participants. This research was approved by the Committee on the Use of Human Subjects, Harvard University's Institutional Review Board and complies with all relevant ethical regulations. Parents provided informed consent before their and their infant's participation.

We recruited 144 typically developing infants from the Greater Boston area (72 females, mean age = 7.2 months, s.d. = 3.1, range: 2.1–14.3). Data from an additional 21 infants were collected but excluded due to infant fussiness ($n = 11$), lack of attention ($n = 1$), technical error ($n = 8$) or experimenter error ($n = 1$).

Nearly all infants were born full-term. Information about language exposure was available from 98 of the participants; of these, none of the languages spoken at home matched those used in the stimuli of this study (see Table 1).

Infants who became fussy and ended their participation partway through the study were included in the analyses if they attended to the first pair of songs and the subsequent preference trial (see Stimuli). Most infants ($n = 125$) contributed data for all four song pairs and preference trials. For compensation, parents received a \$5 gift card and infants were given a prize. All testing took place at the Music Lab at Harvard University.

Stimuli. We chose 16 songs from the Natural History of Song Discography[1] that were originally produced in 15 different societies and languages (Table 1). Eight of the songs were infant-directed, having been used as lullabies (that is, they were originally used to soothe, calm or put an infant/child to sleep) in the societies where they were recorded, according to the anthropologist or ethnomusicologist who collected each recording. The other eight songs were originally produced in the context of expressing love (five), healing the sick (two) or dancing (one).

We chose this particular subset of 16 songs by first limiting the corpus to those songs produced by a single singer with no instrumental accompaniment; then, using adults' ratings of the songs from a previous study[42], we chose a set of lullabies rated as likely to be 'used to soothe a baby' and a set of non-lullabies with low ratings on that item.

We paired the lullabies and non-lullabies from these sets so as to match the perceived sex of the singer as closely as possible, because infants are sensitive to the sex of voices[79]. We ordered the pairs such that those with larger differences on the rating 'used to soothe a baby' were presented first, so as to maximize the measurable differences in responses to lullabies versus non-lullabies in each infant, even if they became inattentive or fussy partway through the study. All recordings were normalized to approximately balance their perceived loudness and were also manually edited to remove background noise and microphone artefacts, using noise reduction filters and equalization.

We generated animations of two characters who lip-synched to each song, giving the impression that they were singing (Fig. 1; videos are available at https://osf.io/2t6cy). Each character sang four songs, such that one exclusively sang lullabies while the other exclusively sang non-lullabies. The videos were counterbalanced on four dimensions: which was the first song heard (lullaby or non-lullaby), which character was the lullaby singer (red or blue), which side the lullaby singer appeared on (left or right, to match character placement during silent preference trials; see Procedure) and the perceived sex of the singer (male or female). This yielded 16 conditions, which we balanced across ages, such that each counterbalancing condition included infants across the full range of ages tested.

Regardless of the counterbalancing condition, we varied the presentation order of lullabies versus non-lullabies, so that they did not appear in strict alternation, which could introduce order effects. This yielded trial orderings that were either P-L-N-P-N-L-P-L-N-P-N-L-P or P-N-L-P-L-N-P-N-L-P-L-N-P, where L denotes a lullaby singing trial, N denotes a non-lullaby singing trial and P denotes a preference trial. Because there were two characters, and each character sang four songs, each infant in the experiment listened to 8 of the 16 songs.

Procedure. Infants sat in a high chair ($n = 62$), recliner ($n = 57$) or a parent's lap ($n = 25$) approximately 150 cm away from a 107.5×60.5 cm^2 television screen; parents chose the seat based on the physical size of the infant and whether the infant was comfortable sitting in it. When infants sat in a high chair or recliner, the parent sat behind them. When infants sat on their parent's lap, the parent listened to masking music through passive noise-cancelling headphones throughout the experiment; we also asked parents to keep their eyes closed. We recorded videos of the infants at ultra-high definition (8-bit 4K at 150 Mbps; Panasonic Lumix GH5S and Lumix G Vario 14–140 mm lens).

