Social Functions of Music in Infancy

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Mehr, Samuel A. 2017. Social Functions of Music in Infancy. Doctoral dissertation, Harvard Graduate School of Education.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:33052842">http://nrs.harvard.edu/urn-3:HUL.InstRepos:33052842</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Social Functions of Music in Infancy

Samuel A. Mehr

Howard Gardner
Catherine Snow
Steven Pinker
Elizabeth Spelke

A Thesis Presented to the Faculty
of the Graduate School of Education of Harvard University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

May, 2017
Acknowledgments

First, I thank the members of my thesis committee, Howard Gardner, Catherine Snow, Steven Pinker, and Elizabeth Spelke, for their support.

Second, I thank Natasha Kalra, Ellyn Schmidt, Rosa Guzman, Bryan Zuluaga, Olivia Sommer, Veri Seo, Amy Bu, Marina Ebert, Alena Egner, Joseph Anthony, Erica Washburn, Rachel Katz, Adena Schachner, Caroline Murphy, Lee Ann Song, and Sophia Sanborn, for their assistance in carrying out this research. For their financial support, I thank the Dana Foundation and the Harvard University Mind/Brain/Behavior Interfaculty Initiative.

Third, I am grateful to the following people, listed in the rough chronological order in which they played a part in my academic life: my parents, Helen Epstein and Patrick Mehr; Jeffrey Leonard; Donna Brink Fox; Richard Aslin and Elissa Newport; Elizabeth Spelke and Adena Schachner; Jesse Snedeker, Shilpa Mody, and Alek Chakroff; Ellen Winner; Howard Gardner; the members of the Laboratory for Developmental Studies; Steven Pinker; David Haig; the members of the Evolutionary Psychology Laboratory; Jerome Kagan; Richard Chang; Max Krasnow; Manvir Singh and Luke Glowacki; Daniel Ketter and Daniel Pickens-Jones; and my partner, Alia Martin.

Last, my heartfelt thanks to my mentors, Howard Gardner, Steve Pinker, Max Krasnow, and especially Elizabeth Spelke, for making me into a scientist.
# Table of Contents

**Acknowledgments**  
Page i

**Table of Contents**  
Page ii

**Abstract**  
Page iii

**Chapter 1. Introduction**  
Page 1

**Chapter 2. For five-month-old infants, melodies are social**  
Introduction  
Page 8  
Experiment 1  
Page 11  
Experiment 2  
Page 20  
Experiment 3  
Page 27  
Experiment 4  
Page 37  
Experiment 5  
Page 41  
General Discussion  
Page 46

**Chapter 3. Shared musical knowledge in 11-month-old infants**  
Introduction  
Page 50  
Method  
Page 51  
Results  
Page 55  
Discussion  
Page 59

**Chapter 4. Conclusion**  
Some methodological concerns  
Page 73  
How to resolve these concerns  
Page 75  
A note on interpretation  
Page 77  
Implications for education  
Page 80

**Bibliography**  
Page 84
Abstract

I explore music's early role in social cognition, testing the hypothesis that infants interpret singing as a social signal. Over six experiments, I examine 5- and 11-month-old infants' social responses to new people who sing familiar or unfamiliar songs to them. I manipulate song familiarity with three training methods: infants learn songs from a parent; from a musical toy; or from an unfamiliar adult who sings first in person and subsequently via video chat. I use two main outcome measures: a test of visual preference for the singer of a familiar song; and, in older infants, a more explicitly social test of selective reaching for objects associated with and endorsed by novel individuals. I also test infants' memory for the songs they hear in these studies. I find that infants garner social information from the songs they hear, which they subsequently act upon in the context of social interaction; when songs are not learned in a social context, infants recall them in great detail after long delays. These results demonstrate a social function of music in early development. Music is not just pleasurable noise: it is a member of a class of behaviors, including language, accent, and food preference, that reliably inform infants' social behavior.
Chapter 1

Introduction

Throughout human history, music has been a ubiquitous piece of human culture, and today, it is difficult to be near other humans without hearing music. Music supports a massive industry with yearly sales of recordings and concert tickets nearing $50 billion (PriceWaterhouseCoopers, 2016). Governments worldwide spend public funds on music education (Parsad & Spiegelman, 2012). Music is so common in television and film that its absence is used as a comedic tool. At the time of this writing, the most popular videos on the internet are music performances¹, content commanding so much interest that it drives the design of web infrastructure: when YouTube's most popular video of all time ("Gangnam Style"; PSY, 2012) was watched 2.15 billion times, the site was forced to upgrade its interface to accommodate 64-bit integers (Savov, 2014). And music is frequently incorporated into non-musical technologies: it is a safe assumption that readers of this thesis have, on their person or within arms' reach, an electronic device capable of storing, playing back, and even composing music.

This ubiquity is not of recent origin. Musical instruments are among our most ancient cultural artifacts, with bone flutes excavated in Germany estimated

¹ As of this writing, YouTube's top 70 videos have been watched 88.4 billion times, which works out to over 275 music video views per second for ten years. This estimate is a lower bound, as it only includes the listing at https://en.wikipedia.org/wiki/List_of_most_viewed_YouTube_videos.
to be about 40,000 years old (Conard, Malina, & Münzel, 2009). The human auditory system is an order of magnitude older than such flutes (Martínez et al., 2004; Quam et al., 2013) and the human vocal production system is similar to that of many other mammals (Fitch, 2006). Given these requisite substrates, Darwin (1871) and Helmholtz (1885) were likely correct in their speculation that vocal music long predated instrumental music. Indeed, while songs leave no fossils, they are found in countless small-scale societies, and exhibit remarkable diversity (Lomax, 1968, 1977; Mehr et al., in prep; Merriam, 1964; Nettl, 2015).

Whereas humans of all ages produce and consume music, a peculiarity of the music faculty is its omnipresence in infancy and childhood. In modern environments, parents sing to their infants (Custodero & Johnson-Green, 2003) and to their children (Mehr, 2014), much like parents in small-scale societies. This practice has been proposed to be a human universal (Brown, 1991) and descriptions of it can be found in the ethnographies of many world cultures. For instance, in a database of ethnography taken from a geographically-stratified random sample of human cultures, text pertaining to singing appears alongside text concerning "Infancy and childhood" in 54 of 60 societies².

---

² This finding can be replicated by conducting a search of the Human Relations Area Files (Murdock et al., 2008; available at http://hraf.yale.edu) for Probability Sample File cultures containing OCM Code 850 ("Infancy and childhood") and the keywords SONG SONGS SING SINGS SINGER SINGERS SINGING SANG SUNG LULLABY LULLABIES. Further details on the Probability Sample File are available in Naroll (1967).
Further, infant-directed song is present in both the Hadza of Tanzania and the !Kung San of Southern Africa’s Kalahari desert, the two most distantly related human groups currently known (Knight et al., 2003). Konner (1972) reports that San mothers sing to their infants in an effort to quiet them:

If the infant is with a caretaker other than the mother, the caretaker will make a brief attempt at quieting and then carry the infant to the mother. Unless the infant is completely satiated the mother almost always responds by trying to nurse him ... If nursing is ineffective or partially effective, rocking and singing are the next responses, often with the infant pressed, front-to-front, against her chest and shoulder. (p. 293, emphasis added)

Marlowe (2010) reports a similar behavior in Hadza fathers:

All caretakers appear to be equally sensitive to fussing and crying, but the mother is far more effective at soothing the child. However, it is usually the father who holds a crying infant in the middle of the night and sings to get the infant to go back to sleep. (pp. 206-207, emphasis added)

Infant-directed song constitutes a distinct category of vocalization, much like infant-directed speech, a well-understood form of language present in many cultures (Broesch & Bryant, 2015; Ferguson, 1964; Fernald, 1984, 1992; Fernald & Simon, 1984; Snow, 1972, 1977; Snow & Ferguson, 1977; Werker, Pegg, & McLeod, 1994). Adult listeners distinguish infant-directed songs from other singing, even when examples are taken from unfamiliar cultures (Trehub, Unyk, & Trainor, 1993a). Moreover, adults attend to the features of infant-directed song in sufficient detail to distinguish reliably between its true form (recordings of a
parent singing directly to an infant) and a simulated form (recordings of the same parent singing alone but imagining that his or her infant is present) (Trehub, Unyk, & Trainor, 1993b). Independent listeners rate such simulated forms as less emotionally engaging than true infant-directed song (Trehub, Hill, & Kamenetsky, 1997).

Moreover, decades of research in the developmental science of music demonstrate that infants are well-prepared to learn from the music they hear (reviews: Patel, 2008; Trehub, 2000, 2001, 2003). Young infants demonstrate impressive musical abilities, decoding auditory input into rhythms (Trehub & Thorpe, 1989) and decoding melodies into relative pitch contours (Chang & Trehub, 1977). As with the development of speech perception and face perception, infant music perception abilities show signs of perceptual narrowing: twelve-month-old infants respond to rhythms in a fashion consistent with adults from their culture, but six-month-old infants demonstrate more culture-general responses (Hannon & Trehub, 2005). Neonates are sensitive to musical meter, as they develop expectations for audible downbeats3, demonstrated by large changes in event-related brain potentials when musical events are removed from...

---

3 The downbeat is the first position of a rhythmic cycle, as in the first beat of a measure in 4/4 time (for example: the words “row”, “gently”, “merrily”, and “life” in the folk song Row, Row, Row Your Boat). Downbeats are often marked by a salient musical event, such as a stable pitch class (i.e., tonic) and/or a rhythmic accent (for discussion, see Lerdahl & Jackendoff 1983).
downbeats (more so than when they are removed from other beats; Winkler, Háden, Ladinig, Sziller, & Honing, 2009). Infants also remember the music they hear: one-month-olds recall a melody heard while in utero, with sufficient detail to discriminate it from a second melody (Granier-Deferre, Bassereau, Ribeiro, Jacquet, & DeCasper, 2011), and long-term memory for music has repeatedly been demonstrated later in infancy (Saffran, Loman, & Robertson, 2000; Trainor, Wu, & Tsang, 2004; Volkova, Trehub, & Schellenberg, 2006). Infants are highly motivated to listen to music, moving spontaneously to it (Zentner & Eerola, 2010), and tolerating listening to it longer than they tolerate listening to either adult-directed or infant-directed speech (Corbeil, Trehub, & Peretz, 2016).

In sum: convergent research from across the cognitive sciences demonstrates that the infant mind is equipped with complex machinery for attending to and processing music; the adult mind is equipped with complex machinery for producing music; and both are inclined to engage with music in the context of parent-infant interaction. Why should this be?

In this thesis I present work exploring the possibility that the answer to these questions lies in the social domain. Specifically, I propose a social function for music in infancy: that infants garner social information from the songs they hear and use that information to guide their future social interactions. If so, music can be counted among other behaviors that inform infants' social behavior,
such as the speech infants hear (e.g., Kinzler, Dupoux, & Spelke, 2007) and the food-related behaviors infants observe (Liberman, Woodward, Sullivan, & Kinzler, 2016). In Chapters 2 and 3, I report experiments testing this hypothesis. In Chapter 4, I conclude by discussing some methodological issues, potential proximate- and ultimate-level explanations for the results, and the relevance of this work to educational research and practice.
Chapter 2

For five-month-old infants, melodies are social

Mehr, Song, & Spelke (2016)4

For 1 to 2 weeks, 5-month-old infants listened at home to one of two novel songs with identical lyrics and rhythms, but different melodies; the song was sung by a parent, emanated from a toy, or was sung live by a friendly but unfamiliar adult first in person and subsequently via interactive video. We then tested the infants’ selective attention to two novel individuals after one sang the familiar song and the other sang the unfamiliar song. Infants who had experienced a parent singing looked longer at the new person who had sung the familiar melody than at the new person who had sung the unfamiliar melody, and the amount of song exposure at home predicted the size of that preference. Neither effect was observed, however, among infants who had heard the song emanating from a toy or being sung by a socially unrelated person, despite these infants’ remarkable memory for the familiar melody, tested an average of more than 8 months later. These findings suggest that melodies produced live and experienced at home by known social partners carry social meaning for infants.

