

Genomic Imprinting Is Implicated in the Psychology of Music



Samuel A. Mehr^{1,2,3,4}, Jennifer Kotler⁵, Rhea M. Howard¹,
David Haig⁵, and Max M. Krasnow¹

¹Department of Psychology, Harvard University; ²Data Science Initiative, Harvard University; ³Graduate School of Education, Harvard University; ⁴School of Psychology, Victoria University of Wellington; and ⁵Department of Organismic and Evolutionary Biology, Harvard University

Psychological Science
2017, Vol. 28(10) 1455–1467
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797617711456
www.psychologicalscience.org/PS
SAGE

Abstract

Why do people sing to babies? Human infants are relatively altricial and need their parents' attention to survive. Infant-directed song may constitute a signal of that attention. In Prader-Willi syndrome (PWS), a rare disorder of genomic imprinting, genes from chromosome 15q11–q13 that are typically paternally expressed are unexpressed, which results in exaggeration of traits that reduce offspring's investment demands on the mother. PWS may thus be associated with a distinctive musical phenotype. We report unusual responses to music in people with PWS. Subjects with PWS ($N = 39$) moved more during music listening, exhibited greater reductions in heart rate in response to music listening, and displayed a specific deficit in pitch-discrimination ability relative to typically developing adults and children ($N = 589$). Paternally expressed genes from 15q11–q13, which are unexpressed in PWS, may thus increase demands for music and enhance perceptual sensitivity to music. These results implicate genomic imprinting in the psychology of music, informing theories of music's evolutionary history.

Keywords

music, parent-offspring conflict, Prader-Willi syndrome, evolution, open data, open materials

Received 1/27/17; Revision accepted 5/3/17

Music is a ubiquitous part of human culture that predates our recorded history. Musical instruments rank among our most ancient artifacts; bone flutes excavated in Germany are estimated to be about 40,000 years old (Conard, Malina, & Münzel, 2009). The human auditory system is older by an order of magnitude (Quam et al., 2013), and the human vocal-production system is similar to those of many other mammals (Fitch, 2006). It is thus likely that vocal music long preceded instrumental music, as Darwin (1871) and Helmholtz (1885) speculated; although songs leave no fossils, they appear in many small-scale societies, with remarkable diversity (Lomax, 1968).

Whereas humans of all ages produce and enjoy music, a peculiarity of the music faculty is its omnipresence in infancy. Infants have robust music perception abilities (for reviews, see Patel, 2008; Trehub, 2003), and parents often sing to their infants (Custodero & Johnson-Green, 2003) and children (Mehr, 2014) in a stereotyped style (Trehub, Unyk, & Trainor, 1993).

Anthropologists have documented infant-directed song across many cultures (Patel, 2008; Trehub & Trainor, 1998), which is consistent with predictions that it is a universal human behavior (Brown, 1991; Trehub, 2001).

Why do people sing to babies? Mehr and Krasnow (2017) proposed that infant-directed song evolved in humans as a signal of caregiver attention to infants, driven by the dynamics of parent-offspring conflict. In the current study, we tested a key prediction of that theory: that responses to music and music perception abilities should be atypical in populations with disorders of *genomic imprinting*. In these rare disorders, imbalances in the typical expression of maternally versus paternally inherited genes create a natural

Corresponding Author:

Samuel A. Mehr, Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138
E-mail: sam@wjh.harvard.edu

experiment ideal for testing predictions generated from theories of parent-offspring conflict.

Intragenomic Conflict and Disorders of Genomic Imprinting

Evolutionary conflict is predicted between parents and offspring over the preferred amount of parental investment (Trivers, 1974)—that is, parental actions that increase an offspring's reproductive fitness at a cost to the parent (Trivers, 1972). Sexual reproduction causes asymmetric fitness interests in a family, because an offspring is more related to himself than to his siblings, while his mother is equally related to all her offspring. All else being equal, an offspring devalues returns to parental investment made in siblings relative to himself, while a mother values returns to parental investment in all her offspring equally. Further, an offspring's maternally and paternally derived genes have unequal probabilities of being found in his mother's other offspring: That is, a mother is equally related to all of her children, whereas the father of a particular child may be unrelated to some or all of her other children. Paternally derived genes are therefore expected to have been subject to selection favoring the extraction of more maternal resources than is optimal for the mother or for maternally derived genes in offspring (Wilkins & Haig, 2003).

Because every autosomal gene is inherited sometimes from mothers and sometimes from fathers, selection should favor unimprinted genes that perform well when averaged across maternal and paternal roles. However, imprinted genes have effects that differ depending on the sex of most recent parental origin. If the optimal level of gene expression is higher when the gene is inherited from fathers than from mothers, then the gene is predicted to behave as a paternally expressed gene (i.e., it is highly expressed when inherited from fathers and minimally expressed when inherited from mothers). Conversely, if the optimal level of gene expression is higher when the gene is inherited from mothers than from fathers, then the gene is predicted to behave as a maternally expressed gene (Wilkins & Haig, 2003). Typical development evolved to depend on a balance between the antagonistic effects of maternally expressed and paternally expressed genes, whereas maladaptive phenotypes are expected in conditions such as Prader-Willi syndrome (PWS), in which expression of imprinted genes is perturbed (Peters, 2014).

PWS provides a window into the dynamics of parent-offspring intragenomic conflict in humans, because maternally expressed genes from chromosome 15q11–q13 are unopposed by paternally expressed genes from

the same region. In contrast to typical development (Fig. 1a), PWS results from deletion of paternally inherited 15q11–q13 (Fig. 1b), duplication of maternally inherited 15q11–q13 (uniparental disomy; Fig. 1c), or by imprinting center defects (Fig. 1d) that lead to the paternally derived chromosome carrying a maternal imprint (Cassidy, Schwartz, Miller, & Driscoll, 2012). Although these errors lead to pathological, maladaptive changes, they provide clues to the selective forces that have acted on smaller changes in expression over time.

On the *conflict theory* of the evolution of genomic imprinting, symptoms of PWS are expected to nonrandomly cluster in the direction of adaptations that reduce parental investment demands on mothers (Haig & Wharton, 2003; Kotler, Balko, Berall, & Haig, 2016). Infants with PWS demonstrate dramatically reduced nursing and crying and high rates of hypotonia and hypersomnolence (Cassidy & Driscoll, 2009; Holm et al., 1993; Peters, 2014), each of which supports this prediction. In the current study, we explored whether comparable effects were present in responses to music, a domain not previously tested in PWS but potentially relevant to parental investment.

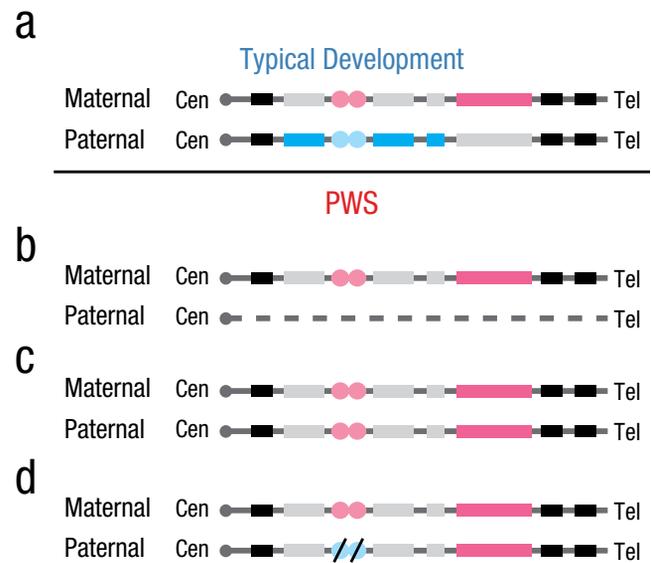


Fig. 1. The genetics of Prader-Willi syndrome (PWS). The diagrams depict chromosome 15q11–q13 in (a) typically developing people and in (b–d) people with PWS. In most cases of PWS, (b) the paternally expressed 15q11–q13 is deleted (approximately 65%–75% of cases) or (c) the maternally expressed 15q11–q13 is duplicated (approximately 20%–30% of cases). In the remaining cases, there are (d) imprinting-center defects (indicated by diagonal black lines), random mutations, or unspecified genetic causes (see Cassidy et al., 2012). Black regions denote unimprinted genes, blue regions denote paternally expressed genes, pink regions denote maternally expressed genes, pink circles denote maternally inherited imprinting centers, blue circles denote paternally inherited imprinting centers, and gray regions denote unexpressed (imprinted) regions. Cen = centromere; Tel = telomere.