Figure 1 depicts the order of events. The experiment began with a 14 s baseline preference trial, in which the two animated characters were presented simultaneously in silence. Four sets of three trials followed, with each set consisting of two singing trials and one preference trial. On the singing trials, one of the animated characters sang a song, appearing alone on the screen next to a screensaver-like animation (to reduce the likelihood that infants would look only at the singer). Each singing trial was 14 s long. The preference trials were identical to the baseline preference trial. Attention-grabbing animations appeared at the centre of the screen before each preference trial. The experiment lasted about five minutes. Experimenters were unaware of the order of presentation of stimuli and thus were unable to influence the infants' behaviour or responses in relation to the music they heard.

Characters on the screen were 25 cm wide. They were presented 45 cm apart when appearing simultaneously during the preference trials. Videos were presented at 4K resolution, and audio played from two speakers (Neumann KH80 DSP) at approximately the height of the infants' ears, 125 cm apart, placed such that the infant was seated at the apex of an equilateral triangle formed with the two speakers. The songs had a maximum volume of approximately 60 dB.

Psychophysiology. We recorded infant heart rate and electrodermal activity with a physiological monitor (Empatica E4) attached to the infant's thigh or calf, depending on the size of the infant, and usually on the left side. The monitor records heart rate via a photoplethysmograph at the site of the device and electrodermal activity via electrodes attached to the side or bottom of the infant's foot (with BIOPAC isotonic gel); it has been successfully validated in adults[80].

Pupillometry. We developed a procedure to manually annotate pupil dilation and applied it to still images from 30 of the infants. We extracted still images of the infant's face from the videos and used the `dlib` face recognition library[81] to automatically rotate the frame, levelling the eye horizontally; and to isolate one of the infant's eyes (we randomly selected either the left or right eye for each infant). Workers on Amazon Mechanical Turk then viewed each eye image (see Supplementary Fig. 1) and were asked (1) to adjust its brightness and contrast, so as to optimize visibility of the pupil; (2) to rate how visible the pupil was (from one of six options: Pupil is clearly visible; Pupil is visible, but it's difficult to see; Pupil is NOT visible, but I could see enough of it to make a guess about its outline; Pupil is NOT visible but the eye is still open; Pupil is NOT visible because the eye is closed; Other); and (3) to draw a superimposed ellipse on the image, surrounding the visible area of the pupil.

We set two qualification criteria for workers based on their performance in ten eye images: (1) a correlation of at least $r = 0.8$ between their annotations (that is, the width and height of their ellipses) and the mean annotations from a pilot study ($N = 46$ workers); and (2) in at least seven of the ten images, a matching visibility rating with the option selected by at least 15% of pilot participants. Workers were not aware of whether or not an image being annotated was counted toward the qualification, but they were told that their performance was being evaluated in real time.

The pool of qualified workers then annotated three images per second of infant video, drawn from the singing trials only and presented in a random order. Each worker annotated approximately 263 images and spent 19.5 s per image, on average (trimming top and bottom 1%). Four images per trial were 'validation images' that were presented more than once to the same worker, providing a measure of internal reliability of the annotations. Reliability was high, as measured in two ways. First, visibility ratings were internally consistent (that is, validation images were generally classified repeatedly in the same fashion by annotators; see confusion matrix in Supplementary Fig. 2). Second, the annotated pupil sizes were internally consistent: validation annotations correlated with the original annotations at $r = 0.88$ (using total pupil area). The degree of reliability varied as a function of how visible the pupil was; validation images marked "Pupil is clearly visible" correlated at $r = 0.95$, whereas those marked "Pupil is NOT visible, but I could see enough of it to make a guess about its outline" correlated less strongly, at $r = 0.86$. Because the workers were unaware of what songs the infants were listening to at the time of each still image (or even that the infants were listening to music at all) they always remained blinded to the conditions of the study.