4 This chapter appears in Psychological Science with the same title as above. For consistency with the published version, I have retained the use of plural first-person pronouns throughout.
Introduction

Music is a putative human universal (Brown, 1991) that is ancient (Conard, Malina, & Münzel, 2009), highly variable (Lomax, 1968), and captivating throughout the life span, including in infancy (Patel, 2008; Trehub, 2003). Human infants learn and remember melodies (Trainor, Wu, & Tsang, 2004), are sensitive to rhythmic detail (Winkler, Háden, Ladinig, Sziller, & Honing, 2009), and move spontaneously to music (Zentner & Eerola, 2010). Parents sing frequently to their infants and children (Custodero & Johnson-Green, 2003; Mehr, 2014), in a style that is identifiable across many cultures (Trehub, Unyk, & Trainor, 1993). Why do parents sing to their infants? Why do infants engage with songs sung by their parents? These questions echo long-standing debates over music’s psychological functions in modern environments and origins in human ancestry (e.g., Darwin, 1871; James, 1890; Spencer, 1857). Here, we explore the hypothesis that melody— a salient feature of vocal music—conveys social information to infants.

As do adults, infants regularly choose whether and how to engage with other individuals, selectively attending to novel people who look at them (Farroni, Csibra, Simion, & Johnson, 2002), produce infant-directed speech (Schachner & Hannon, 2011), or speak their parents’ language (Kinzler, Dupoux, & Spelke, 2007). Such people are likely members of an infant’s social group, and
infants might be sensitive to this information as a result of innate dispositions, early learning, or both. We therefore examine whether infants selectively attend to novel singers of melodies learned in social settings.

Melody could signal social affiliation for two reasons. First, before the advent of sound recordings, melodies were learned only from other people. Because melodies are complex and highly variable, members of disjoint groups are unlikely to invent the same melodies. The ability and propensity to sing a familiar song could therefore carry social information: A novel person who sings a known melody is more likely to be connected to one’s social group than is a novel person who sings an unknown melody. Second, in many cultures, people sing together in social contexts (Savage, Brown, Sakai, & Currie, 2015). Infants may be attentive to this fact, and these social experiences may lead them to endow music with social meaning.

Thus, we hypothesized that infant-directed songs convey social meaning to infants. In the experiments reported here, we examined infants’ preferences for novel individuals, one who had sung a familiar melody and another who had sung an unfamiliar melody. The infants were familiar with one of the songs as a result of exposure to it in one of three forms: (a) live song from a parent in the home, (b) recorded song emanating from a toy presented and activated by a
parent in the home, or (c) live song presented by a novel adult both in person in the lab and via live, interactive video at home.

These three forms of song exposure have different social implications. A musical toy presented by a parent divorces the source of music from the source of social interaction, while maintaining a socially engaging context provided by the parent, who activates the toy and modulates the infant’s engagement with it. Live, interactive song presented by an unfamiliar adult involves social interaction between the singer and infant but divorces the source of music from any history of social interaction; when the singing episodes occur at home over interactive video, they are also divorced from other social interactions in the home.

Nevertheless, each of these forms of song exposure may carry social meaning. Infants reliably learn from both musical recordings (e.g., Trainor et al., 2004) and socially contingent nonmusical video interactions (e.g., Roseberry, Hirsh-Pasek, & Golinkoff, 2014). Many infants in modern societies encounter music produced by sound recordings as frequently as music produced by a parent (e.g., Mehr, 2014). Indeed, parents may believe that recordings provide higher-quality musical experiences than their own singing, given widespread beliefs that music listening and music education can have a positive influence on cognitive skills (see Mehr, 2015). Thus, we tested for social effects of music under
all three learning conditions and investigated whether the magnitude of these effects varied by condition.

The infants in our experiments learned one of two lullabies that were adapted to equate their rhythms and lyrics; they differed only in melody and sounded highly similar (see Fig. 1a). The infants were randomly assigned to a 1- to 2-week familiarization with one song or the other, presented by one of the methods just described. At test, we measured the duration of the infants' looking at two novel people before and after they each sang one of the songs. If melodies convey information about social affiliation, then the infants would be expected to attend selectively to the new singer of the familiar song.

Experiment 1

Method. In Experiment 1, parents with little or no musical expertise learned to sing one of the two songs. They sang that song to their infants on a regular basis and visited a Web site where they could listen to the song, to help them faithfully reproduce it at home. To maximize ecological validity, we instructed the parents to "sing the song to your baby as much as you like." After 1 to 2 weeks of song exposure, the infants returned to the lab for a selective-attention test (described later in this section). Finally, the parents were recorded as they sang to their infants, so that expert raters could judge their pitch accuracy.
Participants. We recruited 38 full-term infants and their parents from the greater Boston area. The parents were given a $5 travel reimbursement at each lab visit, and the infants were given a toy or other small reward after completing the study. All testing took place at the Laboratory for Developmental Studies at Harvard University. Data from 6 infants were excluded because they were fussy \((n=1)\) or inattentive \((n=1)\), or because their parents failed to produce a recognizable song \((n=4)\). Thus, analyses included 32 infants \((17 \text{ females}; \text{ mean age} = 5.61 \text{ months}, SD = 0.31, \text{ range: 5.06–6.11})\). Typically, the infants were brought to the lab by one parent, who participated in the experiment; when both parents were present, one elected to serve as the participating parent, who

Figure 1. The lullabies (a) and testing procedure (b) used in Experiments 1 through 3. During the baseline trial of the selective-attention test, two novel individuals stood silently smiling, directly gazing at the infant, for 16 s. Next, during the two familiarization trials, each individual sang one of the two lullabies in turn while gazing at the infant \((22 \text{ s per trial}; \text{ i.e., one person sang the lullaby that was familiar to the infant, and the other sang the unfamiliar lullaby})\). Finally, during the test trial, the two individuals again stood silently smiling, directly gazing at the infant.
completed all surveys, was given the music lesson, and sang to the infant at home and in the lab. Participating parents were predominantly female (26 female, 6 male). Some parents reported that other adults learned the song and sang it to their infant as well.

**Statistical power.** The target sample size of 32 was determined before the experiment began, to ensure adequate power to detect a positive selective-attention effect. A similar experiment testing effects of language rather than music (Kinzler et al., 2007) obtained an effect size ($d$) of 0.54, and a sample of 32 had .84 power to detect an effect of this magnitude.

**Musical content.** We adapted two obscure lullabies from folk repertoires ("Babushka Baio" and "Shady Grove, My Little Love"; Feierabend, 2000) to create two songs with identical rhythms and lyrics but different melodies (see Fig. 1a). In pilot testing, the ratings of approximately 30 adults demonstrated that the songs were equally pleasant; half preferred one song, and half preferred the other. Additionally, twelve 5-month-old infants in a pilot study attended comparably to videos of novel individuals singing the two songs.

**Music instruction and assessment of music aptitude.** At the first lab visit, each participating parent was randomly assigned to learn one or the other song (using the True Random Number Generator at www.random.org) and was given a 10- to 30-min music lesson, conducted without music notation and with the aid
of a keyboard. The lesson concluded when the parent indicated that he or she felt confident in being able to reproduce the song without assistance. Parents were given access to a Web site that provided two recordings of their song (with and without lyrics), as well as the printed text of the lyrics; they were encouraged to practice by singing along with these recordings, but were also told that their infant should hear the song only from live individuals (i.e., that they should never play the recorded music on the Web site to their infants or record their own voice and play the recording for their infants). Compliance with these instructions was high: Thirty-one of the 32 parents visited the Web site at least once and spent enough time on the site listening to the song, on average, to listen to it a total of approximately 15 times over the course of the study. Total time spent on the site listening to the song was comparable across the two song conditions (Song 1: $M = 5.73$ min, $SD = 5.62$; Song 2: $M = 5.23$ min, $SD = 6.02$), $t(29.9) = 0.24, p = .81$ (Satterthwaite's $t$ test). Thus, the two songs appear to have been comparably easy for untrained musicians to learn.

Parents also completed a standardized assessment of auditory perception skill, the Advanced Measures of Music Audiation (AMMA; Gordon, 1989), so that we could test for effects of parents' music aptitude on their infants' behavior.

**Assessment of singing accuracy.** At the second lab visit, we recorded the parents singing the song from the experiment to their infants during a free-play
session. The audio recordings were presented to three hypothesis-blind professional musicians, who independently judged the number of accurate pitches (of 25 total pitches) in each performance. Interrater reliability (computed from the raw number of accurate pitches) was high (Cronbach’s α = .90). When any one rater was unable to identify the song from a parent’s performance, the data from the parent and infant were excluded from analyses. This occurred in four cases. With the exception of these participants, parents' pitch accuracy (proportion of correct pitches) was comparable across the two song conditions (Song 1: M = .87, SD = .13, range: .56–1; Song 2: M = .83, SD = .15, range: .52–1), t(29.7) = 0.76, p = .45 (Satterthwaite’s t test).

Survey. After the first lab visit, we e-mailed parents daily with a brief survey, in order to determine the approximate number of times each infant heard the song each day. The rate of survey completion was high (92.3%). To estimate the amount of the infants' song exposure, we took the mean of each parent's daily responses to the question "About how many times did you sing your new lullaby to your baby today?" and multiplied it by the number of days of that family's participation in the study. These estimates thus accounted for both incomplete survey responses and variability in study length across participants.

Selective-attention test. At the second lab visit, we tested each infant’s attention to two novel individuals, one who sang the song that was familiar to
the infant and another who sang the unfamiliar song. The infant sat on his or her
parent’s lap, approximately 5 ft away from a 55- × 40-in. projection screen. The
parent closed his or her eyes and wore noise-canceling headphones that played
masking music throughout the experiment. The selective-attention test had four
trials (see Fig. 1b). First, the infant viewed side-by-side high-definition video
recordings of the two unfamiliar individuals, smiling with direct gaze at the
infant, for 16 s (baseline trial). Then, the infant viewed, in turn, one 22-s video of
each of the two individuals singing one of the two songs while continuing to
look and smile at the infant (familiarization trials). Finally, the infant viewed a
silent 16-s test trial that was identical to the baseline trial. A looming object with
an attractive sound effect brought the infant’s eyes to the center of the screen
before the baseline and test trials.

The videos were dubbed so that the two unfamiliar individuals sang in
the same voice, to control for potential differences in singing quality across the
two song conditions. The song-to-singer pairing, order of the familiarization
trials, and presentation location (left or right side of the screen) were fully
counterbalanced. Each infant’s gaze in each of the four trials was recorded with a
hidden high-definition camera and was coded frame by frame, at 30 frames per
second, by a coder blind to which song the infant was familiar with and to the
presentation location of each song. A second person recoded gaze for all the
infants, and interrater reliability (computed by correlating the raw proportion of looking toward the singer of the familiar song) was high ($r = .91$). The coders viewed the baseline and test trials before the familiarization trials, so that no differences in the infants’ behavior during the singing could influence the coding of attention to the individuals during the test trial.