Infant-Directed Song as Attentional Investment

The natural environment contains many hazards that human infants are relatively helpless to manage themselves. Instead, infants rely on their caregivers' attentional vigilance to keep them safe from harm, a state of affairs that likely was true for much of human ancestry. But parents have other uses for their attention beyond ensuring infant safety; attention, like food, can be considered a form of parental investment subject to parent-offspring conflict. The amount of parental attention that is optimal is likely to be greater from an infant's perspective than from his or her parent's perspective. Infants are thus expected to attempt to elicit more attention from a parent than the parent wants to give (Trivers, 1972, 1974).

Attention, however, is a covert property of the parent's mind and is not directly perceptible by offspring, and so it must be signaled. Just as arms-race coevolution between peacocks and peahens over the overt signal (plumage) of an unobservable property (good genes) crafted extravagant displays in peacocks and choosiness in peahens (Grafen, 1990), arms-race coevolution between parents and infants over overt signals of attention should result in both elaborate signals in parents and a choosy, demanding appetite for those signals in infants. Studies of people with Angelman syndrome, a genomic imprinting disorder in which paternally expressed genes from 15q11–13 are unopposed (i.e., the pattern opposite that of PWS), support this prediction: Children with Angelman syndrome smile more and make more adult-directed eye contact than age- and cognitively matched children, both of which evoke higher levels of adult attention (Oliver et al., 2007).

One fashion in which parents could signal their attention to infants is by singing to them (for a full discussion, see Mehr & Krasnow, 2017). Song may constitute an honest signal of attention because of opportunity costs (the parent is not interacting with someone else), structural costs (the use of the vocal apparatus for song is incompatible with other strenuous activities), and/or structural features (the parent's proximity and orientation can be inferred when the parent is not fully visible, as at night). If, ancestrally, parental song was a signal of attentional investment, subject to parent-offspring conflict, then the psychology of music could be a target of parent-of-origin effects in intragenomic conflict.

Summary and Predictions

The musical phenotype of people with PWS is thus predicted to differ from that of typically developing people in a fashion consistent with reduced demand

for attentional investment (Mehr & Krasnow, 2017). Reduced demand for attentional investment could manifest itself in one of two ways. First, increased potency of engagement with music and response to music could enable parents to invest relatively less while maintaining the same degree of effect on infants. Second, reduced demand for attention could translate to an overall reduction in musical interest, thereby producing a reduced demand for music. In the present experiments, we tested these hypotheses against each other and against a null hypothesis that response to music in people with PWS is comparable with that in typically developing people.

Two existing findings raise the possibility of an altered musical phenotype associated with PWS, but do not yet distinguish between these hypotheses. When choosing from a list of activities, parents of people with PWS report lower rates of music listening and music performance than do parents of people with Williams or Down syndromes (Sellinger, Hodapp, & Dykens, 2006). When freely listing "nonsport activities" that their children enjoy, parents of people with PWS are less likely to report musical activities than are parents of people with Williams syndrome or Down syndrome (Rosner, Hodapp, Fidler, Sagun, & Dykens, 2004). It is unclear, however, whether reduced parental reporting of active musical activities reflects a general lack of interest in music or a more specifically altered psychology of music in PWS. Indeed, musical behavior is not a unitary domain; it includes a suite of psychological effects and motivations that are mediated by musical ability and a variety of psychophysiological responses.

Thus, we aimed to characterize the musical phenotype in PWS by measuring three variables: (a) increased motion while listening to music, representing an engagement response; (b) reduced heart rate after listening to music, representing a relaxation response; and (c) tone deafness, representing pitch-perception ability. The first hypothesis predicts that people with PWS will engage more with music relative to typically developing people, and the effects should be limited to music (i.e., other pleasant, vocally produced auditory stimuli should not elicit the same effects). The second hypothesis predicts that people with PWS will engage less with music, relative to typically developing people, and the effect should be again limited to music. Both hypotheses predict that people with PWS will relax more following music listening (but not following other auditory stimuli) than would typically developing people. Finally, a deficit in pitch perception would be consistent with both hypotheses, for two different reasons: because reduced discrimination between pitches might produce less choosy listeners, which would reduce demand for high-quality singing (first hypothesis) or

because impaired pitch perception might reduce interest in music in general (second hypothesis).

Method

Subjects

This research was approved by the Committee on the Use of Human Subjects at Harvard University. Subjects with PWS were recruited through the Latham Centers (Brewster, MA), a residential care facility specializing in the treatment of PWS; consequently, clinical characterization of subjects with PWS was provided by Latham Centers along with brief medical histories. Full genetic testing was not available for all subjects; thus, we report analysis of the PWS cohort in the aggregate, as opposed to distinguishing among the three known forms of PWS (i.e., paternal deletion, maternal uniparental disomy, and imprinting-center defects; see Discussion). Data from subjects in the PWS cohort were collected on site with a mobile laboratory.

Typically developing adults were Harvard University undergraduates, graduate students, staff members, and Harvard Summer School students. Typically developing children were recruited via Harvard's Laboratory for Developmental Studies, which obtains subject information from public birth records and recruits via hospitals, schools, museums, and other public spaces. All typically developing subjects were tested in our laboratories at Harvard. All subjects were provided with small gifts in exchange for their participation.

Samples. We studied 39 people with PWS (12 female; mean age = 28.2 years, $SD = 10.1$, age range = 13.7–45.7 years) and compared them with several cohorts of typically developing children and adults (overall $n = 589$; 355 female; mean age = 17.2 years, $SD = 5.75$, age range = 5.16–52.1 years). Both the rarity of PWS in the general population and the associated difficulties of psychological testing in this population precluded us from determining a sample size before conducting the experiments; rather, we tested the maximum number of subjects with PWS that were available to us through the Latham Centers.

We tested children and adults with PWS, as opposed to infants with PWS, because although the musical effects we predict are targeted toward infancy, they likely persist through childhood and adulthood (see discussion in Mehr & Krasnow, 2017). Further, the rarity of PWS (between 1 in 10,000 and 1 in 30,000 live births; Cassidy et al., 2012) makes the testing of infants prohibitively difficult. Data from the PWS cohort ($n = 39$) were collected in two sessions that were about 6 months apart. The first session consisted of pilot testing of a

pitch-perception measure and a listening session; the second session consisted of a battery of cognitive tests and a new pitch test.

Separate typically developing cohorts were recruited for the listening session ($n = 153$; 116 female; mean age = 20.0 years, $SD = 4.96$, age range = 15.5–52.1 years) and for the cognitive battery session (adults: $n = 248$, 146 female, mean age = 20.6, $SD = 2.72$, age range = 17.8–47.9 years; children: $n = 188$, 93 female; mean age = 10.5, $SD = 2.52$, age range = 5.16–17.0 years).

Data exclusions. For heart rate analyses, we excluded data from 2 people with PWS (1 for poor placement of the physiology monitor and 1 for technical error) and from 6 typically developing people (1 for poor placement of the physiology monitor and 5 for technical error). Thus, the sample for the listening-session data included 184 subjects (PWS: $n = 37$; typically developing: $n = 147$), before cleaning of the heart rate data (see the Physiological Monitoring section). Motion analyses were conducted on 160 of the 192 subjects (PWS: $n = 36$; typically developing: $n = 124$) because video and/or stimulus-synchronization data were not available from 29 typically developing subjects and 3 subjects in the PWS cohort, before cleaning of the motion data (see the Motion Analyses section).