To produce the data used in analyses, we computed a relative pupil size measure by dividing the pupil area by the full eye area, in pixels and within-participants, so as to adjust for increases in visible pupil size due to motion toward or away from the camera (which would erroneously increase or decrease the visible area of the pupil, respectively). Last, we removed all observations above the 99th percentile and below the 1st percentile; these appeared to be impossibly large or small values due to face recognition errors in the automated image extraction.

Gaze and blinking. We manually annotated infant gaze and blinks, frame-by-frame at 60 fps using Datavyu[82]. Annotators always worked with the audio muted, so that they remained unaware of the songs each character sang, and thus were blinded to the conditions of the study.

For gaze, we randomly selected 20% of the videos, which a second person then annotated, independently of the first set of annotations. We assessed reliability by correlating trial-wise durations of gaze toward the two locations on the screen across pairs of annotators for each infant. Reliability was high (median $r = 0.98$, interquartile range: 0.90–0.99).

For blinks, which are more difficult to annotate, and, given their sparsity, are more likely to produce internally unreliable annotations, we used a slightly different procedure. Two annotators independently annotated all the videos, and we assessed the reliability of each video's annotations by correlating the two annotators' trial-wise counts of blinks for each infant. The distribution of correlations was strongly left-skewed, with approximately ten low outliers ($rs < 0.6$). The annotators revisited those ten videos and corrected any evident errors, or elected to drop these infants from analyses, because they disagreed about the timing and frequency of the blinking. The decision to drop these participants was made blind to the results of any analyses. Among the remaining participants ($n = 140$) reliability was high (median $r = 0.94$, interquartile range: 0.85–1).

Parent measures. After the infant completed the experiment, parents viewed videos of singing trials for the eight songs which their infant had not heard during the study, using a tablet. For each pair, we asked parents to choose the song they would prefer to sing if their baby were fussy (and assuming the parent already knew how to sing both songs). We analysed all available data, regardless of whether or not the parent's infant had completed the experiment. Parents also completed a survey concerning their infant's home musical environment, for use in a separate study.

Statistical power. We used data from a similar listening experiment in typically developing adults[28] to compute a plausible within-subjects effect size, based on the difference in mean heart rate during speech versus song in people with Prader–Willi syndrome ($d = 0.36$). We chose a target sample size of $N = 144$ before running the experiment to provide power greater than 0.99 for the main planned comparison (that is, mean heart rate during lullaby trials relative to non-lullaby trials). We also chose this sample size to facilitate even counterbalancing of stimuli across a wide range of infant ages, maximizing our ability to measure age effects while avoiding effects of stimulus ordering.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data reported here are available at https://github.com/themusiclab/infant-relax. Stimuli are available at https://osf.io/2t6cy. Audio excerpts from the Natural History of Song Discography are available at https://osf.io/vcybz; the full corpus can be explored interactively at https://themusiclab.org/nhsplots. For assistance with data or materials, please contact M.B., C.M.B. and S.A.M.

Code availability

A reproducible version of this manuscript, including all analysis and visualization code, is available at https://github.com/themusiclab/infant-relax. For assistance with code, please contact M.B., C.M.B. and S.A.M.

Received: 17 January 2020; Accepted: 11 September 2020;
Published online: 19 October 2020