**Results.** No variables differed between infants exposed to Song 1 and those exposed to Song 2 ($p > .1$); thus, we present the data in aggregate. In the selective-attention test, the infants showed no preference for either individual at baseline. The proportion of time that they looked toward the person who sang the familiar song did not differ from chance (i.e., $.5; M = .521, SD = .177, 95% confidence interval, or CI = [.457, .585]), $t(31) = 0.67, p = .51$ (one-sample t test; see Fig. 2a). Moreover, the infants attended highly and equally to the two singers during the familiarization trials, as each singer appeared by herself and sang a song (duration of looking toward the singer of the familiar song: $M = 15.6$ s, $SD = 5.07$; duration of looking toward the singer of the unfamiliar song: $M = 15.3$ s, $SD = 5.10$), $t(31) = 0.28, p = .78$ (paired t test). At test, however, the infants selectively attended to the now-silent singer of the song with the familiar melody; the proportion of time during which they looked toward her was greater than chance ($5; proportion of looking: M = .593, SD = .179, 95% CI = [.529, .658]), $t(31) = 2.96, p = .0059, d = 0.52$ (one-sample t test), and greater than the proportion at
baseline (difference in proportion of looking: $M = .072$, $SD = .169$, 95% CI = [.011, .133]), $t(31) = 2.42$, $p = .022$, $d = 0.43$ (paired t test; see Fig. 2a).

We used simple linear regression to investigate whether the degree of infants’ increase in attention to the singer of the familiar song from baseline to test depended on their level of in-home exposure to that song. Parents reported singing regularly to their infants (median of 9 performances per day,
interquartile range = [4, 11]; estimated total number of song performances: \( M = 76, SD = 56 \). After a log\(^2\) transformation (because of strong curvature), song exposure was a significant predictor of the within-subjects main effect (Fig. 2b), \( \chi^2(1) = 7.53, p = .006, R^2 = .14 \) (Wald test). A doubling of the approximate number of parental performances corresponded to an estimated 0.37-SD increase in attention to the novel person who sang that song.

We also tested the predictive power of two characteristics of parents' musical abilities: the objective accuracy of their song performances (proportion of correct pitches, as judged by the expert raters) and their music perception skills (as measured by the AMMA). Both measures showed considerable variation across the parents in the sample, but neither predicted infants' attentional preferences—singing accuracy: \( \chi^2(1) = 0.25, p = .61 \); music aptitude: \( \chi^2(1) = 0.85, p = .36 \) (Wald tests)\(^5\).

**Discussion.** The infants selectively attended to the novel person who previously sang the song they had learned from their parents, and the amount of in-home exposure to the song predicted the size of this effect. In Experiment 2, we asked whether infants would show a similar effect if their parents presented

---

\(^5\) All regressions reported in this article were bootstrapped with 40,000 replications and subsequently validated with sensitivity analyses, to ensure that their results were not attributable to the presence of influential observations.
them with a novel melody via a sound recording, emanating from a toy—that is, when the source of the melody was not a known social partner.

**Experiment 2**

**Method.** We repeated Experiment 1 with a second cohort of infants, altering the means of song exposure. In Experiment 2, parents were instructed never to sing the song to their child; instead, they were given a stuffed animal that produced music. When a parent squeezed this toy, it played a recording of an expert vocalist singing one of the two songs in an infant-directed fashion. Thus, song exposure in Experiment 2 was highly compelling and occurred in a social context, but the source of the song was an inanimate object. The procedure for recruiting participants, target sample size and statistical power, musical content, daily surveys (with minor changes in wording to account for the difference in exposure type, and a comparably high completion rate of 90.6%), selective-attention test, and analyses were the same as in Experiment 1. Thus, only the type of song exposure differed: In lieu of taking a music lesson in the lab and then singing the song in the home, parents in Experiment 2 were given a musical toy in the lab and used it with their infants at home. (Parent's pitch accuracy and music aptitude were not measured, because the parents in this experiment never sang the song to their infants.)
Participants. We recruited 35 full-term infants. Data from 3 infants were excluded because they were fussy \( n = 1 \), failed to attend to the test stimuli \( n = 1 \), or had prior familiarity with an actor in the test stimuli \( n = 1 \). Analyses included 32 infants (18 females; mean age = 5.49 months, \( SD = 0.31 \), range: 4.99–6.14). Participating parents were predominantly female (27 females).

Song exposure: musical toy. We adapted an age-appropriate stuffed animal (Jellycat Inc., London, United Kingdom) to play a recording of the song. The toy was no longer publicly available for purchase, and no parent indicated that his or her infant was familiar with the toy at the outset of the experiment. We removed stuffing from each toy and inserted a sound module (Invite By Voice LLC, Eden Prairie, MN) that played a recording of the song when the toy was squeezed. The recorded singers were two professional vocalists (gender matched to the participating parent), instructed to sing in an infant-directed fashion; neither vocalist was the singer in the selective-attention test. The toy played the song at approximately 55 dB at a distance of 5 in., a comfortable volume comparable to that of other commercially available musical toys intended for young infants. The infants lacked the dexterity to activate the toy themselves, and thus required a parent or another individual to do so. As in Experiment 1, the infants were randomly assigned to learn one song or the other.
At the first visit, we introduced the infants and parents to the musical toy and demonstrated its use. We instructed parents to "treat the toy as if you purchased it and use it as much as you like," indicating that the amount of song exposure and the nature of the infants' toy-directed actions were at the parents' discretion. Parents were instructed never to sing the song to their infants, and to relay this instruction to any other individuals who would come into contact with the toy. Compliance with this instruction was high: We asked parents at the second lab visit if they had sung the song during the exposure period. At most, parents reported a total of three or four instances of singing or humming the song, usually "by accident." Parents completed online surveys to report the amount of the infants' song exposure, as in Experiment 1.

**Selective-attention test.** Parents returned the toys to the lab at the second visit, when the infants participated in a testing session identical to that of Experiment 1. Interrater reliability, computed in the same fashion as in Experiment 1, was high ($r = .96$).

**Results.** The infants' degree of song exposure was comparable across the two experiments: The estimated total number of song performances was similar in Experiment 1 ($M = 76.3$, $SD = 56.2$) and Experiment 2 ($M = 81.8$, $SD = 50.5$), $t(61.3) = 0.41$, $p = .68$ (Satterthwaite's $t$ test), and the exposure periods were of comparable duration (in both experiments, $Mdn = 7$ days, range: 7–14). Given
that a song performance lasted 22 s, we estimated that the infants in Experiment 2 received about 30 min of song exposure, on average; this was more than the exposure in a previous study showing that 6-month-olds remembered songs presented to them via audio recordings (21 min in Trainor et al., 2004).

Figure 3. Main effects in Experiment 2. The box plots (a) show the proportion of time in the baseline and test trials during which the infants looked toward the singer of the familiar song. The dotted line indicates chance level (.5), the Xs indicate the means, the horizontal lines indicate the medians, the boxes indicate the interquartile ranges, and the vertical lines indicate the full ranges. The scatterplot in (b) shows each infant’s increase in proportion of looking toward the singer of the familiar song from baseline to test, along with the predicted effect of the amount of song exposure on the within-subjects main effect (i.e., the increase in the proportion of looking toward the singer of the familiar song from baseline to test), from a bootstrapped model with 40,000 replications. The dashed lines represent ±2 bootstrap standard errors, and chance (0) is indicated by the dotted line. Note that the x-axis is displayed on a log scale.
In the selective-attention test, the infants looked equally toward the two unfamiliar adults at baseline. The proportion of time that they looked toward the person who sang the familiar song did not differ from chance (i.e., .5; \( M = .512, SD = .175, 95\% \text{ CI} = [.449, .575] \)), \( t(31) = 0.39, p = .70 \) (one-sample \( t \) test; see Fig. 3a).

In addition, the infants attended equally to the two singers during familiarization (duration of looking toward the singer of the familiar song: \( M = 16.2 \text{ s}, SD = 5.18 \); duration of looking toward the singer of the unfamiliar song: \( M = 15.2 \text{ s}, SD = 6.69 \)), \( t(31) = 0.95, p = .35 \) (paired \( t \) test). The levels of attention to the two novel adults in Experiment 2 were comparable to the levels in Experiment 1 both during the baseline period, \( t(62.0) = 0.20, p = .84 \) (Satterthwaite's \( t \) test), and during the familiarization trials—familiar song: \( t(62.0) = 0.49, p = .63 \); unfamiliar song: \( t(57.9) = 0.06, p = .95 \) (Satterthwaite's \( t \) tests).

Despite the high appeal of the toy and professional quality of the singing, the selective-attention effects found in Experiment 1 were not present in Experiment 2. At test, the proportion of time during which the infants looked toward the singer of the familiar song (\( M = .537, SD = .181, 95\% \text{ CI} = [.472, .602] \)) did not differ from chance (.5), \( t(31) = 1.17, p = .25 \) (one-sample \( t \) test), or from the proportion at baseline (difference in proportion of looking: \( M = .025, SD = .165, 95\% \text{ CI} = [−.034, .085] \)), \( t(31) = 0.86, p = .40 \) (paired \( t \) test). Further, the amount of song exposure did not predict the infants' change in attentional preference.
toward the new individual who sang the familiar song (Fig. 3b), $\chi^2(1) = 0.01, p = .92$ (Wald test).

**Comparison with Experiment 1.** We used multiple regression to test the effects of exposure type (via a parent, in Experiment 1, vs. via a toy, in Experiment 2) and amount of song exposure on the within-subjects increase, from baseline to test, in looking toward the singer of the familiar song. As in Experiment 1, we performed a log$_2$ transformation on the parents’ self-reported number of song performances. We used no-constant models because the population intercept is known to be 0: If infants have never heard either song (i.e., 0 song performances), on average, they should not show any increase in looking toward either singer from baseline to test. Model building began with a simple no-constant model predicting the overall increase in looking from exposure quantity only; we then added the predictor of exposure type, and finally an Exposure Type $\times$ Exposure Quantity interaction.

The final model was statistically significant, $\chi^2(3) = 12.0, p = .0076, R^2 = .16$ (Wald test), as was the Exposure Type $\times$ Exposure Quantity interaction, $b = 0.059$, 95% CI = [0.014, 0.105], $z = 2.53, p = .011$ ($z$ test of the bootstrapped regression coefficient). The inclusion of exposure type and the interaction term significantly increased model fit, $\chi^2(2) = 6.63, p = .036$ (nested test). At the grand mean of exposure quantity (79 song performances), infants in Experiment 1 showed a
larger increase in attention toward the singer of the familiar song from baseline to test than did infants in Experiment 2, $b = 0.072$, 95% CI = $[-0.013, 0.156]$, $z = 1.67$, $p = .047$, $\beta = 0.43$ (general linear hypothesis test, one-tailed). At the 75th percentile of exposure quantity (101 song performances), the difference between the two experiments was larger, $b = 0.093$, 95% CI = $[-0.001, 0.187]$, $z = 1.95$, $p = .025$, $\beta = 0.55$ (general linear hypothesis test, one-tailed).

**Discussion.** Despite the use of recorded music, a common form of musical exposure for infants, the results of Experiment 2 differed markedly from those of Experiment 1. Infants selectively attended to a novel singer of a song originally sung by a parent, but not to a novel singer of a song originally played via a musical toy. This contrast suggests that infants do not prefer novel individuals who are associated with *any* familiar melody presented in a positive social context, despite long-standing findings that people prefer familiar sounds, objects, and patterns over unfamiliar ones, in both adulthood (Zajonc, 2001) and infancy (Bornstein, 1989).

Live and recorded singing differ in many respects, however. Although parents in Experiment 2 activated the toys, the toys themselves were inert: Unlike live singers, they did not interact contingently with the infants, vary the style and content of their song production, fine-tune their singing to the infants' affective state, or move in synchrony with their singing. In Experiment 3, we
presented infants with singing that incorporated these characteristics, but the source of the song was an unfamiliar adult who sang to the infants primarily over live, interactive video.