From the battery of cognitive tests, data were not available from 2 of the 39 subjects with PWS: 1 declined to participate and the other was hospitalized at the time of testing. We excluded none of the typically developing subjects, but receptive vocabulary scores were missing from 14 subjects because of technical error. Thus, the overall sample for the cognitive battery included 473 subjects (PWS: $n = 37$; typically developing adults: $n = 248$; typically developing children: $n = 188$), but analyses using receptive vocabulary data included 459 subjects (PWS: $n = 37$; typically developing adults: $n = 238$; typically developing children: $n = 184$).

Listening session

Subjects listened to a series of high-fidelity recordings of a female vocalist singing or speaking the lyrics to 12 songs. Songs were chosen so that each subject heard familiar and unfamiliar material in the genres of lullaby, play song, and folk song. Songs varied in length ($M = 45$ s, $SD = 7.5$), but the pre- and poststimulus silent periods were held constant at 10 s each. Each subject listened to 12 recordings of 6 sung tracks and 6 spoken tracks. The 12 recordings were presented in a random order. We designed this method so that we could test whether effects were specific to music listening in particular as opposed to one or more nonmusical features of the auditory stimulus.

Musical content. We selected 12 songs, 8 from infant-directed styles (lullabies or play songs) and 4 from an adult-directed style (folk songs). For half of the repertoire in each style, we chose relatively obscure songs; we thus predicted that half of the songs would be familiar to subjects. In the typically developing cohort, 38% of the songs were reported to be familiar. This percentage was higher in the PWS cohort (50%), but it is unlikely that the PWS cohort's reporting regarding familiarity was accurate: Subjects often reported knowing all the songs, including one that was composed for use in an unrelated experiment (Mehr, Song, & Spelke, 2016), so no subject could possibly have heard it before.

A professional female vocalist with extensive experience in jazz performance and voice acting in productions for children (e.g., cartoons) recorded all stimuli. The recording session was engineered by the technical director of the Harvard University Studio for Electroacoustic Composition and was designed so as to provide as transparent a reproduction of the singer's voice as possible. Recordings were produced in a fully sound-isolated room at high fidelity (96 kHz, 24 bits), using a preamplifier (HV-3D; Millennia Media, Diamond Springs, CA) with two microphones matched in gain and centered in a stereo mix. The microphones were a large-diaphragm condenser (Mouse; Blue Microphones, Westlake Village, CA) and a small-diaphragm condenser (SR40; Earthworks Audio, Milford, NH). This approach gave us the ability to substantially edit the stimuli in postproduction without introducing audio artifacts that would detract from the subjects' experience of hearing a realistic, natural voice during the study.

In this fashion, we produced two versions of each song: The vocalist performed the song in a style suited to the material—a lullaby or play song in an infant-directed manner and a folk song in an adult-directed manner. The vocalist also spoke the lyrics of each song alone, without the melody, also in the style suited to the material. In postproduction, working from the uncompressed audio (Audition; Adobe, San Jose, CA), we made slight adjustments to the lengths of each version, removed breaths, adjusted pitches and rhythms, and performed other miscellaneous edits to improve their quality. Final versions of each stimulus for use in the experiment were created as 320 kbps constant-bit-rate MP3 files to ensure ease of presentation in the mobile laboratory.

Subjects listened to one of four sets of songs that counterbalanced which particular songs were sung and which were spoken. Songs were presented in a random order, with 10 s of silence before and after each song and a short break between trials. During this break, the experimenter asked the subjects to report whether the song was familiar and whether they had enjoyed the

song. All subjects sat in a comfortable chair such that head position was approximately at the apex of a 24-in. equilateral triangle formed with two powered speakers (AV40; M-Audio, Cumberland, RI). The speakers were placed on isolation stands (IsoAcoustics, Markham, ON, Canada), and driven by a Forte audio interface (Focusrite, High Wycombe, Buckinghamshire, United Kingdom). A large photo of a nature scene was placed behind the speakers to maintain visual attention (image courtesy of Shimon Gonen Photography).

Motion analyses. All subjects were recorded with a front-facing camera at 60 frames per second at a resolution of $1,920 \times 1,080$ pixels (1080p). We down-rendered these videos, which resulted in video clips of 10 frames per second at a resolution of 426×240 pixels (240p), and computed a graph-based map of visual saliency (Harel, Koch, & Perona, 2006) for each frame of the video in MATLAB (Figs. 2a and 2b). We computed the absolute value of the change in saliency (i.e., a value from 0 to 1 for each pixel of each frame) and summed these values for each frame. The camera was static, centered on the subject's face, and the subject was seated in front of a static background; thus, changes in saliency values represent subject motion, measured at 10 Hz (Fig. 2c). We also analyzed videos at higher resolutions ($1,280 \times 720$ pixels and 853×480 pixels) to ensure that findings were not attributable to the resolution of the rendering; results were comparable at other resolutions, although we found that at these higher resolutions, large changes in saliency were strongly exaggerated, which necessitated transformations. Thus, we used the 240p videos for final analyses.

To characterize motion during the stimulus, we normalized saliency values to the 10 s before the stimulus, which yielded units interpretable as standard deviations of change in motion. To reduce the impact of extreme values, we removed all observations with z scores greater than 5 or less than -5 ; this resulted in removal of 1.9% of the data but did not alter the total number of subjects. Altering this threshold for trimming did not substantially affect the direction or magnitude of the results. All computations were run on the Odyssey supercomputing cluster supported by the Faculty of Arts and Sciences Division of Science, Research Computing Group at Harvard University.

Physiological monitoring. We measured heart rate using a wrist-mounted physiological monitor (E4; Empatica, Boston, MA) that yields clinical-grade heart rate via photoplethysmography measured at 64 Hz (Garbarino, Lai, Tognetti, Picard, & Bender, 2014). All but 1 subject with PWS tolerated this device. Interbeat intervals were computed by Empatica's proprietary algorithm, which