References

1. Mehr, S. A. et al. Universality and diversity in human song. *Science* **366**, 957–970 (2019).
2. Jacoby, N. & McDermott, J. H. Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr. Biol.* **27**, 359–370 (2017).
3. Jacoby, N. et al. Universal and non-universal features of musical pitch perception revealed by singing. *Curr. Biol.* **29**, 3229–3243 (2019).
4. Ravignani, A., Delgado, T. & Kirby, S. Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* **1**, 0007 (2016).
5. Mendoza, J. K. & Fausey, C. M. Everyday music in infancy. *PsyArXiv* https://doi.org/10.31234/osf.io/sqatb (2019).
6. Custodero, L. A., Rebello Britto, P. & Brooks-Gunn, J. Musical lives: a collective portrait of American parents and their young children. *J. Appl. Dev. Psychol.* **24**, 553–572 (2003).
7. Custodero, L. A. & Johnson-Green, E. A. Passing the cultural torch: musical experience and musical parenting of infants. *J. Res. Music Educ.* **51**, 102–114 (2003).
8. Mehr, S. A. Music in the home: new evidence for an intergenerational link. *J. Res. Music Educ.* **62**, 78–88 (2014).
9. Trehub, S. E., Hill, D. S. & Kamenetsky, S. B. Parents' sung performances for infants. *Can. J. Exp. Psychol.* **51**, 385–396 (1997).
10. Granier-Deferre, C., Bassereau, S., Ribeiro, A., Jacquet, A.-Y. & DeCasper, A. J. A melodic contour repeatedly experienced by human near-term fetuses elicits a profound cardiac reaction one month after birth. *PLoS ONE* **6**, e17304 (2011).
11. Zentner, M. R. & Kagan, J. Perception of music by infants. *Nature* **383**, 29 (1996).
12. Winkler, I., Háden, G. P., Ladinig, O., Sziller, I. & Honing, H. Newborn infants detect the beat in music. *Proc. Natl Acad. Sci. USA* **106**, 2468–2471 (2009).
13. Hannon, E. E., Schachner, A. & Nave-Blodgett, J. E. Babies know bad dancing when they see it: older but not younger infants discriminate between synchronous and asynchronous audiovisual musical displays. *J. Exp. Child Psychol.* **159**, 159–174 (2017).
14. Hannon, E. E. & Trehub, S. E. Tuning in to musical rhythms: infants learn more readily than adults. *Proc. Natl Acad. Sci. USA* **102**, 12639–12643 (2005).
15. Mehr, S. A., Song, L. A. & Spelke, E. S. For 5-month-old infants, melodies are social. *Psychol. Sci.* **27**, 486–501 (2016).
16. Mehr, S. A. & Spelke, E. S. Shared musical knowledge in 11-month-old infants. *Dev. Sci.* **21**, (2017).
17. Trainor, L. J., Wu, L. & Tsang, C. D. Long-term memory for music: infants remember tempo and timbre. *Dev. Sci.* **7**, 289–296 (2004).
18. Volkova, A., Trehub, S. E. & Schellenberg, E. G. Infants' memory for musical performances. *Dev. Sci.* **9**, 583–589 (2006).
19. Hrdy, S. B. *Mothers and Others: The Evolutionary Origins of Mutual Understanding* (Harvard University Press, 2009).
20. Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B. G.) 136–179 (Aldine, 1972).
21. Soltis, J. The signal functions of early infant crying. *Behav. Brain Sci.* **27**, 443–458 (2004).
22. Mehr, S. A. & Krasnow, M. M. Parent–offspring conflict and the evolution of infant-directed song. *Evol. Hum. Behav.* **38**, 674–684 (2017).
23. Mehr, S. A., Krasnow, M., Bryant, G. A. & Hagen, E. H. Origins of music in credible signaling. *Behav. Brain Sci.* https://doi.org/10.1017/S0140525X20000345 (2020).
24. Haig, D. & Wharton, R. Prader–Willi syndrome and the evolution of human childhood. *Am. J. Hum. Biol.* **15**, 320–329 (2003).
25. Ubeda, F. Evolution of genomic imprinting with biparental care: implications for Prader–Willi and Angelman syndromes. *PLoS Biol.* **6**, 1678–1692 (2008).
26. Cassidy, S. B. & Driscoll, D. J. Prader–Willi syndrome. *Eur. J. Hum. Genet.* **17**, 3–13 (2008).
27. Williams, C. A. et al. Angelman syndrome 2005: updated consensus for diagnostic criteria. *Am. J. Med. Genet.* **140**, 413–418 (2006).
28. Mehr, S. A., Kotler, J., Howard, R. M., Haig, D. & Krasnow, M. M. Genomic imprinting is implicated in the psychology of music. *Psychol. Sci.* **28**, 1455–1467 (2017).