**Experiment 3**

**Method.** We repeated Experiment 1 with a third cohort of infants, again altering the means of song exposure. As in Experiment 2, parents were asked never to sing the song to their infants, but the parents and infants were introduced to a friendly adult (a university student with musical training) who sang the song to the infants in the lab. Parents then were given an iPad, by means of which this adult interacted daily with the infants, using interactive video chat (Skype). The procedure for recruiting participants, sample size and statistical power, musical content, daily surveys (with minor changes in wording to account for the difference in exposure type, and a comparably high completion rate of 90.0%), and selective-attention test were the same as in Experiment 1.

**Participants.** We recruited 39 infants. Data from 7 infants were excluded because they were fussy ($n = 1$), failed to attend to test stimuli ($n = 1$), or did not complete all the interactive video sessions ($n = 2$), or because of experimenter error ($n = 3$). Analyses included 32 infants (12 females; mean age = 5.82 months, $SD = 0.49$, range: 5.03–6.47). Participating parents were predominantly female (24 females).
Song exposure: interactive video sessions. Parents were provided with an iPad equipped with a flexible stand, the Skype application, a carrying bag, and a charger, all of which they kept for the duration of the study. At the first lab visit, the infants met the new adult, who played with them and sang the song directly to them six to eight times. The adult and parents then scheduled daily appointments for singing over interactive video. The parents were asked to choose times during which they expected the infants to be comfortable and attentive; across the cohort, appointment times were spread throughout the day, but within subjects they tended to occur consistently at the same time (e.g., every day after breakfast). The adult kept in regular contact with each parent via e-mail so as to ensure that the infant and parent were both present for each video appointment.

The infants participated in 6 to 11 interactive video sessions at home (Mdn = 7), each lasting roughly 10 min, during which the singer sang the song 4 to 11 times (Mdn = 7). Because the singing was live, there was natural variability in performance both within and between sessions. When the adult was not singing, she talked to the infants and parents. At the beginning of each session, she confirmed that the infant could see and hear her, and during the session, she asked the parent to reposition the iPad to maintain the infant's line of sight, as needed. The parents were invited to use their discretion in positioning the
infants; we chose to encourage this because, during pilot testing, we determined that most families were already accustomed to using interactive video with their infants; indeed, 86% of the infants in Experiment 3 had previously used Skype, FaceTime, Hangouts, or some other form of interactive video chat before participating in the study. Thus, there was variability in the infants' positioning during the sessions: Some infants sat in a parent's lap, others sat in a high chair with the iPad on the stand, and so on.

The infants were randomly assigned to learn one song or the other, and the parents were instructed that they should never sing the song to their infants and should relay this instruction to other individuals who were present during the video sessions (who might learn the song incidentally). Compliance with this instruction was high, with only two parents in the study reporting any live singing of the song.

Selective-attention test. Parents returned the iPad to the lab at the second visit, when the infants participated in a selective-attention test identical to that of Experiment 1. Interrater reliability, computed in the same fashion as in Experiments 1 and 2, was high ($r = .95$).

Assessment of the infants' responses to song exposure. The presentation of songs via interactive video allowed us to analyze the infants' responses to song exposure in a fashion that was not possible in Experiments 1 and 2: In
Experiment 3, all video sessions were recorded. These videos enabled us to compare the infants’ responses to the song and speech in a natural setting and to test for the relation between the infants’ engagement with the song during learning and their selective attention to a novel singer at test.

We analyzed three video sessions per infant, from the beginning, middle, and end of the exposure period. Videos in which the infants were fussy were not used (exclusion decisions were made before any coding was performed, to avoid confirmation bias). Two coders viewed the videos at 10 frames per second and recorded the durations of four categories of events: infant smiling, infant gazing toward the screen (where the adult was always visible), adult singing, and adult speaking. One of the three videos from each infant was randomly selected for double coding, and interrater reliability (computed as the frame-by-frame percentage of agreement across the two coders) was high (82%–99%).

Although infants’ reactions to songs performed via video chat have never been systematically assessed, infants find singing highly enjoyable (for a review, see Patel, 2008). We expected that if the infants were sufficiently engaged with the adult during the interactive video sessions, they would smile more and attend longer to her when she was singing than when she was speaking.

**Results.** Parents’ estimates of the total number of song performances ($M = 47.7, SD = 17.0$) were slightly lower than the actual number, as determined from
the recorded video sessions ($M = 48.7$, $SD = 16.9$); this difference was not significant, $t(31) = 0.59$, $p = .56$ (paired $t$ test). The total duration of song exposure ($M = 17.5$ min, $SD = 6.23$; extrapolated from the three videos that were coded for singing duration) was slightly lower than in a previous study in which 6-month-old infants were taught songs and subsequently remembered them (21 min; Trainor et al., 2004). Parents reported significantly less song exposure in

Figure 4. Main effects in Experiment 3. The box plots (a) show the proportion of time in the baseline and test trials during which the infants looked toward the singer of the familiar song. The dotted line indicates chance level (.5), the Xs indicate the means, the horizontal lines indicate the medians, the boxes indicate the interquartile ranges, and the vertical lines indicate the full ranges. The scatterplot in (b) shows each infant’s increase in proportion of looking toward the singer of the familiar song from baseline to test, along with the predicted effect of the amount of song exposure on the within-subjects main effect (i.e., the increase in the proportion of looking toward the singer of the familiar song from baseline to test), from a bootstrapped model with 40,000 replications. The dashed lines represent ±2 bootstrap standard errors, and chance (0) is indicated by the dotted line.
Experiment 3 than in the other two experiments—comparison with Experiment 1: $t(36.6) = 2.75, p = .009$; comparison with Experiment 2: $t(38.0) = 3.6, p = .001$ (Satterthwaite’s $t$ tests)—although the study duration was comparable across all three experiments (Experiment 1: $Mdn = 7$ days, range: 7–14; Experiment 2: $Mdn = 7$ days, range: 7–14; Experiment 3: $Mdn = 7$ days, range: 7–12). We address possible bias in these reports later in this section.

In the selective-attention test, the infants showed no preference for either of the two unfamiliar adults at baseline. The proportion of time that the infants looked toward the person who sang the familiar song did not differ from chance (i.e., .5; $M = .479, SD = .183, 95\% CI = [.413, .545]), t(31) = 0.64, p = .53$ (one-sample $t$ test; see Fig. 4a). Similarly, the infants attended equally to the two singers during the familiarization trials (duration of looking to the singer of the familiar song: $M = 18.4 s, SD = 4.75$; duration of looking to the singer of the unfamiliar song: $M = 17.2 s, SD = 4.93$), $t(31) = 1.28, p = .21$ (paired $t$ test). The infants in Experiments 1 and 3 exhibited comparable levels of attention during the baseline period, $t(61.9) = 0.93, p = .36$, and during the familiarization trial for the unfamiliar song, $t(61.9) = 1.47, p = .15$. However, the infants in Experiment 3 attended longer to the familiarization trial for the familiar song than did the infants in Experiment 1, $t(61.7) = 2.26, p = .027$ (all Satterthwaite’s $t$ tests).
The selective-attention effects found in Experiment 1 were not present in Experiment 3. The proportion of time that the infants looked toward the novel person who sang the familiar song did not differ from chance (i.e., .5; \( M = .488, SD = .232, 95\% CI = [.404, .571] \)), \( t(31) = 0.30, p = .77 \) (one-sample \( t \) test), or from the proportion at baseline (difference in proportion of looking: \( M = .009, SD = .198, 95\% CI = [-.063, .080] \)), \( t(31) = 0.24, p = .81 \) (paired \( t \) test; see Fig. 4a). Further, the amount of song exposure did not predict the infants' change in attentional preference toward the new singer of the familiar song (Fig. 4b), \( \chi^2(1) = 0.46, p = .50 \) (Wald test).

**Comparison with Experiment 1.** Because parents may have paid greater attention to songs they themselves sang than to songs sung by a toy or another person, their reports of song exposure may have been depressed in Experiment 3, relative to Experiments 1 and 2. For this reason, we used somewhat different methods to compare the results of Experiments 1 and 3 than we used to compare the results of Experiments 1 and 2. We began by testing for differences between the experiments (exposure type) without adjusting for differences in the amount of song exposure. Then we proceeded with a regression model that differed in one respect from that used to compare Experiments 1 and 2: Instead of predicting selective attention from the estimated total number of song performances, after a log\(_2\) transformation, we standardized those values by dividing them by the
standard deviation in each experiment in an effort to reduce the differences in parents' self-reporting bias across the two experiments. The predictor can thus be interpreted as standard-deviation units of song exposure.

Both approaches revealed differences between the two experiments. The first showed that at test, the infants attended significantly longer to the singer of the familiar song in Experiment 1 than in Experiment 3, $t(58.2) = 2.04, p = .046$ (Satterthwaite’s $t$ test). The difference between the experiments in the increase in proportion of looking toward the singer of the familiar song from baseline to test was smaller, but in the same direction, $t(60.5) = 1.39, p = .086$ (Satterthwaite’s $t$ test, one-tailed). We continued with the second approach, which included the standardized measure of song exposure as a predictor. The model's omnibus test yielded a significant result, $\chi^2(3) = 11.0, p = .01, R^2 = .14$ (Wald test), as did a test of the Exposure Type × Exposure Quantity interaction, $b = 0.061, 95\% \text{ CI} = [0.006, 0.115], z = 2.18, p = .029$ (z test of the bootstrapped regression coefficient). The inclusion of exposure type and the interaction term significantly increased model fit, $\chi^2(2) = 6.87, p = .032$ (nested test). Thus, although the comparison of the experiments was complicated by the introduction of bias in the measure of song exposure, the infants’ behaviors at test clearly differed between Experiments 1 and 3. As in Experiments 1 and 2, all regressions were bootstrapped with 40,000
replications and validated with sensitivity analyses to ensure that findings were not attributable to the presence of influential observations.

**Infants' responses to song and speech presented via interactive video.**

Across the three sessions that were coded for each infant, we computed four variables: duration of smiling while the adult sang, duration of looking at the video screen while the adult sang, duration of smiling while the adult spoke, and duration of looking at the video screen while the adult spoke. Each variable was computed as a proportion of singing time or speaking time. We then used planned comparisons to test whether the infants smiled and gazed at the video differentially, within subjects, depending on whether the adult was singing or speaking.

When the adult sang, the infants smiled at nearly twice the rate ($M = .098, SD = .106, 95\% CI = [.059, .136]$) that they did when she spoke ($M = .054, SD = .062, 95\% CI = [.032, .077]$), $t(31) = 2.85, p = .008$ (paired $t$ test). Likewise, they attended to the screen at a significantly higher rate during her singing ($M = .623, SD = .184, 95\% CI = [.557, .689]$) than during her speaking ($M = .547, SD = .130, 95\% CI = [.500, .594]$), $t(31) = 4.02, p = .0003$ (paired $t$ test).

Given that these positive responses might reflect differences in infants' experiences during song exposure between Experiments 1 and 3, we conducted two exploratory analyses to test whether these measures predicted the infants'
preferences regarding the novel individuals at test. The results were negative: The rate of smiling during singing episodes were not related to the within-subjects main effect, $\chi^2(1) = 0.23, p = .63$ (Wald test), and neither was the rate of looking during singing episodes, $\chi^2(1) = 0.50, p = .48$ (Wald test). Thus, although the infants responded more positively to songs than to speech, as evidenced by two measures, the magnitude of these positive responses to the singing did not predict the infants' attentional preferences between the novel singers at test.

**Discussion.** The contrasting findings of Experiments 1 through 3 suggest that parents' singing has different effects on infants than do either musical toys or singing by a minimally familiar adult over interactive video. Did these differing effects stem from infants' differential attention to and learning of the songs in the three conditions? Although we know that the infants in Experiment 3 were highly attentive to the singer, the infants in Experiments 2 and 3 might have failed to learn the song well enough to recognize it at test.