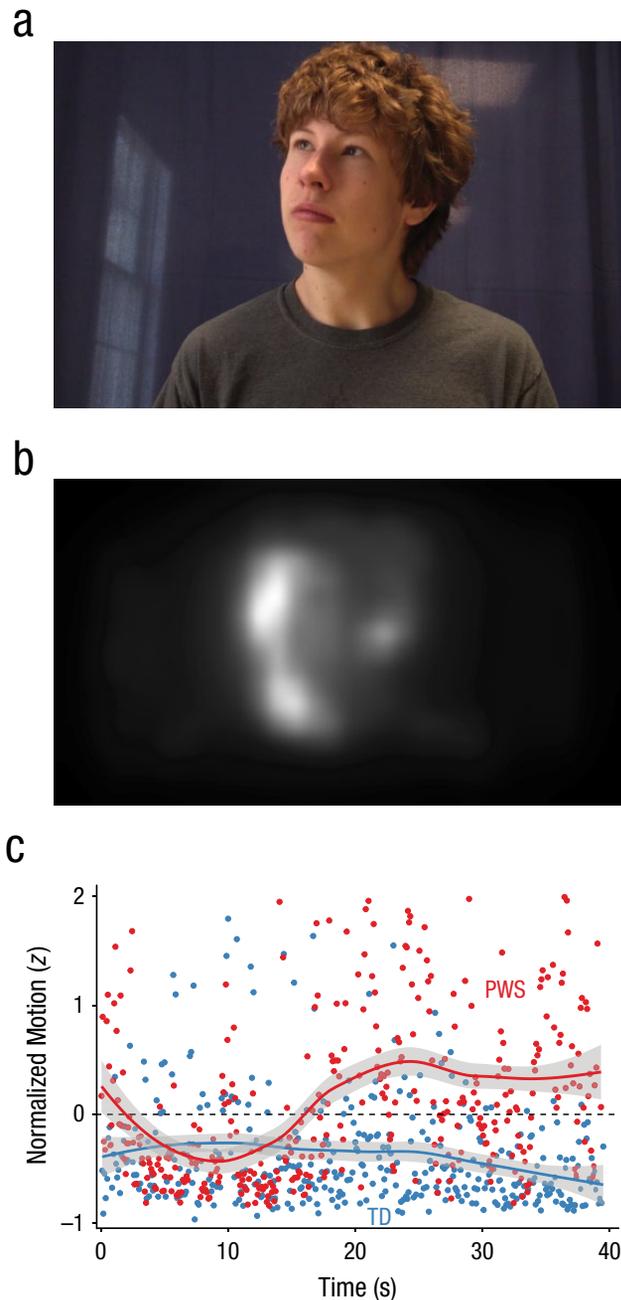


Fig. 2. The method for assessing motion during the listening session. Motion was characterized by measuring the change in total graph-based visual saliency over time. Each frame of a video recorded during each subject's listening session was converted to a saliency map, in which each pixel had a saliency value from 0 to 1. Videos were processed at 10 Hz, and we computed the absolute difference between the sum of each frame's total saliency and that of the previous frame. These values were normalized relative to the 10-s period before the stimulus presentation; thus, larger values indicate more motion. The subject in (a), pictured in one frame from his video, has Prader-Willi syndrome (PWS), and the saliency map in (b) is for that frame, with each pixel's saliency value depicted in grayscale from 0 (black) to 1 (white). The scatterplot in (c) represents the detected motion for both this subject (red dots) and a typically developing subject (blue dots) while each listened to the same lullaby. The lines represent loess regressions for each subject's motion and the shaded areas around the lines represent 95% confidence intervals of the mean.

automatically imputes missing data from the photoplethysmograph signal and corrects for motion artifacts. Subjects wore the device on their nondominant hand and sat without their legs crossed (to maximize blood flow). To minimize motion, we asked subjects to place the device-equipped wrist on the table in front of them. Because the measure of interest was the change in heart rate from before music listening to after music listening, we normalized heart rate values in the 10-s poststimulus period to those in the 10-s prestimulus period from the same trial. Scores may thus be interpreted as standard deviation changes in heart rate as a result of listening to the stimulus.

Heart rate data were missing more often in some subjects than in others because the quality of the physiological monitor's signal was variable across subjects. This necessitated data cleaning. We did so using two methods. First, when fewer than five heart rate observations were present in the prestimulus period, we dropped data from the corresponding poststimulus periods. This decision was based on the assumption that computing a standardized score with fewer than five observations would not yield enough variance to make the standardized score interpretable. This cleaning reduced the sample by 3 subjects in each cohort.

Second, as with the motion data, to reduce the impact of extreme values, we trimmed the raw data set (i.e., before collapsing heart rate data across music and speech trials within each subject) by removing all observations with z scores greater than 5 or less than -5 . This resulted in removal of 1.2% of the data and a single subject in the typically developing cohort. Thus, the final sample of heart rate data was reduced by 3 in the PWS cohort and by 4 in the typically developing cohort, which resulted in final samples of 31 and 140, respectively. Adjusting thresholds for exclusion and trimming did not substantially affect the direction or magnitude of the results.

Whereas some data suggest that PWS is associated with abnormal autonomic nervous system function (e.g., DiMario, Dunham, Burleson, Moskovitz, & Cassidy, 1994), all heart rate measures that we undertook are within subjects; that is, for a given subject, we compared average heart rate during a given interval with average heart rate during the immediately preceding interval. This approach helped to rule out cohort-level bias toward or against our hypotheses.

Assessment of tone deafness

Predicting it unlikely that subjects with PWS would agree to participate in the Montreal Battery for the Evaluation of Amusia (MBEA; Peretz et al., 2008), a commonly used test of tone deafness that takes about

45 min to administer, we piloted a short-form, adaptive test (Loui, Guenther, Mathys, & Schlaug, 2008) with our PWS cohort. This test proved too difficult for subjects, however, and yielded a high rate of attrition. Thus, we designed a new test with reduced length and difficulty, so as to minimize attrition, and with design elements that aimed to maximize the likelihood that even subjects with severe cognitive delays would understand it.

On each trial, a reference tone (A4, or 440 Hz) was played three times, followed by a comparison tone that was either higher or lower in pitch than the reference tone. Subjects were asked “Did that sound go UP or DOWN?” Demonstration trials were included for all subjects, along with practice trials that included corrective feedback. Subjects could repeat practice trials multiple times, to ensure that they understood the test; they began the test only after correctly answering both items. Thirty-two test trials followed, at eight difficulty levels; the size of the interval varied, from three semitones (i.e., 300 cents or 83.3 Hz) to a 64th tone (i.e., 3.125 cents or 0.795 Hz), decreasing in size by half at each difficulty level (in units of cents). The first eight trials were presented at the two easiest difficulty levels, in random order, and the remaining 24 followed, also in random order.

The direction of the pitch difference (up or down) was randomized at every trial. This resulted in an even distribution of trial types when weighted by subject (number of up trials: *Mdn* = 16, interquartile ratio, or IQR = 14–18) and no difference in the distribution of trial types across cohorts (subject-wise number of up trials—PWS: *Mdn* = 17, IQR = 15–18; typically developing cohort: *Mdn* = 16, IQR = 14–18; $z = 1.39$, $p = .165$, Wilcoxon rank-sum test). Moreover, performance was unrelated to the number of up trials in either cohort (PWS cohort: model $\chi^2(1) = 0.28$, $p = .596$; typically developing cohort: model $\chi^2(1) = 2.82$, $p = .093$; ordinal logistic regressions).

Because of the random assignment of direction at each trial, it was also possible that subjects would have different numbers of *switch points* across cohorts (i.e., because, by chance, some subjects necessarily had more switches back and forth between up trials and down trials). To ensure that this randomization did not confound our pitch-test analyses, we tested whether there was any difference in the number of switch points across cohorts. As would be expected in random assignment of 32 trial directions across many subjects, the median number of switch points was 16 (IQR = 14–17), and the rates were comparable across cohorts (PWS cohort: *Mdn* = 16, IQR = 15–17; typically developing cohort: *Mdn* = 15, IQR = 14–17; $z = 1.19$, $p = .236$, Wilcoxon rank-sum test). Moreover, there was no relation between the number of switch points and overall performance on the pitch test in either cohort (PWS cohort: model

$\chi^2(1) = 0.99$, $p = .319$; typically developing cohort: model $\chi^2(1) = 0.18$, $p = .669$; ordinal logistic regressions).

There were two differences in test administration across the typically developing and PWS cohorts: (a) Typically developing subjects completed the test on a computer, with instructions printed in text, whereas subjects with PWS did so with the assistance of an experimenter who read the text aloud; and (b) all typically developing subjects wore headphones to listen to the tones, whereas 8 of the 37 subjects with PWS listened to the tones on speakers, because they expressed discomfort with headphones. No difference in overall performance was detectable in the PWS cohort across listening format, independent-samples $t(35) = 0.62$, $p = .54$. When subjects wore headphones, the experimenter was unaware of the correct response.

Validation study. To validate the new tone deafness test, we administered it along with the MBEA to 271 members of the typically developing cohort ($M = 19.6$ years, $SD = 3.76$, age range = 10.3–47.9 years). Performance on the two tests was significantly correlated, $r = .27$, $p < .0001$. Using the recommended tone deafness diagnostic criterion of 75.9% accuracy on the MBEA (Peretz et al., 2008), we also split the typically developing cohort into two groups to test whether typically developing people who were identified as tone deaf on that test (low-MBEA group; $n = 51$) performed worse on the pitch test than their peers who were not tone deaf (high-MBEA group; $n = 230$). They did, performing an average of 0.52 SD worse on the pitch test: The mean proportion of correct responses on the pitch test for the low-MBEA group was .808 ($SD = .096$, 95% confidence interval, or CI = [.781, .835]); for the high-MBEA group, the mean was .851 ($SD = .079$, 95% CI = [.841, .861]), independent-samples $t(279) = 3.41$, $p = .0008$. Thus, despite its short, simplified format, the pitch test was able to distinguish between subjects who performed well on the MBEA and those who did not, capturing variance attributable to tone deafness.