29. Kotler, J., Mehr, S. A., Egner, A., Haig, D. & Krasnow, M. M. Response to vocal music in Angelman syndrome contrasts with Prader–Willi syndrome. *Evol. Hum. Behav.* **40**, 420–426 (2019).
30. Maynard Smith, J. & Harper, D. *Animal Signals* (Oxford Univ. Press, 2003).
31. Morton, E. S. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* **111**, 855–869 (1977).
32. Owren, M. J. & Rendall, D. Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evol. Anthropol.* **10**, 58–71 (2001).
33. Endler, J. A. Some general comments on the evolution and design of animal communication systems. *Philos. Trans. R. Soc. B* **340**, 215–225 (1993).
34. Filippi, P. et al. Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proc. R. Soc. B* **284**, (2017).
35. Moser, C. J. et al. Acoustic regularities in infant-directed vocalizations across cultures. *bioRxiv* https://doi.org/10.1101/2020.04.09.032995v1 (2020).
36. Fernald, A. Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Dev.* **60**, 1497–1510 (1989).
37. Kuhl, P. K. et al. Cross-language analysis of phonetic units in language addressed to infants. *Science* **277**, 684–686 (1997).
38. Piazza, E. A., Iordan, M. C. & Lew-Williams, C. Mothers consistently alter their unique vocal fingerprints when communicating with infants. *Curr. Biol.* **27**, 3162–3167 (2017).
39. Broesch, T. & Bryant, G. A. Fathers' infant-directed speech in a small-scale society. *Child Dev.* **89**, e29–e41 (2018).
40. Bryant, G. A. & Barrett, H. C. Recognizing intentions in infant-directed speech: evidence for universals. *Psychol. Sci.* **18**, 746–751 (2007).
41. Trehub, S. E., Unyk, A. M. & Trainor, L. J. Adults identify infant-directed music across cultures. *Infant Behav. Dev.* **16**, 193–211 (1993).
42. Mehr, S. A., Singh, M., York, H., Glowacki, L. & Krasnow, M. M. Form and function in human song. *Curr. Biol.* **28**, 356–368 (2018).
43. Trainor, L. J., Clark, E. D., Huntley, A. & Adams, B. A. The acoustic basis of preferences for infant-directed singing. *Infant Behav. Dev.* **20**, 383–396 (1997).
44. Trehub, S. E. et al. Mothers' and fathers' singing to infants. *Dev. Psychol.* **33**, 500–507 (1997).
45. Gomez, P. & Danuser, B. Relationships between musical structure and psychophysiological measures of emotion. *Emotion* **7**, 377–387 (2007).
46. Rock, A., Trainor, L. & Addison, T. Distinctive messages in infant-directed lullabies and play songs. *Dev. Psychol.* **35**, 527–534 (1999).
47. Tan, X., Yowler, C. J., Super, D. M. & Fratianne, R. B. The interplay of preference, familiarity and psychophysical properties in defining relaxation music. *J. Music Ther.* **49**, 150–179 (2012).
48. Zajonc, R. B. Mere exposure: a gateway to the subliminal. *Curr. Dir. Psychol. Sci.* **10**, 224–228 (2001).
49. Cirelli, L. K., Jurewicz, Z. B. & Trehub, S. E. Effects of maternal singing style on mother–infant arousal and behavior. *J. Cogn. Neurosci.* **37**, 1213–1220 (2019).
50. Garunkstiene, R., Buinauskiene, J., Uloziene, I. & Markuniene, E. Controlled trial of live versus recorded lullabies in preterm infants. *Nord. J. Music Ther.* **23**, 71–88 (2014).
51. Cirelli, L. K. & Trehub, S. E. Familiar songs reduce infant distress. *Dev. Psychol.* **56**, 861–868 (2020).
52. Corbeil, M., Trehub, S. E. & Peretz, I. Singing delays the onset of infant distress. *Infancy* **21**, 373–391 (2016).
53. Richards, J. E. & Cronise, K. Extended visual fixation in the early preschool years: look duration, heart rate changes, and attentional inertia. *Child Dev.* **71**, 602–620 (2000).
54. Laeng, B., Sirois, S. & Gredebäck, G. Pupillometry: a window to the preconscious? *Perspect. Psychol. Sci.* **7**, 18–27 (2012).
55. Bradley, M. M., Miccoli, L., Escrig, M. A. & Lang, P. J. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **45**, 602–607 (2008).