The best test of this possibility would be to compare the infants' later memory for the melody across the three experiments, but this comparison was not possible. Many of the parents in Experiment 1 continued to sing the song after the experiment ended, whereas the return of the toy at the end of

*Parents volunteered this information when they returned to our lab to participate in unrelated studies.*
Experiment 2 and the return of the iPad at the end of Experiment 3 made it unlikely that the infants heard the song again. Thus, in Experiments 4 and 5, we investigated long-term retention of the melody among the participants from Experiments 2 and 3, respectively.

Three to 12 months after the conclusion of those experiments (more than 8 months later, on average), we tested whether the infants could discriminate the familiar melody from the other melody while listening in the same context in which they had originally learned the song. In Experiment 4, we tested infants' attention to two visually identical toys from Experiment 2; one toy played the familiar song, and the other played the unfamiliar song. In Experiment 5 we tested infants' attention to two videos of the singer from Experiment 3; she sang the familiar song in one video and the unfamiliar song in the other.

Experiment 4

Method.

Participants. We attempted to test all 32 infants from Experiment 2, but we failed to reach three families, two families declined to participate, and one family had moved away; in addition, 4 infants participated but were excluded because they were fussy during the lab session. Thus, our analyses included 22 full-term infants (11 females; mean age = 13.9 months, $SD = 1.50$, range: 10.8–17.1), who were tested an average of 8.63 months after they had last heard the
recorded song in Experiment 2 ($SD = 1.51$, range: 5.82–11.6 months). During this interim, some of these infants had returned to the lab for unrelated experiments, but none had received any additional exposure to the toy or the recording of the familiar song. We asked the parents if they had sung the toy’s song after the original study ended. Parents of 5 of the 22 infants reported singing the song “once or twice” during the weeks following Experiment 2, but not afterward.

**Procedure.** The infants sat in a high chair while we presented them with toys from Experiment 2, at a distance of approximately 66 cm. Parents sat next to their infants, facing away from the toys, and wore noise-canceling headphones that played masking music for the duration of the experiment. The experiment comprised four trials. On each trial, the experimenter, who was unaware of which song had been presented to the infant during the original experiment, placed a toy to one side in front of the infant. He said, "Look at this toy, [baby’s name]!” and activated the song (see Fig. 5a). The trials alternated between two visually identical toys, one that played the song the infant had heard in Experiment 2 and one that played the other song used in Experiment 2, which the infant had never heard except during the selective-attention test in that experiment. Thus, each infant heard the familiar song twice and the unfamiliar song twice. A given song was presented on the same side both times.
On each trial, we measured the infant’s gaze toward the toy during the singing (22 s) and afterward in silence (16 s). The recordings were identical to those used in Experiment 2, and did not differ in lyrics, rhythm, or the identity of the singer. Because the two toys were visually identical, the only basis for distinguishing between them was the melodies they played. The order of

Figure 5. Procedure and results of Experiment 4. On each of four trials (a), one of two visually identical toys was presented to one side in front of the infant for the duration of one performance of either the familiar or the unfamiliar song and during a 16-s delay afterward. The two songs were presented in alternation. The box plots (b) show the duration of looking toward each of the toys during the first phrase of the song. The Xs indicate the means, the horizontal lines indicate the medians, the boxes indicate the interquartile ranges, and the vertical lines indicate the full ranges. The asterisk indicates a significant difference (*p < .05).
presentation (familiar vs. unfamiliar song first) and the location of each toy (left side vs. right side) were counterbalanced, though lower-than-expected recruitment led to a slight imbalance in presentation order, such that 13 infants heard the familiar song on Trials 1 and 3, and 9 heard it on Trials 2 and 4 (we report a test for order effects in the next section). Each infant’s gaze was recorded with a hidden high-definition camera and coded at 30 frames per second by an experimenter who was blind to the infant’s song familiarity. A second experimenter recoded the videos of all participants, and interrater reliability (computed by correlating raw looking times, four trials per infant) was high ($r = .97$).

**Results.** Despite the lengthy delay between initial song exposure and test, the infants looked longer, on average, at the toy that played the song they had heard at home more than 8 months earlier than at the toy that played the unfamiliar song. This effect was strongest during the first phrase of the songs (Fig. 5b; difference in looking time: $M = 0.815$ s, $SD = 1.61$, 95% CI = [0.10, 1.53]), $t(21) = 2.38$, $p = .027$ (paired t test), which is consistent with findings in a previous study of musical memory (Saffran, Loman, & Robertson, 2000), but it remained marginally detectable over a longer interval (first half of the songs; difference in looking time: $M = 1.28$ s, $SD = 3.66$, 95% CI = [−0.35, 2.90]), $t(21) = 1.63$, $p = .059$ (one-tailed). The first-phrase effect was no larger for the 5 infants whose parents
reported singing the familiar song after Experiment 2 than for the 17 infants whose parents did not report such exposure (mean difference = 1.46 s, 95% CI = [−1.00, 3.91]), \( t(5.13) = 1.51, p = .19 \) (Satterthwaite’s \( t \) test). The size of the main effect did not vary with presentation order (familiar vs. unfamiliar song first; mean difference = 0.95 s, 95% CI = [−0.53, 2.43]), \( t(16.0) = 1.37, p = .19 \) (Satterthwaite’s \( t \) test). In an exploratory analysis, we tested whether the amount of song exposure in Experiment 2 predicted the strength of the memory effect in Experiment 4. The result was negative: Infants with more song exposure were no more likely, on average, to show a greater preference for the toy playing the familiar song over the toy playing the unfamiliar song, \( \chi^2(1) = 0.51, p = .47 \) (Wald test).

In sum, Experiment 4 provides evidence that the infants in Experiment 2 learned the song from the toy, discriminated that song from a song with identical lyrics and timing but a different melody, and retained this distinction between the two highly similar songs for 6 to 12 months (\( M > 8 \) months).

**Experiment 5**

**Method.**

**Participants.** We attempted to test all 32 infants from Experiment 3, but we failed to reach six families, and three families declined to participate. The sample comprised 23 full-term infants (8 females; mean age = 5.86 months, \( SD = \)
0.50, range: 5.03–6.47), who were tested an average of 8.48 months after they had last heard the familiar song from Experiment 3 ($SD = 2.91$, range: 2.99–12.1).

During the interim, some of the infants had returned to the lab for other experiments, but none had received any additional exposure to the familiar song from either the singer or interactive video. Parents of 6 of the 23 infants reported singing the song during the first weeks following Experiment 3, but not afterward. Thus, exposure conditions in Experiment 5 were comparable to those in Experiment 4.

**Procedure.** The infants sat on the lap of a parent, who had closed eyes and wore noise-canceling headphones that played masking music for the duration of the experiment. They viewed high-definition videos of the adult singer from Experiment 3. On four alternating trials, she sang the familiar and unfamiliar songs (Fig. 6a). Because the same person sang both songs, the familiar song was not identifiable from the identity of the singer's voice. Because infants are highly attentive to videos of singing, we presented each song video alongside a video of a looming object that moved out of synchrony with the song, so as to avoid ceiling effects on the infants' gaze toward the face of the singing adult. The order of the songs (familiar vs. unfamiliar song first) and location of the singer (left vs. right side) were fully counterbalanced. The infants' gaze was recorded with a hidden high-definition camera and coded at 30 frames per second by two
experimenters, both of whom were unaware of which song was familiar to each infant. Interrater reliability (computed by correlating raw looking times, four trials per infant) was high ($r = .93$).

**Results.** The infants attended longer to the video of the adult when she sang the same song as in Experiment 3 than when she sang the other song (Fig. 6b; difference in looking time: $M = 2.23$ s, $SD = 4.80$), $t(22) = 2.27$ $p = .033$ (paired $t$ test). Preferential looking on familiar-song trials was no higher for the 6 infants whose parents reported singing the familiar song than for the 17 infants whose parents did not report such exposure (mean difference = 2.67 s, 95% CI = [−2.78, 8.12]), $t(7.89) = 1.13$, $p = .29$ (Satterthwaite’s $t$ test). Thus, these findings provide evidence that the infants in Experiment 3 learned the song via interactive video, discriminated that song from a song with identical lyrics and timing but a different melody, and retained this distinction for more than 8 months, on average.

In contrast to Experiment 4, Experiment 5 revealed a memory effect that was not largest in the first phrase of the song. Instead, effects were detectable through the entirety of the song. In addition, the latter two of the four trials (Trials 3 and 4) drove the overall difference in attention to the singing of the familiar versus unfamiliar song (mean difference between Trials 3 and 4 = 2.18 s, 95% CI = [0.47, 3.89]), $t(22) = 2.61$, $p = .015$ (paired $t$ test). This difference between
the effects observed in the two experiments is likely due to differences in the format of the memory test: In Experiment 4, the toy never moved on its own, and its recorded song was always performed identically. Visual attention to the toy therefore dropped off quickly after the beginning of each test trial. In contrast, in Experiment 5, the infants viewed videos of the person who had previously sung

Figure 6. Procedure and results of Experiment 5. On each of four trials (a), the infants viewed a video of the person who had previously Skyped with them. She sang either the familiar or the unfamiliar song (on alternating trials) and appeared alongside a green looming object. The box plots (b) show the duration of looking toward the videos of the singer while she sang the familiar song and while she sang the unfamiliar song. The Xs indicate the means, the horizontal lines indicate the medians, the boxes indicate the interquartile ranges, and the vertical lines indicate the full ranges. The asterisk indicates a significant difference (*p < .05).
to them in a variable context, and whose behavior varied throughout the test; thus, their attention was better sustained throughout each trial.

In an exploratory analysis, we tested whether the amount of song exposure in Experiment 3 predicted the strength of the memory effect in Experiment 5. As we found in the corresponding analysis in Experiment 4, the result was negative: Infants with more song exposure were no more likely, on average, to show a greater preference for the video of the adult singing the familiar song over the video of her singing the unfamiliar song, $\chi^2(1) = 0.04, p = .84$ (Wald test). Finally, we tested whether a relation between song exposure and strength of the memory effect might be obtained by collapsing across the two experiments. In a simple logistic regression, the likelihood of a difference in looking time between the two songs was unrelated to the degree of prior song exposure, $\chi^2(1, N = 45) = 0.08, p = .77$ (Wald test). Thus, the degree of exposure did not contribute to the size of the memory effects in Experiments 4 and 5.

**Supplementary analysis.** Because song exposure was video-recorded in Experiment 3, we were able to test whether the infants’ levels of engagement while they learned the song in that experiment predicted their memory in Experiment 5. We computed the size of the memory effect in Experiment 5 as the raw difference in duration of looking toward the person singing the familiar song and duration of looking toward the person singing the unfamiliar song; we
tested for relations between this measure and the measures of smiling and looking during song exposure in Experiment 3. No model showed a clear predictive relation—smiling only: $\chi^2(1) = 1.32, p = .25$; looking only: $\chi^2(1) = 2.31, p = .13$; both smiling and looking: $\chi^2(2) = 2.28, p = .32$; smiling, looking, and their interaction: $\chi^2(3) = 1.62, p = .66$. Although the infants remembered the song from Experiment 3 after a long delay, their levels of engagement while they learned the song did not predict the degree of the memory effect.

**General Discussion**

Infants selectively attended more to a novel individual who sang a song learned from a parent’s singing than to a novel individual who sang a contrasting song with the same words and rhythms but a different melody. Infants displayed no such preference when they learned the song from a recording emanating from an inanimate toy or from live video interactions with a singing adult whom they had met only briefly.