Battery of cognitive tests. We administered the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007), a standardized test of receptive vocabulary, along with a modified version of Panamath (Halberda, Mazocco, & Feigenson, 2008), a test of approximate number acuity, as control measures. We built the modified Panamath test specifically to match the new pitch test in design. Instead of presenting two arrays of dots simultaneously, as is typical of Panamath (Halberda et al., 2008), we presented a reference array three times, followed by a comparison array, and asked “Were there MORE dots or LESS dots?” As in the pitch test, there were demonstrations and required practice trials with corrective feedback, followed by 32 test trials, 4 each of eight difficulty levels. The first

8 trials, presented in random order, were taken from the two easiest difficulty levels (2 dots:3 dots and 3 dots:4 dots) and the remaining 24 trials were presented in random order, in ratio sizes varying from 5 dots:6 dots to 9 dots:10 dots. Like the pitch test, the test of number acuity was completed with the assistance of an experimenter in the PWS cohort and with printed instructions on a computer in the typically developing cohort. For both cohorts, the experimenter was unable to see the subjects' computer screens during the test of number acuity and thus was unaware of the correct answer and unable to influence subjects' accuracy.

The test of receptive vocabulary was administered according to the instructions provided, but on paper with the assistance of an experimenter for subjects with PWS and on a computer for typically developing subjects. During the testing of the PWS cohort, the experimenter was always aware of the subject's accuracy, to adapt the length of the test in real time (in keeping with the design of the test of receptive vocabulary). Because the test is age-standardized in typically developing populations, not in populations with developmental disabilities, we followed the test authors' recommendations to conduct analyses only on raw test scores (i.e., the total number of correct answers after a baseline criterion set was achieved, stopping when performance was at chance level; Dunn & Dunn, 2007).

Analysis strategy

Statistical analyses were conducted in Stata. For all three types of outcome measures (motion, heart rate, and pitch-perception skill), we used linear regression to examine the relation between cohort and the outcome measure in question. For motion and heart rate analyses, no covariates were used, but for pitch-perception analyses, we included scores on other cognitive tests as covariates (see the Results section; this decision was made before data analysis began). All regressions were bootstrapped with 40,000 replications and were stratified by cohort (PWS cohort vs. typically developing cohort) so as to adjust for the cohorts' different sizes. Last, to ensure that no findings were attributable to the presence of influential observations, we validated all findings with planned sensitivity analyses (Cohen, Cohen, West, & Aiken, 2003).

Results

Engagement response to music

We computed a map of visual saliency (Harel et al., 2006) for each frame of each subject's video and computed the absolute difference in total saliency from frame to frame as a measure of the amount of motion,

tracked at 10 Hz throughout the listening session (Figs. 2a, 2b, and 2c). All subjects' movement, normalized relative to the 10-s prestimulus period (as described in the Method section), was reduced during music listening ($M = -0.20$, $SD = 0.19$, 95% CI = $[-0.24, -0.16]$, in units of standard deviations of change in saliency). The reduction in movement was significantly different from zero, $t(141) = 10.4$, $p < .0001$. This is likely due to the fact that in the baseline period, subjects sat in silence and were given no instruction, so subjects from both cohorts often fidgeted and shifted in their seats. The comparison with baseline serves as a within-subjects normalization to enable a direct comparison of the degree of motion across cohorts during music listening.

People with PWS moved significantly more during music listening ($M = -0.11$, $SD = 0.30$, 95% CI = $[-0.22, -0.02]$) than did typically developing people ($M = -0.23$, $SD = 0.23$, 95% CI = $[-0.27, -0.19]$), model $\chi^2(1) = 4.26$, $p = .039$, $R^2 = .034$, $\beta = 0.44$ (Fig. 3a; for detail on the choice of regression models throughout the Results section, see the Analysis Strategy section). This cohort difference was not present during speech recordings (PWS cohort: $M = -0.23$, $SD = 0.22$, 95% CI = $[-0.30, -0.15]$; typically developing cohort: $M = -0.21$, $SD = 0.21$, 95% CI = $[-0.25, -0.17]$), model $\chi^2(1) = 0.12$, $p = .72$ (Fig. 3b), however, and the within-subjects difference in motion between song and speech recordings also differed between the PWS cohort and the typically-developing cohort, model $\chi^2(1) = 6.20$, $p = .013$, $R^2 = 0.05$, $\beta = 0.51$. Sensitivity analyses revealed one highly atypical observation in the PWS cohort that influenced results toward our predicted effect. To reduce the influence of that observation, we Winsorized it; the analyses presented earlier in this paragraph and in Figure 2 include the Winsorized observation. Analyses using the original value were slightly weaker, but remained statistically significant.

To ensure that the automatic motion detection reflected engagement with music, as opposed to fidgeting or otherwise unmeaningful motion, human coders who were unaware of both our hypothesis and the motion findings reported in the preceding paragraphs viewed all music trials from both the PWS and typically developing cohorts. For each trial, the coders indicated whether the subject moved in a fashion suggesting engagement with music (e.g., swaying to the beat). Trials on which coders specified that such engagement occurred had substantially higher automatically detected motion ($M = 0.16$, $SD = 0.75$, 95% CI = $[0.03, 0.30]$, in units of absolute change in total saliency) than trials on which coders specified that no engagement occurred ($M = -0.24$, $SD = 0.39$, 95% CI = $[-0.26, -0.21]$), independent-samples $t(134.5) = 5.80$, $p < .0001$, $d = 0.84$. Moreover, subjects with PWS were rated as moving in

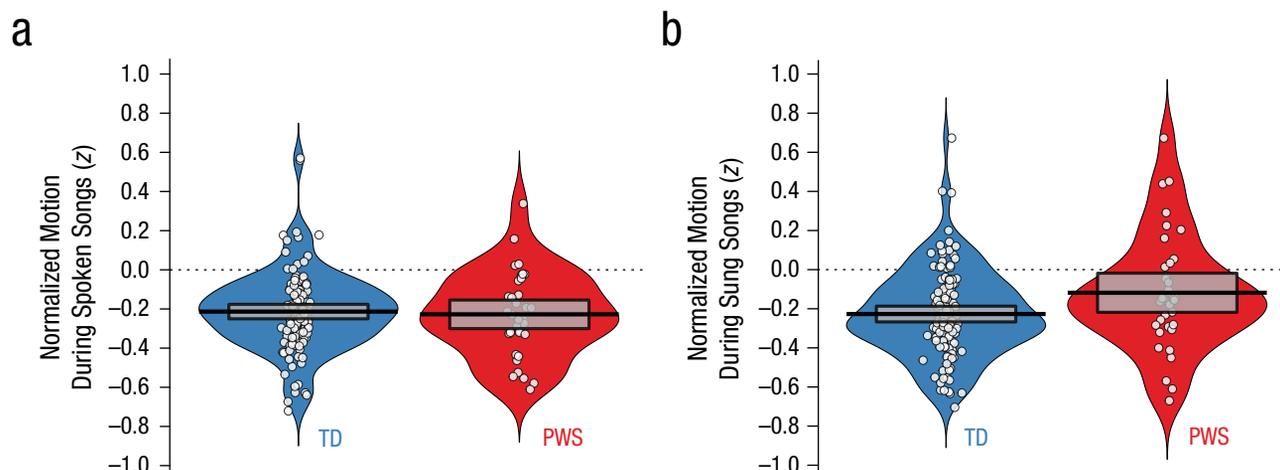


Fig. 3. Normalized motion while subjects (a) listened to spoken songs and (b) listened to sung songs. Results are shown separately for the typically developing cohort and the Prader-Willi syndrome (PWS) cohort. The violin plots depict kernel-density estimations of the normalized motion values, averaged by subject. The horizontal lines indicate the group means. The shaded boxes depict the 95% confidence intervals of the means. The circles show values for individual subjects and are jittered to avoid overlap.

response to music on a higher proportion of trials ($M = .196$, $SD = .280$, 95% CI = [.105, .287]) than were typically developing subjects ($M = .119$, $SD = .195$, 95% CI = [.085, .154]). This comparison was weaker than the automated analysis, given the relatively coarse human-coded measure: Using a common analysis with relaxed assumptions, it trended toward significance, independent-samples $t(163) = 1.92$, $p = .057$, but the result was weaker using a more conservative (but more strictly accurate) ordinal logistic regression, model $\chi^2(1) = 2.54$, $p = .11$, odds ratio = 1.74. (Note that the manually coded analysis includes all 39 subjects in the PWS cohort, whereas the automatic motion-detection analysis was restricted to 36 subjects.) Together, these results demonstrate that subjects with PWS displayed more music-specific engagement than did typically developing subjects.