56. Laeng, B., Eidet, L. M., Sulutvedt, U. & Panksepp, J. Music chills: the eye pupil as a mirror to music's soul. *Conscious. Cogn.* **44**, 161–178 (2016).
57. Widmann, A., Schröger, E. & Wetzel, 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).
58. Lakens, D., Scheel, A. M. & Isager, P. M. Equivalence testing for psychological research: a tutorial. *Adv. Methods Pract. Psychol. Sci.* **1**, 259–269 (2018).
59. Shultz, S., Klin, A. & Jones, W. Inhibition of eye blinking reveals subjective perceptions of stimulus salience. *Proc. Natl Acad. Sci. USA* **108**, 21270–21275 (2011).
60. Trehub, S. E. Musical predispositions in infancy. *Ann. N. Y. Acad. Sci.* **930**, 1–16 (2001).
61. Hannon, E. E. & Trainor, L. J. Music acquisition: effects of enculturation and formal training on development. *Trends Cogn. Sci.* **11**, 466–72 (2007).
62. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
63. Sperber, D. & Hirschfeld, L. A. The cognitive foundations of cultural stability and diversity. *Trends Cogn. Sci.* **8**, 40–46 (2004).
64. Blumstein, D. T., Bryant, G. A. & Kaye, P. The sound of arousal in music is context-dependent. *Biol. Lett.* **8**, 744–747 (2012).
65. Trainor, L. J. Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behav. Dev.* **19**, 83–92 (1996).
66. Sievers, B., Polansky, L., Casey, M. & Wheatley, T. Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. Natl Acad. Sci. USA* **110**, 70–75 (2013).
67. Cowen, A. S., Fang, X., Sauter, D. & Keltner, D. What music makes us feel: at least 13 dimensions organize subjective experiences associated with music across different cultures. *Proc. Natl. Acad. Sci. USA* **117**, 1924–1934 (2020).
68. Trehub, S. E. & Trainor, L. Singing to infants: lullabies and play songs. *Adv. Infancy Res* **12**, 43–78 (1998).
69. Spokes, A. C. & Spelke, E. S. The cradle of social knowledge: infants' reasoning about caregiving and affiliation. *Cognition* **159**, 102–116 (2017).
70. Soley, G. & Sebastián-Gallés, N. Infants prefer tunes previously introduced by speakers of their native language. *Child Dev.* **86**, 1685–1692 (2015).
71. Tsang, C. D., Falk, S. & Hessel, A. Infants prefer infant-directed song over speech. *Child Dev.* **88**, 1207–1215 (2017).
72. Hole, J., Hirsch, M., Ball, E. & Meads, C. Music as an aid for postoperative recovery in adults: a systematic review and meta-analysis. *Lancet* **386**, 1659–1671 (2015).
73. Richard-Lalonde, M. et al. The effect of music on pain in the adult intensive care unit: a systematic review of randomized controlled trials. *J. Pain Symptom Manage.* **59**, 1304–1319 (2019).
74. Bieleninik, Ł., Ghetti, C. & Gold, C. Music therapy for preterm infants and their parents: a meta-analysis. *Pediatrics* **138**, e20160971 (2016).
75. Bo, L. K., Mn, B. & Callaghan, P. Soothing pain-elicited distress in Chinese neonates. *Pediatrics* **105**, e49–e49 (2000).
76. Keith, D. R., Russell, K. & Weaver, B. S. The effects of music listening on inconsolable crying in premature infants. *J. Music Ther.* **46**, 191–203 (2009).
77. Filippa, M., Devouche, E., Arioni, C., Imberty, M. & Gratier, M. Live maternal speech and singing have beneficial effects on hospitalized preterm infants. *Acta Paediatr.* **102**, 1017–1020 (2013).
78. Fancourt, D. & Perkins, R. Could listening to music during pregnancy be protective against postnatal depression and poor wellbeing post birth? Longitudinal associations from a preliminary prospective cohort study. *BMJ Open* **8**, e021251 (2018).
79. Miller, C. L. Developmental changes in male/female voice classification by infants. *Infant Behav. Dev.* **6**, 313–330 (1983).
80. van Lier, H. G. et al. A standardized validity assessment protocol for physiological signals from wearable technology: methodological underpinnings and an application to the E4 biosensor. *Behav. Res. Methods* **52**, 607–629 (2020).
81. King, D. E. Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res.* **10**, 1755–1758 (2009).
82. Datavyu Team. *Datavyu: a video coding tool* (Databrary Project, New York Univ., 2014).