Strikingly, the infants in the latter two conditions remembered the melody to which they had been exposed for an average of more than 8 months, in sufficient detail to discriminate it from a second, highly similar melody (see Fig. 1a). Moreover, analyses of video recordings of the infants as they learned the song in Experiment 3 revealed that they exhibited substantially more positive engagement with the adult while she sang than while she spoke. Thus, 5-month-
old infants enjoy melodies that are sung by a variety of people under different conditions, and they show long-term retention of melodies learned from a variety of sources. Nevertheless, only a melody produced live by a parent leads infants to display an attentional preference for a new person who sings that melody.

The effect of parents’ singing was robust to variation in their musical skills: Infants readily identified familiar melodies sung by novel individuals even when their parents' renditions of those melodies only roughly matched the new performances. This finding speaks to parents and early-childhood educators who favor high-quality, professionally recorded music as a source of infants' song exposure. Caregivers with low confidence in their musical abilities need not worry that the effects of their live singing are reduced by their lack of extensive musical training: In our experiments, we could not predict infants' behaviors at test from their parents' musical abilities.

Our findings suggest an early link between live song and social engagement that is independent of songs' semantic content: Social responses were driven by melody alone. This link may be attributable to the early experiences of infants within their families, evolved predispositions to view songs as signals of a social connection, or both; our experiments do not distinguish between these interpretations.
Moreover, the present experiments do not reveal whether the selective-attention effect found in Experiment 1 was driven by the song being sung by a parent or by its being sung in person. It is possible that infants show social preferences for new singers of familiar songs only when those songs have been learned from family members. Alternatively, infants may show social preferences for new singers of songs learned from any friendly adult, but only when the adult sings to them in person. Further experiments using the same general methods of the present experiments could distinguish between these possibilities. For example, might infants display social preferences for novel singers of songs learned from family members who sing only over interactive video? We predict that they will. If so, further analyses of singing over live video could then determine the particular performance features that elicit social interpretations of music. For example, songs may convey social information to infants more effectively the more the singers are attuned to the infants’ affective states, varying their singing style in a fashion coordinated with the infants’ responses.

Whatever the findings of such studies, the present research demonstrates that the social information conveyed by a melody depends on its original source. We found that a melody conveys social information about its singer when it is sung by a parent, during the course of parent-infant interactions. In contrast,
such social information is not conveyed when the melody is produced by a musical toy, even if the parent is highly engaged with the infant while playing with the toy, or when the melody is sung by an interactive but socially unrelated adult, even if the infant is highly engaged with that adult in interactive video sessions.

Why might melodies serve a social function? In small-scale human societies, child rearing is conducted by multiple individuals who communicate with one another, sharing language and music (e.g., Hrdy, 2009). Infants might do well to identify individuals who could care for them when a parent is not available, and attending to other people's speech and singing might facilitate such identification. Indeed, some researchers have proposed that the human music faculty evolved in the context of child rearing (e.g., Trehub, 2003). Our studies did not test any particular evolutionary theory, but this speculative interpretation of our findings is consistent with the possibility that in ancestral environments, infants' caregivers reliably produced melodies, and infants reliably listened to and remembered them.
Chapter 3

Shared musical knowledge in eleven-month-old infants

Mehr & Spelke (2017)\textsuperscript{7}

Five-month-old infants selectively attend to novel people who sing melodies originally learned from a parent, but not melodies learned from a musical toy or from an unfamiliar singing adult, suggesting that music conveys social information to infant listeners. Here, we test this interpretation further in older infants with a more direct measure of social preferences. We randomly assigned 64 eleven-month-old infants to 1–2 weeks' exposure to one of two novel play songs that a parent either sang or produced by activating a recording inside a toy. Infants then viewed videos of two new people, each singing one song. When the people, now silent, each presented the infant with an object, infants in both conditions preferentially chose the object endorsed by the singer of the familiar song. Nevertheless, infants' visual attention to that object was predicted by the degree of song exposure only for infants who learned from the singing of a parent. Eleven-month-olds thus garner social information from songs, whether learned from singing people or from social play with musical toys, but parental

\textsuperscript{7} This chapter appears in Developmental Science with the same title as above. As in Chapter 2, for consistency with the published version, I have retained the use of first-person pronouns throughout.
singing has distinctive effects on infants’ responses to new singers. Both findings support the hypothesis that infants endow songs with social meaning. These findings raise questions concerning the types of music and behavioral contexts that elicit infants’ social responses to those who share music with them, and they support suggestions concerning the psychological functions of music both in contemporary environments and in the environments in which humans evolved.

Introduction

Infants are avid music listeners. They discriminate consonant from dissonant intervals and detect musical beats at only a few days of age (Perani et al., 2009; Winkler, Háden, Ladinig, Sziller, & Honing, 2009); they tolerate repetitive, unfamiliar melodies longer than infant-directed speech (Corbeil, Trehub, & Peretz, 2015); they remember melodies heard in the womb after they are born (Granier-Deferre, Bassereau, Ribeiro, Jacquet, & DeCasper, 2011); and they discriminate highly similar songs on the basis of their melodies alone, long after last hearing them (Mehr, Song, & Spelke, 2016). These are a few highlights of a rich literature on infants’ music cognition (review: Patel, 2008), but one basic question has received relatively little attention: Why do infants care about the music they hear?

One possibility is that infants garner social information from music, as they do from language and accent (Kinzler, Dupoux, & Spelke, 2007, 2012), from
direct social overtures (Csibra & Gergely, 2009; Schachner & Hannon, 2011), and from food choices (Liberman, Woodward, Sullivan, & Kinzler, 2016). Recently, we reported that five-month-old infants selectively attend to the singers of familiar songs when they had originally learned those songs from a parent (Mehr et al., 2016); that attentional preference was found during a silent test, after singing was completed, and can thus be interpreted as a preference for the singer, rather than for the song itself. In contrast, infants displayed no such preference if they learned the same song from a recording embedded in a toy or from an initially unfamiliar adult, who sang to them both live and by interactive video. Infants attended to the song under these conditions and remembered it after long delays, but they did not prefer new singers who performed it.

These findings raise the possibility that a psychological function of music lies in the social domain. On this hypothesis, songs are more than pleasurable noise: Because they are produced by and learned from other people, primarily in social contexts, songs may convey social information about their singers, as two people who know the same song are more likely to be socially connected than those who do not. This social signal is potentially useful to infants, who could benefit from attending to and eliciting care from those people who are most likely to provide it.
Here, we explore this hypothesis through a new experiment that differs from our earlier studies in four ways. First, we use a more direct measure of infants’ social preferences, based on research investigating infants’ social preferences between native- and foreign-language speakers (Kinzler, Dupoux & Spelke, 2012). After hearing two unfamiliar people sing different songs, one of which was learned from a parent who sang or activated a toy for the infant, infants were presented with two new objects of different types and each person, now silent, endorsed a different object for the infant, whose looking at and reaching for the objects was measured. If infants prefer the person who sang the song they had learned, they should reach more for the object endorsed by that person. Second, we present infants with play songs rather than the lullabies used in our past research, to test for the generality of the preference effect. Third, we use songs that differed from one another in melody, lyrics, and rhythms, unlike the original study, which presented songs with different melodies but identical lyrics and rhythms. This change addresses an alternative explanation of our original result: infants may have perceived the singer of the novel song as a less competent singer of the original song, and then exhibited a preference for the more competent individual.

Finally, we test 11-month-old infants, in contrast to the 5-month-old infants tested previously, as older infants may attribute social meaning to songs
under a greater range of conditions. As infants approach their first birthday, they begin to incorporate objects into their social interactions (reviews: Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Tomasello, 2008). They share their social partners’ attention to objects (Brooks & Meltzoff, 2005; Liszkowski, Carpenter, Henning, Striano, & Tomasello, 2004) and they treat gift-giving, imitation of object-directed actions, and helping others to attain objects as socially meaningful acts (Agnetta & Rochat, 2004; Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013). Older infants therefore may show social preferences for new singers of songs even when the parent presented the song by activating a toy rather than by singing.

Thus, we randomly assigned infants to learn one of two obscure play songs, either from a parent or from a recording embedded in a stuffed animal (as in Experiments 1 and 2 of Mehr et al., 2016). After 1–2 weeks of exposure, infants returned to the lab, where they viewed videos of novel people, each of whom sang one of the two songs and then presented one of two objects to the infant (after Kinzler et al., 2012). We reasoned that infants would interpret the behavior of the two adults either as invitations to share attention to an object (see, e.g., Tomasello, 2008) or as attempts to convey information about an object (Csibra & Gergely, 2009). In either case, we predicted that the infants who learned the song from a singing parent would demonstrate a social preference for the singer of a
familiar song by attending to and reaching for the object endorsed by that person. For infants in the parent-activated toy condition, predictions were less clear. When a parent activates the musical toy, infants may interpret the music as being offered or endorsed by the parent. Because the parent is not the direct source of the music, however, the social significance of the music may be attenuated, relative to the parent singing condition. Thus, 11-month-old infants may or may not show a social preference for the novel singer of the familiar song in the toy condition.

**Method**

**Participants.** All testing took place at the Laboratory for Developmental Studies at Harvard University. We recruited 79 full-term infants and their parents from the greater Boston area. Incentives were $10 in travel reimbursements for parents and toys or other small rewards for infants. Data from 15 infants were excluded because they were fussy during testing ($n = 4$) or failed to reach for an object on any test trials ($n = 11$). These exclusion criteria were determined before the experiment began. Thus, analyses included 64 infants (29 females; mean age = 11.2 months, $SD = 0.25$, range: 10.8–12.0), a sample size chosen before the experiment began to match our previous work. Parents accompanying their infants to the lab were predominantly female (52
female); when a male parent was present, we asked that he participate in the study as the primary parent, as in our previous work.

**Musical exposure.** Families were randomly assigned to present a new song under one of two exposure conditions ($n = 32$ each) lasting 1-2 weeks ($Med. = 10$ days, IQR [7.5, 11]). In the parent singing condition, parents were taught one of the two songs during an initial visit to the laboratory, without music notation and with the aid of a keyboard (see Exp. 1 in Mehr et al., 2016), and they were given a recording of the song to take home for refresher training. In the musical toy condition, parents were given a stuffed animal (a green alligator), adapted to play a recording of one of the two songs when squeezed (see Exp. 2 in Mehr et al., 2016). The recorded singers were two research assistants (gender matched to the participating parent) who sang in an infant-directed manner; neither vocalist was heard in subsequent testing (described below). Parents were asked to present the toy to their infants but not to sing the song. Compliance with this instruction was high: at most, parents reported singing the song once or twice. In both conditions, they were told to present the song as much or as little as they liked. All families returned to the lab for a second visit.

---

*Because the songs in this study were considerably easier for parents to learn than those we used in previous work, the music lesson was brief, most parents reported not using the refresher recording, and most parents reproduced the song with accuracy at or near ceiling during a subsequent visit to the lab.*
Assessment of song exposure. Between the two lab visits, we contacted parents each evening with a survey (completion rate: 87%), asking "About how many times did your baby hear the song today?" We multiplied each parent’s mean responses by the duration of their participation in the study, yielding an estimate of the amount of song exposure that was not skewed by missing data or variation in study length.

Musical content. In both musical exposure conditions, infants were randomly assigned to learn one of two obscure children's songs (Feierabend, 1986). We wrote new lyrics for both songs, aimed to make them attractive to infants (Figure 1A).

Social preference test. During a 15-min test, infants viewed videos of two novel people who, over four familiarization trials, each sang the two verses of one of the two songs (Figure 1B). A silent test followed, in which both people appeared side by side (Figure 1C), lifted different objects in synchrony with one another, looked toward their object while smiling, looked at the infant, nodded, moved the object back and forth, and pointed downward. Immediately beneath the screen were physical replicas of the two objects, giving the illusion that the people in the videos were pointing at and endorsing them. The video froze during the pointing, and the infant’s high chair, which was fitted with wheels
and a track, was then pushed forward so that the infant could reach for the objects.