Relaxation responses to music

All subjects displayed a marked drop in heart rate (normalized relative to the 10-s prestimulus period, as described in the Method section) after listening to singing ($M = -0.55$, $SD = 0.67$, 95% CI = [-0.60, -0.40], in units of standard deviations of heart rate). The change in heart rate was significantly different from zero, $t(171) = 10.1$, $p < .0001$. This drop was significantly larger for the PWS cohort ($M = -0.72$, $SD = 0.67$, 95% CI = [-0.97, -0.48]) than for the typically developing cohort ($M = -0.46$, $SD = 0.65$, 95% CI = [-0.56, -0.35]), model $\chi^2(1) = 4.19$, $p = .041$, $R^2 = 0.025$, $\beta = 0.41$ (Fig. 4a); this analysis used the same regression approach as in the motion analyses. No such difference in heart rate

after listening to speech recordings was found between the PWS cohort ($M = -0.44$, $SD = 0.75$, 95% CI = [-0.70, -0.18]) and the typically developing cohort ($M = -0.50$, $SD = 0.60$, 95% CI = [-0.60, -0.40]), model $\chi^2(1) = 0.17$, $p = .68$ (Fig. 4b). The within-subjects difference in change in heart rate after song listening compared with after speech listening differed significantly between the PWS and typically developing cohorts, model $\chi^2(1) = 3.19$, $p = .037$ (one-tailed), $R^2 = .02$, $\beta = 0.43$. Sensitivity analyses revealed no atypical observations. Thus, subjects with PWS displayed an increased relaxation response after listening to music, but not after listening to highly similar nonmusical auditory stimuli.

Prevalence of tone deafness

The mean proportion of correct answers on the pitch test was substantially lower in the PWS cohort ($M = .657$, $SD = .136$, 95% CI = [.612, .702]) than in the adults in the typically developing cohort ($M = .825$, $SD = .095$, 95% CI = [.817, .834]), model $\chi^2(1) = 70.3$, $p < .0001$, $R^2 = .339$, $\beta = 1.73$, but reviews of medical charts revealed no known hearing impairments in the PWS cohort, and they performed well above chance, $t(36) = 7.05$, $p < .0001$ (comparison with a proportion correct of .5).

However, for receptive vocabulary, the raw scores for the PWS cohort ($M = 167$, $SD = 29$, 95% CI = [158, 177]) were also lower than those for the adults in the typically developing cohort ($M = 212$, $SD = 6.62$, 95% CI = [211, 212]), model $\chi^2(1) = 87.5$, $p < .0001$, $R^2 = .606$, $\beta = 2.28$. A similar pattern was found for approximate number acuity: The proportion correct was lower for

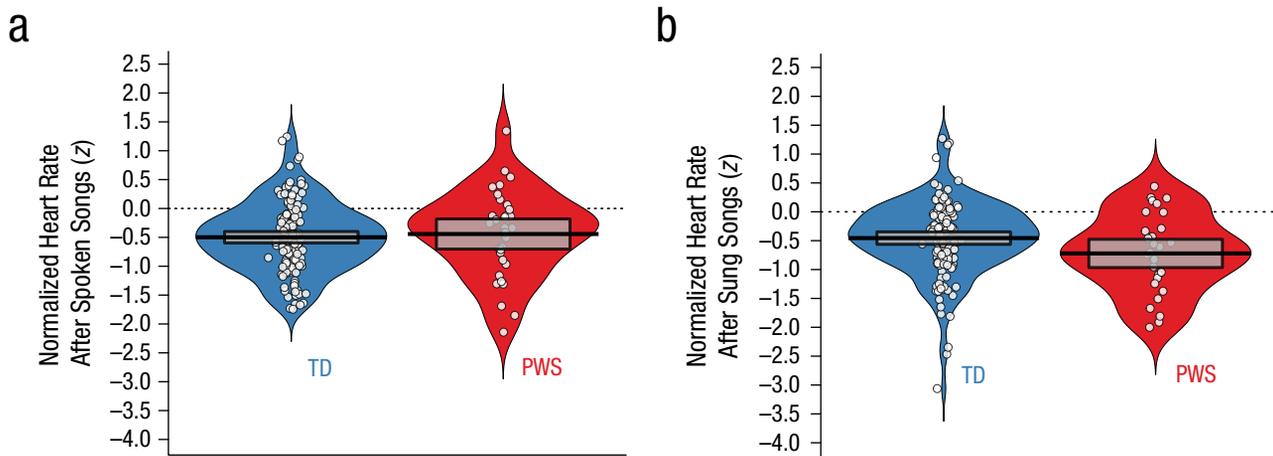


Fig. 4. Normalized heart rate after subjects (a) listened to spoken songs and (b) listened to sung songs. Results are shown separately for the typically developing cohort and the Prader-Willi syndrome (PWS) cohort. Heart rate values were normalized relative to the 10 s before each trial, as in the motion analyses; lower values thus indicate larger reductions in heart rate. The violin plots depict kernel density estimations of the normalized heart rate values, averaged by subject. The horizontal lines indicate the group means. The shaded boxes depict the 95% confidence intervals of the means. The circles show values for individual subjects and are jittered to avoid overlap.

the PWS cohort ($M = .661$, $SD = .131$, 95% CI = [.618, .705]) than for the adults in the typically developing cohort ($M = .892$, $SD = .079$, 95% CI = [.882, .902]), model $\chi^2(1) = 114$, $p < .0001$, $R^2 = .442$, $\beta = 1.97$.

Consequently, we tested an additional group of typically developing children (age range = 5–17 years; $n = 188$) with the same measures to assess the degree to which low performance on the pitch test was attributable to general cognitive deficits in the PWS cohort. This new group's scores for receptive vocabulary ($M = 173$, $SD = 28.5$, 95% CI = [169, 177]) were comparable with those of the PWS cohort ($M = 167$, $SD = 29.1$, 95% CI = [158, 177]), model $\chi^2(1) = 1.32$, $p = .251$. However, on the pitch test, the proportion correct for the PWS cohort ($M = .657$, $SD = .136$, 95% CI = [.612, .702]) was substantially lower than that of the typically developing children ($M = .800$, $SD = .104$, 95% CI = [.784, .815]), model $\chi^2(1) = 37.6$, $p < .0001$, $R^2 = .191$, $\beta = 1.17$. This result held after we included receptive vocabulary score as a covariate (Fig. 5a), $\chi^2(2) = 103$, $p < .0001$, $R^2 = .347$. In this model, PWS was associated with an average reduction of 1.07 SD in performance on the pitch test, Wald $\chi^2(1) = 39.5$, $p < .0001$. In this analysis, the five lowest scores on the test of receptive vocabulary among the typically developing children were clustered in an atypical fashion compared with the 183 other subjects in the typically developing cohort. This atypical cluster reduced fit in the overall model by 9.8%; given the large size of the typically developing cohort, we opted to remove the data for these 5 subjects from the analyses reported earlier in this paragraph and shown in Figure 5 instead of Winsorizing them (see Cohen et al., 2003).

Models retaining the original values were statistically significant.