Acknowledgements

We thank the infants and parents who participated in this research; J. Kominsky, N. Soja, W. Pepe, E. Spelke and S. Carey for their support with participant recruitment; H. Alton, A. Bergson, A. Bitran, G. Jessani, A. Keomurjian and B. Milosh for research assistance; and C. Payne and three anonymous reviewers for detailed and constructive feedback on the manuscript. This research was supported by the NIH Director's Early Independence Award DP5OD024566, the Harvard University Department of Psychology and the Harvard Data Science Initiative. The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Author contributions

S.A.M., S.A. and C.M.B. designed the research, supported with ideas from A.M. J.Y., C.M.B. and S.A. led data collection, assisted by M.B., L.Y., K.L. and F.X., under the supervision of S.A.M. M.B., J.S. and S.A.M. analysed the data. J.S. and S.A.M. designed the pupil annotation method. S.A.M. provided funding. M.B., J.Y., C.M.B. and S.A.M. wrote the manuscript, and all authors approved it.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-00963-z.

Correspondence and requests for materials should be addressed to C.M.B., M.B. or S.A.M.

Primary handling editor: Charlotte Payne.

nature research

Corresponding author(s): Samuel Mehr
Last updated by author(s): Aug 11, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Datavyu, QuickTime, Empatica Connect
Data analysis	R, Stata

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Participant-level data and study materials are available at http://osf.io/2t6cy. Audio excerpts from the *Natural History of Song* Discography are available at https://osf.io/vcybz. For assistance with data or materials, please contact S.A.M.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative observational
Research sample	Infants from Cambridge, MA
Sampling strategy	Convenience sample. Power analysis reported in paper to determine sample size.
Data collection	See Methods for full details. Video and audio equipment, parent was present during data collection but blinded to study condition. Researchers were always blind to condition.
Timing	Data were collected between 13 June 2018 and 14 May 2019, continuously.
Data exclusions	All data exclusions are documented in the paper.
Non-participation	No participants dropped out of the paper voluntarily, but participants were excluded on the basis of their behavior (see paper).
Randomization	No random assignment; full counterbalancing into study conditions.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above.
Recruitment	Participants were recruited via web-based advertising, physical flyers, and an existing participant database. Parents could selfselect into this particular study rather than other studies for infants from collaborating labs.
Ethics oversight	Harvard University Committee on the Use of Human Subjects in Research

Note that full information on the approval of the study protocol must also be provided in the manuscript.