This process was repeated four times, yielding 16 familiarization and 4 test trials. The position and identity of the singer, along with the order of song presentation, was fully counterbalanced. The objects on trials 1 and 3 were a small stuffed lion and bear; on trials 2 and 4 they were realistic models of an apple and a pear. The location of these objects (and hence, their pairings with the two singers) was swapped across trials, such that on trials 3 and 4, both objects had been previously endorsed by both singers. Thus, primary analyses focus on

Figure 1. One of the two songs (A) was either sung by a parent or produced when a musical toy was activated during the exposure period. The testing procedure included four familiarization trials (B); each of two novel adults sang one of the two songs, one verse at a time, such that infants heard each adult sing both verses of her song. Then, each adult endorsed one of two objects (C) in the reaching test. In synchrony with one another, both adults smiled at the infant, presented and looked at a different object, showed it to the infant, and pointed down, toward a replica of that object that had been placed on a table before the familiarization trials. Infants then were moved forward so they could reach for either of the objects. The full sequence [i.e., (B) followed by (C)] was repeated two times, and then two further sequences were presented in which the pairings of people and objects were reversed.
the first two trials, in which each object was endorsed by only one singer (see Results).

Infant behavior was monitored by a hidden high-definition camera at 30 frames per second. Four coders viewed all footage independently of one another and blind to which song was familiar to the infant, to how the infant had learned that song, and to which singer sang which song (i.e., with footage muted). They coded gaze to each person and each object and touching of each object, in two passes per infant using Datavyu (2014), split across coders such that all data were independently coded twice. Inter-coder reliability, computed as the % agreement on a frame-by-frame basis and weighted by infant, was 97.6%.

Results

Because infants were randomly assigned to learn one of the two songs, and because there were no differences in reaching or gaze behavior across the two song conditions (ps > .4), we collapsed over the two songs and compared responses to the person who sang the familiar song.

Object choice. In the first two sets of familiarization trials, infants attended highly and comparably to both singers (across all eight trials, familiar song: $M = 46.2$ s, $SD = 30.8$ s; unfamiliar song: $M = 45.2$ s, $SD = 28.2$ s; $t(63) = 0.66$, $p = .51$). This pattern was comparable to five-month-old infants' responses to the same style of familiarization trials in Mehr et al. (2016), as well as 10-month-old
infants’ responses to unfamiliar people who previously spoke in a native vs. foreign language (Kinzler et al., 2012). In the main analyses, we computed for each infant a difference score between the number of target object reaches and the number of non-target object reaches on the first two trials. Given a priori, directional predictions, reported $p$-values are one-tailed except when analyses are specified as exploratory. In both music exposure conditions, infants reached
more for the target object than the non-target object (Figure 2; difference in number of reaches, parental song condition: Med. = 0.5, IQR = [-0.5, 2], z = 1.73, p = .042; musical toy condition: Med. = 0, IQR = [0, 1], z = 1.76, p = .039; Wilcoxon signed-rank tests here and below unless otherwise specified). These rates of reaching did not differ from one another (z = 0.58, p = .56; Wilcoxon-Mann-Whitney test) and a combined analysis of both conditions demonstrated that overall, infants reached more for the target object than the non-target object (Med. = 0, IQR = [0, 2]; z = 2.41, p = .008).

Secondary analyses included the data from all four trials. During familiarization trials, infants across both music exposure conditions showed equal looking to the two singers as they sang the songs (across all 16 trials, familiar song: M = 81.5 s, SD = 41.4 s; unfamiliar song: M = 83.7 s, SD = 42.0 s; t(63) = 1.00, p = .32). Reaching results were comparable, though somewhat weaker: infants in the musical toy condition reached significantly more to the target object than the non-target object (difference in number of reaches: Med. = 1, IQR = [0, 2], z = 2.30, p = .011), while results in the in the parental song condition were not significant (Med. = 0, IQR = [-2, 2.5], z = 1.25, p = .11). These rates of reaching did not differ from one another (z = 0.39, p = .70; Wilcoxon-Mann-Whitney test). As in the main analysis, across both conditions infants reached
more for the target object than the non-target object (Med. = 1, IQR = [-1, 2]; z = 2.29, p = .011).

Two exploratory analyses focused on the object choice data. First, we asked whether the main effects were attributable to infants' failure to attend to the non-target object. They were not: infants held both objects for comparable durations during the first two trials (target object: $M = 10.7$ s, $SD = 10.1$ s; non-target object: $M = 8.64$ s, $SD = 8.29$ s; $t(63) = 1.46$, $p = .15$) and across all trials (target object: $M = 22.2$ s, $SD = 16.4$ s; non-target object: $M = 22.1$ s, $SD = 18.2$ s; $t(63) = 0.04$, $p = .97$). They also looked comparably long at both objects during the first two reaching trials (target object: $M = 14.1$ s, $SD = 6.77$ s; non-target object: $M = 13.3$ s, $SD = 6.13$ s; $t(63) = 0.87$, $p = .39$) and across all reaching trials (target object: $M = 26.6$ s, $SD = 8.57$ s; non-target object: $M = 27.6$ s, $SD = 11.8$ s; $t(63) = 0.60$, $p = .55$). Unsurprisingly, the duration of gaze to an object and duration of holding that object were correlated in the first two trials (target object: $r = .50$, $p < .001$; non-target: $r = .46$, $p < .001$) and across all trials (target: $r = .32$, $p = .01$; non-target: $r = .63$, $p < .001$).

Second, we asked whether the strength of the main effect of reaching toward the target object differed on the basis of the type of objects presented (fruits or stuffed animals). On fruit trials (i.e., trials 2 and 4) infants reached significantly more frequently to the target object than to the non-target object
(Med. = 0, IQR = [0, 1], z = 2.47, p = .014). No such difference was found on animal trials (i.e., trials 1 and 3; Med. = 0, IQR = [-1, 1], z = 1.35, p = .18). Rates of reaching to the target object did not differ across object types, however (z = 1.12, p = .26).

**Predictive effect of song exposure.** Because parents were not given a quota for how often to present the song, the estimated degree of song exposure was variable across infants. The mean estimates did not differ across conditions (estimated number of song performances, parental song condition: \(M = 77.5, SD = 41.5\); musical toy condition: \(M = 71.9, SD = 34.6\); \(t(62) = 0.59, p = .56\)), nor was there any significant difference in the amount of variance on this variable across conditions (\(F(31,31) = 0.69, p = .32\), variance ratio test).

To correct substantial right skew, we log-transformed the exposure variable, and tested whether the log-transformed degree of song exposure predicted infants’ reaching for the target object. It did not: ordered logistic regressions predicting infants’ difference scores for reaching to the target vs. non-target objects on trials 1 and 2, from the amount of song exposure, were not significant in the parent singing condition (\(\chi^2(1) = 0.92, p = .34\)), the musical toy condition (\(\chi^2(1) = 0.04, p = .85\)), or overall (\(\chi^2(3) = 1.07, p = .30\)). Song exposure therefore appears to be unrelated to reaching behavior at test.

However, the proportion of reaching to target objects is necessarily a low-variance measure, in contrast to the continuous selective attention measure used
in previous work (Mehr et al., 2016). Thus, in exploratory analyses, we tested whether the amount of song exposure was predictive of the duration of gaze to the target object during the reaching trials: a measure more directly comparable to that used with five-month-old infants. Because the looking measure was continuous and approximately normally distributed, we used bootstrapped multiple linear regression, verified by sensitivity analyses.

Across all four trials, song exposure predicted the duration of gaze to the target object in the parental song condition (Figure 3A; $\chi^2(1) = 6.22, p = .013, R^2 = .21; \text{Wald test}$), such that a doubling of the amount of song exposure corresponded with an estimated 0.88 SD increase in the duration of gaze to the target object.

**Figure 3.** The scatterplots show infants’ duration of looking toward the target object, in (A) the parental song condition and (B) the musical toy condition, along with the predictive effects of the amount of song exposure (solid lines) +/- 2 standard errors (dashed lines) from a bootstrapped model with 40,000 replications. Note that the x-axes are on log₂ scales.
target object in that condition. In contrast, song exposure did not predict gaze to the target object in the musical toy condition (Figure 3B; $\chi^2(1) = 2.18, p = .14, R^2 = .05$). Thus, we continued by modeling the two conditions together, to test their interaction. The overall model was significant ($\chi^2(3) = 8.50, p = .037, R^2 = .16$), as was the condition by song exposure interaction term ($z = 2.81, p = .005$). These results held when including as a covariate the duration of gaze to the non-target object ($\chi^2(4) = 9.60, p = .048, R^2 = .17$), the difference in infants’ number of reaches to the target vs. non-target objects ($\chi^2(4) = 9.63, p = .047, R^2 = .18$), or both measures ($\chi^2(5) = 13.3, p = .021, R^2 = .22$). Tests of the interaction terms in each of these models survived a Bonferroni correction for six tests (i.e., adjusted alpha level of .0083); we corrected for six tests given the six exploratory models we ran here (i.e., each condition separately, both conditions together, and the three models with covariates). Crucially, these effects were driven by gaze during the portion of the trial after the actors pointed to the objects: the overall model of infants’ gaze before the pointing occurred yielded no significant effects ($\chi^2(3) = 1.05, p = .79, R^2 = .01$), but it held for gaze after the actors' pointing ($\chi^2(3) = 8.95, p = .030, R^2 = .156$), with a significant condition by song exposure interaction ($z = 2.81, p = .005$).

In sum, while reaching to the target object did not differ across song exposure conditions, gazing to the target object varied as a function of the
amount of song exposure in the parental song condition, but not in the musical toy condition.

Discussion

Infants preferentially reached to objects endorsed by a new person who sang a familiar song. Because this act is more explicitly interpretable as a social preference than the visual preferences observed in younger infants, infants' selective reaching replicates and extends previous findings concerning infants' social preferences for the singers of familiar songs (Mehr et al., 2016). In contrast to those findings, however, we observed preferential reaching to the target object regardless of whether the infant learned the song from a parent who sang or who activated a musical toy for them. Exploratory analyses nevertheless suggested differences in infant social preferences across these two song exposure conditions: infants' looking at the target object varied with song exposure only if that exposure came from a singing parent.

Why did infants show a social preference for the singer of the familiar song in the toy condition, whereas the five-month-old infants in our previous research did not? The previous negative finding is not likely attributable either to sampling error or to low sensitivity of the looking time measure, because the original experiments were well powered (power of .84), and experiments using the same measure have reliably detected social preferences in smaller samples of
young infants across multiple domains (e.g., Farroni, Csibra, Simion, & Johnson, 2002; Kinzler et al., 2007; Schachner & Hannon, 2011). The negative finding also is not due to younger infants' failure to attend to or remember toy-produced recorded songs: indeed, 5-month-old infants remembered the song produced by the toy many months later (Mehr et al., 2016).