We also compared the pitch-test performance of the PWS cohort with that of a subsample of typically developing children ($n = 46$; mean age = 7.05 years, $SD = 1.03$, range = 5.16–8.53 years). The subsample was of the lowest-performing typically developing children on the test of approximate number acuity, such that their scores ($M = .690$, $SD = .120$, 95% CI = [.654, .725]) were comparable with those of the PWS cohort ($M = .661$, $SD = .131$, 95% CI = [.618, .705]), model $\chi^2(1) = 1.06$, $p = .303$. This comparison is particularly informative, because we built the test of number acuity to match the structure of the pitch test (e.g., with three reference arrays and a comparison-test array; see the Method section). Here again, the PWS cohort scored significantly lower on the pitch test ($M = .657$, $SD = .136$, 95% CI = [.612, .702]) than did these typically developing children ($M = .744$, $SD = .129$, 95% CI = [.706, .782]), model $\chi^2(1) = 8.93$, $p = .0028$, $R^2 = .099$, $\beta = 0.63$ (Fig. 5b). This result held when we included number acuity as a covariate (Fig. 5b), model $\chi^2(2) = 27.6$, $p < .0001$, $R^2 = .199$. In this model, PWS was associated with an average reduction of 0.56 SD in performance on the pitch test, Wald $\chi^2(1) = 7.22$, $p = .007$. Sensitivity analyses revealed no atypical observations.

Finally, the results of a multiple regression comparing performance on the pitch test between the full cohort of typically developing children and the PWS cohort, adjusted for performance on both of the nonmusical cognitive tests, was significant, model $\chi^2(3) = 90.9$, $p < .0001$, $R^2 = .328$; PWS was associated

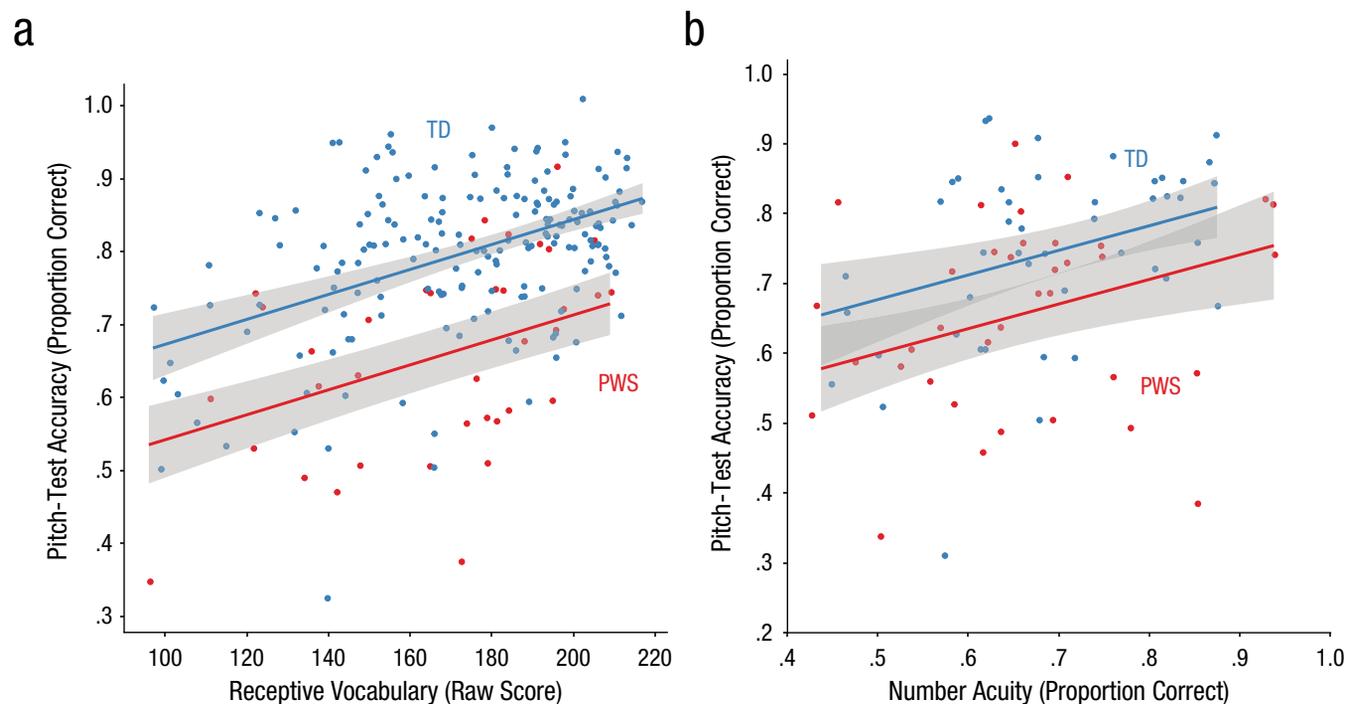


Fig. 5. Scatterplots showing the relation between performance on the pitch test and (a) receptive-vocabulary score and (b) approximate-number-system acuity in the Prader-Willi syndrome (PWS) cohort (red dots) and two cohorts of typically developing children (blue dots). The graphed lines indicate the predicted values for each group, derived from a bootstrapped multiple regression with stratified resampling; the shaded areas above and below the lines indicate the bootstrapped 95% confidence bands for each regression.

with an average reduction of 0.99 *SD* in performance on the pitch test, Wald $\chi^2(1) = 23.2$, $p < .0001$.

Together, these results suggest that the reduction in pitch-perception ability associated with PWS was not fully accounted for by receptive-vocabulary level or by numerical ability, as assessed using methods highly similar to those of the pitch test. Moreover, the relation between our pitch test and a standardized measure of tone deafness in a validation sample (see the Method section) suggests that PWS is associated with increased prevalence of tone deafness.

Discussion

People with PWS showed increased engagement (more motion) during music listening and increased relaxation (lower heart rate) after music listening, along with reduced pitch-perception ability. This confirms a key theoretical prediction concerning the evolution of infant-directed song (Mehr & Krasnow, 2017, p. 680). Engagement and relaxation effects are interpretable as traits that decrease demand for music, such that the same amount of song produces larger effects in subjects with PWS than in typically developing subjects. Reduced pitch-perception ability has a more speculative interpretation: Infants with less sensitivity to pitch may be

less choosy about the quality of the singing required to elicit a calming response, further reducing demands for parental investment. Together, these effects implicate intragenomic conflict in the psychology of music.

Given the association between tone deafness and self-reports of overt musical behaviors and of skill in music performance (Peretz et al., 2008), future research aimed at fully characterizing the range of music perception deficits in people with PWS should include tests of music production in addition to pitch-perception and rhythm-perception tests. Indeed, in the course of this work, we obtained several anecdotal reports of poor music performance skill in people with PWS.

Moreover, demonstrating a converse effect in a population with a disorder of paternally derived expression, especially Angelman syndrome, would offer strong support for the role of intragenomic conflict in the evolution of infant-directed song. Future work using the same methods to test musical response and perception in people with nonimprinted genomic disorders (e.g., Down syndrome) would also help determine the specificity of the effects seen in PWS. In particular, this would help to determine the robustness and specificity of the deficit of pitch perception.

If music effectively increases relaxation in people with PWS, then clinicians, educators, and families may

want to incorporate more music into their behavioral management and care. In health-care settings, listening to music improves well-being in typically developing people (Hole, Hirsch, Ball, & Meads, 2015), and it may be particularly well suited to the treatment of people with PWS, especially considering the need for interventions addressing aggressive and violent behavioral outbursts (Butler, Lee, & Whitman, 2006). A more complete understanding of the ultimate mechanisms responsible for music's effects on arousal and behavior will provide a basis for developing evidence-based music therapies grounded in basic science.

Action Editor

Steven W. Gangestad served as action editor for this article.

Author Contributions

S. A. Mehr and M. M. Krasnow conceived of the study and designed it with J. Kotler and D. Haig. S. A. Mehr and M. M. Krasnow wrote code for the experiments and data acquisition. S. A. Mehr and J. Kotler collected listening-session data with the cohort with Prader-Willi syndrome (PWS). S. A. Mehr administered the cognitive battery with the PWS cohort. R. M. Howard coordinated the research and supervised research assistants who collected all data from the typically developing cohort. S. A. Mehr and M. M. Krasnow designed the analyses. S. A. Mehr conducted the analyses. All the authors wrote the manuscript and approved the final version for publication.

Acknowledgments

We thank all the subjects and their families for taking part in this research. We thank the staff and leadership of the Latham Centers of Brewster, Massachusetts, including Anne McManus, Gerry Pouillot, Scott Fitzgerald, Brittni Taylor, Kathleen Van Esselstyn, and Kara McDowell for supporting this work and coordinating the testing of subjects with PWS; James Cuff, Robert Freeman, Plamen Krastev, Brian White, and the staff at Harvard's Odyssey supercomputing cluster for assistance with motion analysis; April Hall and Seth Torres for performing and engineering the recording session; Dan Bender and Empatica Research for assistance with physiological monitoring; and Alena Egner, Zachary Smith, Erin Mernoff, Adrianna Ratajska, Emma Hardimon, Marcella Montagnese, Terry Lee, Katie Tutrone, and Mark Liu for research assistance.

Declaration of Conflicting Interests

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

Open Practices



All data and materials have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/gzyth> and <https://osf.io/dvmjr>, respectively.

The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797617711456>. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <https://www.psychologicalscience.org/publications/badges>.

References

- Brown, D. E. (1991). *Human universals*. Philadelphia, PA: Temple University Press.
- Butler, M. G., Lee, P. D. K., & Whitman, B. Y. (2006). *Management of Prader-Willi syndrome*. New York, NY: Springer-Verlag.
- Cassidy, S. B., & Driscoll, D. J. (2009). Prader-Willi syndrome. *European Journal of Human Genetics*, *17*, 3–13. doi:10.1038/ejhg.2008.165
- Cassidy, S. B., Schwartz, S., Miller, J. L., & Driscoll, D. J. (2012). Prader-Willi syndrome. *Genetics in Medicine*, *14*, 10–26. doi:10.1038/gim.0b013e31822bead0
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge.
- Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature*, *460*, 737–740. doi:10.1038/nature08169
- Custodero, L. A., & Johnson-Green, E. A. (2003). Passing the cultural torch: Musical experience and musical parenting of infants. *Journal of Research in Music Education*, *51*, 102–114. doi:10.2307/3345844
- Darwin, C. (1871). *The descent of man*. London, England: Watts.
- DiMario, F. J., Dunham, B., Burleson, J. A., Moskovitz, J., & Cassidy, S. B. (1994). An evaluation of autonomic nervous system function in patients with Prader-Willi syndrome. *Pediatrics*, *93*, 76–81.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Minneapolis, MN: Pearson Assessments.
- Fitch, W. T. (2006). Production of vocalizations in mammals. *Visual Communication*, *3*, 145–175.
- Garbarino, M., Lai, M., Tognetti, S., Picard, R., & Bender, D. (2014). Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In K. S. Nikita, N. Bourbakis, B. Lo, D. I. Fotiadis, Y. Hao, & A. Kiourti (Eds.), *4th International Conference on Wireless Mobile Communication and Healthcare* (pp. 39–42). doi:10.4108/icst.mobihealth.2014.257418
- Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*, 517–546. doi:10.1016/S0022-5193(05)80088-8
- Haig, D., & Wharton, R. (2003). Prader-Willi syndrome and the evolution of human childhood. *American Journal of Human Biology*, *15*, 320–329. doi:10.1002/ajhb.10150
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668. doi:10.1038/nature07246

- Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. In B. Schölkopf, J. C. Platt, & T. Hoffman (Eds.), *Proceedings of the 19th International Conference on Neural Information Processing Systems* (pp. 545–552). Retrieved from <http://resolver.caltech.edu/CaltechAUTHORS:20160315-111145907>
- Helmholtz, H. von. (1885). *The sensations of tone as a physiological basis for the theory of music*. London, England: Longmans, Green.
- Hole, J., Hirsch, M., Ball, E., & Meads, C. (2015). Music as an aid for postoperative recovery in adults: A systematic review and meta-analysis. *The Lancet*, *386*, 1659–1671. doi:10.1016/S0140-6736(15)60169-6
- Holm, V. A., Cassidy, S. B., Butler, M. G., Hanchett, J. M., Greenswag, L. R., Whitman, B. Y., & Greenberg, F. (1993). Prader-Willi syndrome: Consensus diagnostic criteria. *Pediatrics*, *91*, 398–402.
- Kotler, J., Balko, K., Berall, G., & Haig, D. (2016). Nutritional phases in Prader-Willi syndrome: Evolutionary and clinical interpretations. *Journal of Evolutionary Medicine*, *4*, 1–7. doi:10.4303/jem/235968
- Lomax, A. (1968). *Folk song style and culture*. Washington, DC: American Association for the Advancement of Science.
- Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, *18*, R331–R332. doi:10.1016/j.cub.2008.02.045
- Mehr, S. A. (2014). Music in the home: New evidence for an intergenerational link. *Journal of Research in Music Education*, *62*, 78–88. doi:10.1177/0022429413520008
- Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution & Human Behavior*, *38*, 674–684.
- Mehr, S. A., Song, L. A., & Spelke, E. S. (2016). For 5-month-old infants, melodies are social. *Psychological Science*, *27*, 486–501. doi:10.1177/0956797615626691
- Oliver, C., Horsler, K., Berg, K., Bellamy, G., Dick, K., & Griffiths, E. (2007). Genomic imprinting and the expression of affect in Angelman syndrome: What's in the smile? *Journal of Child Psychology and Psychiatry*, *48*, 571–579. doi:10.1111/j.1469-7610.2007.01736.x
- Patel, A. D. (2008). *Music, language, and the brain*. New York, NY: Oxford University Press.
- Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., . . . Bouchard, B. (2008). Online identification of congenital amusia. *Music Perception*, *25*, 331–343. doi:10.1525/mp.2008.25.4.331
- Peters, J. (2014). The role of genomic imprinting in biology and disease: An expanding view. *Nature Reviews Genetics*, *15*, 517–530. doi:10.1038/nrg3766
- Quam, R. M., de Ruiter, D. J., Masali, M., Arsuaga, J.-L., Martínez, I., & Moggi-Cecchi, J. (2013). Early hominin auditory ossicles from South Africa. *Proceedings of the National Academy of Sciences, USA*, *110*, 8847–8851. doi:10.1073/pnas.1303375110
- Rosner, B. A., Hodapp, R. M., Fidler, D. J., Sagun, J. N., & Dykens, E. M. (2004). Social competence in persons with Prader-Willi, Williams and Down's syndromes. *Journal of Applied Research in Intellectual Disabilities*, *17*, 209–217. doi:10.1111/j.1468-3148.2004.00200.x
- Sellinger, M. H., Hodapp, R. M., & Dykens, E. M. (2006). Leisure activities of individuals with Prader-Willi, Williams, and Down syndromes. *Journal of Developmental and Physical Disabilities*, *18*, 59–71. doi:10.1007/s10882-006-9006-8
- Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences*, *930*, 1–16.
- Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, *6*, 669–673. doi:10.1038/nn1084
- Trehub, S. E., & Trainor, L. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research*, *12*, 43–78.
- Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Adults identify infant-directed music across cultures. *Infant Behavior & Development*, *16*, 193–211. doi:10.1016/0163-6383(93)80017-3
- Trivers, R. L. (1972). Parental investment and sexual selection. In B. G. Campbell (Ed.), *Sexual selection and the descent of man* (pp. 136–179). Chicago, IL: Aldine.
- Trivers, R. L. (1974). Parent-offspring conflict. *American Zoologist*, *14*, 249–264. doi:10.1093/icb/14.1.249
- Wilkins, J. F., & Haig, D. (2003). What good is genomic imprinting: The function of parent-specific gene expression. *Nature Reviews Genetics*, *4*, 359–368. doi:10.1038/nrg1062