Two remaining differences between the past and present experiments may account for their differing findings in the conditions with toy-produced songs. First, the style of music differed across the two experiments: our previous work used slow, soothing lullabies, whereas the current studies used faster, upbeat play songs. Infants may have been more socially engaged by the play songs than by lullabies, and also more predisposed to associate play songs than lullabies with toys. It is difficult to evaluate this explanation because data on the relative effects of lullabies vs. play songs are scarce. Second, decades of research have shown that infants' understanding of social interaction skyrockets around one year of life, especially in the context of interacting with social partners about objects (reviews: Carpenter et al., 1998; Tomasello, 2008). Thus, older but not younger infants may construe the musical toy as part of their social play with a parent, and a novel person singing the same song may inherit social meaning from this parent-infant social experience.
Both possibilities remain open, and further experiments may distinguish between them. In particular, it would be worthwhile to repeat the present experiments with younger and older infants while varying the types of music that infants hear. If the original finding is robust to changes in music type, it would lend support to the idea that 11-month-old infants are more apt than 5-month-old infants to garner social information from their interactions with parents in the context of recorded music. Further experiments also could unpack the particular social aspects of live song that drive infants' social preferences. Live singing includes a variety of rich behaviors that recorded song does not, including reciprocal gaze, smiling, and contingent interaction. It is not yet known whether the lack of such behaviors in current or previous musical toy conditions, or their reduced presence in the previous interactive video condition, could account for infants' social preferences (see Mehr et al., 2016, for further discussion).

Although we obtained positive and equal effects on the object choice test in the parent singing and musical toy conditions, exploratory analyses revealed that infants who had received more exposure to parental song (but not to toy-produced song) gazed longer, on average, at the object endorsed by the singer of the familiar song than those whose parents sang less frequently. This positive association in the parental song condition and its absence in the musical toy
condition is similar to effects obtained in the selective attention test used with five-month-old infants. In contrast to our previous findings, however, the effects did not reflect a preference for one object or the other, as infants looked to the two objects for comparable durations.

This difference may be attributable to differences in the timing of the measures. Preferential reaching for a toy is a function of the person who recommends the toy, the infant’s existing preference for one type of toy over another, and the infant’s previous experience with those toys. This measure is most directly comparable to our previous results, where infants' gaze represented a forced choice between the two actors' faces: a preference for one person over the other. Gaze toward a toy that a potential new social partner endorses may differ in meaning: if infants interpreted the actor's pointing as an invitation to share attention toward an object, the degree of their attention to that object might reflect their immediate social engagement with the actor, as opposed to a preference for that actor. The degree of this engagement was moderated by infants' degree of familiarity with the actor's song — but only when that song had been previously presented by a parent.

We suggest two interpretations of this finding. First, it may reflect differences in the speed with which infants habituate to live vs. recorded singing over time. Whereas a recording played repeatedly over 1–2 weeks becomes more
and more predictable, repetition in live singing may maintain or increase interest, as the musical features live song can be intentionally varied in response to the audience's interests and expectations. Indeed, musical repetition and redundancy have been proposed to be human universals (Brown, 1991). Thus, parents who sang more to their infants may have enjoyed the song more, enhancing its social value for infants.

Second, one of us has hypothesized that infant-directed song functions as parental investment in the form of attention (Mehr & Krasnow, 2017). Live singing on the part of a parent requires continuous investment of effort and attention throughout the song's duration, whereas toy-produced singing in the present experiment required only a single press of a button at the song's onset. By this hypothesis, therefore, the strength of infants' social preferences for new people who sing the same songs as their parents do might vary as a function of the effort the parent exerted in producing the song. If a new person demonstrates shared musical knowledge with a parent who has been providing reliable and frequent parental investment in the form of song, that new person is likely to provide future investment to the infant. These interpretations are speculative, however, and further research is needed to ascertain the sources of the music exposure effect.
Two further questions are raised but not answered by our findings. The first concerns a deflationary account of the present results. Might infants prefer a person associated with any familiar behavior, as in mere-exposure effects in adults (e.g., Zajonc, 2001)? We cannot yet rule out this account, but three points weigh against it. First, in the present experiments, while infants reached to the target objects at comparable rates across the parental song and musical toy condition, their patterns of visual attention to the target objects differed across conditions as a function of the degree of their exposure. Second, in our previous work, infants who learned a song from a friendly but otherwise unfamiliar singer via Skype demonstrated no visual preference for a new actor who sang that song, despite having learned it well enough to distinguish it from a second song with the same words and rhythms, but a different melody, some eight months later (Mehr et al., 2016). Third, in a previous study using the same reaching methods used here, White infants reached for objects endorsed by White or Black adults at comparable rates, despite having far more familiarity with White adults than with Black adults (Kinzler & Spelke, 2011). Each of these findings suggest that raw familiarity does not fully account for infants’ social preferences. Future experiments might more tightly control the song exposure conditions while directly manipulating the degree of that song’s social meaning to the infant; for
instance, infants might learn a song from a research assistant who visits the home regularly, but is otherwise unknown to the infant.

The second question concerns the uniqueness of the social effects of music. Is music "special," or might infants have shown comparable social preferences for new adults who shared infants' knowledge in other, non-musical domains? In these experiments we did not test the strength of music's ability to convey social information relative to other domains that are known to have similar effects, such as language (Kinzler et al., 2007) or food choice (Liberman et al., 2016). Infants' observations of behavior in these and other domains may well interact to produce varied responses to new social partners.

But our claim is more general. In our view, music is one of a class of behaviors that reliably inform infants' preferences for new social partners, because songs are learned from other people and are often sung in social contexts. Not all behaviors are expected to fall into this class: instrumental actions (e.g., breaking a rock), self-directed actions (e.g., scratching one's head), and unlearned actions (e.g., yawning), for example, should not. The present findings, taken together with our previous work (Mehr et al., 2016), support this view.
Chapter 4

Conclusion

This research provides evidence that music carries a social function in infancy. Five-month-old infants preferentially attended to a new person after she had sung a song the infants had learned from a parent; the degree of infants’ attentional preference was predictable from the amount of their parent’s singing. Despite their long-term memory for the song itself in both conditions, no such effects were found when infants had learned songs from a sound recording or from a new person’s singing via Skype. Eleven-month-old infants preferentially reached for objects presented and endorsed by new people who sang familiar songs, regardless of how infants had originally learned them (i.e., from a parent or from a sound recording). But the degree of their attention to those objects was predictable only from the amount of parental singing, and not from the number of times they had heard the sound recording.

Some methodological concerns

To justify the interpretation of these findings, four issues remain to be sorted out. First, the song learning conditions in Chapter 2 vary on many dimensions, not only on the key manipulation, that is, whether the original source of the song has any social relevance to the infant. While infants indeed learned the song in the musical toy and Skype conditions, their experience of
learning it was necessarily less rich than the experience of those infants in the parental song condition. Live song is more interactive, higher fidelity, more variable, and more multimodal than both recorded song and songs presented via Skype. These factors may have reduced infants' visual preferences for the singer of the familiar song in these conditions.

Second, the experiments in Chapter 2 do not disambiguate between the effects of live song and parental song. Because of logistical concerns, these factors were confounded in Experiment 3: in contrast to the singer in Experiment 1 (a parent), the singer who appeared once in person and subsequently via Skype was an adult whom the infant had never met. While the infant did meet this adult in person at the outset of the experiment, in the context of social interaction (e.g., along with the parent, while playing with toys), the obvious difference in the degree of social affiliation between a parent and a friendly but unfamiliar adult clouds the interpretation of infants' behavior at test.

Third, infants' social behavior at test in Chapter 2 was more unequivocally different across singing conditions than that of infants in Chapter 3. In Chapter 2, five-month-olds' selective attention contrasted sharply across song conditions, with positive effects only after parental song; in contrast, eleven-month-olds in the musical toy condition preferentially reached for objects endorsed by the singer of a familiar song regardless of exposure condition. This result may be
attributable to differences in the types of music used across these experiments (i.e., lullabies in Chapter 2 vs. play songs in Chapter 3); or to the fact that eleven-month-olds may have considered the recorded song to be a part of their social play with parents (given their more developed understanding of social interaction with objects; see, e.g., Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998); or to both.

Fourth, given the ambiguity between the main effects in Chapters 2 and 3, the present experiments do not yet fully rule out the possibility that raw familiarity (e.g., Zajonc, 2001) with songs — as opposed to the more complex social inferences of those songs that I posit as the focus of this work — can account for infants’ behavior. Several factors weigh against this interpretation (see Discussion in Chapter 3) but the present experiments leave this possibility open.

**How to resolve these concerns**

With the development of a replicable method for providing in-home music listening, reported in this thesis, each of these concerns can be directly evaluated. To address the possibility of differences in the quality of song exposure across parental song, sound recordings, and interactive video, one might repeat Experiment 1 in Chapter 2, and then one could compare its results to a new experiment in which infants learn the song from an unfamiliar adult
who sings live in the home, but has no other social interactions with the infant or with the infant’s group. If the social affiliation of the source of the song drives infants’ later social preferences, a negative effect is expected in the latter condition. That is, live song should not be sufficient to elicit social preferences. Live parental song might also be compared to songs presented via Skype by a known social partner of the infant who is not physically present (e.g., a known relative living abroad): if the social affiliation of the original singer is the key factor driving infants’ social preferences, then a positive effect can be expected in this condition.

To address the possibility that the eleven-month-olds in Chapter 3 interpreted their play with the parent and musical toy as conveying social meaning, the experiments should be repeated with efforts to minimize the social cues present in the toy condition. One way in which this could be carried out is by presenting the song without any physical stimulus: instead of an attractive toy with arms, legs, and a face (as in the present work), a parent’s smartphone could play back the song recording while displaying a static image of a shape, a boring screen saver, or the like. Parents could be instructed not to interact with
the infant while the song is playing, as in previous research\(^9\), further reducing the song’s social valence, but without impeding infants' ability to learn the song. Infants’ reaching would then be predicted to reflect the effects in Chapter 2.

Two further concerns are straightforward to address via replication: the possibility that song style and/or content drives infant social preferences, and the possibility that raw familiarity effects underlie the present results; that is, infants may simply prefer a new person who does anything that is familiar (see Zajonc, 2001). Direct replications of Chapter 2 experiments, replacing the lullabies with play songs, and including the same memory tests, would mitigate against both concerns.

**A note on interpretation**

Why should music serve a social function in infancy? Three mutually inclusive lines of reasoning provide some explanation. First, with Max Krasnow, I have proposed that music can function as parental investment in the form of attention (Mehr & Krasnow, 2017). Human infants are helpless and need parents to attend to them, ensuring their safety. Natural selection has thus designed the infant mind to elicit attention from parents, to be sated by attentional investment,

---

\(^9\) Studies of infant memory for music often take this approach, because a parent's interaction with the infant during exposure could alter infants' recall ability (positively or negatively). For instance, in Volkova, Trehub, and Schellenberg (2006), parents played recordings on a CD and were instructed not to interact with their infants or draw infants' attention to the music while it played; infants subsequently demonstrated robust memory for that music.
and to be sensitive to the honesty of the attentional signals that parents send (Dawkins & Krebs, 1979; Fernald, 1992; Trivers, 1972). Infants should thus be more calmed by costly forms of parental attention that are difficult to fake and more likely to constitute honest attention. Infant-directed song may have evolved as a form of attention fitting these criteria. It requires planning, memory, and physical effort; specialized and exclusive use of the vocal apparatus; and it includes features tailored to infants’ affective states in real time. On this hypothesis, infants should not only be attracted to infant-directed song, but should also be interested in future social interaction with people who produce it. This is especially so when new people sing songs that a parent has previously sung: this act demonstrates that the new adult is a potential source of the very same investment that a parent has reliably produced in the past, and an adult who provides investment once may well do so again in the future\(^{10}\).

Second, for much of ontogeny, infants have few communicative skills at their disposal. As a result, in the course of their daily social interaction, they must use receptive cues to make inferences about the people they meet. A fundamental question about these people is whether or not they are members of the infant’s social group. A growing body of work suggests that infants use a

\(^{10}\) In human ancestry, parental investment was shared across multiple adults who might not all have been related to a given infant (e.g., Hrdy, 2009). The infant is thus predicted to be motivated by parental investment whether or not the current source of that investment is, in fact, his parent.
variety of cues to make social inferences, including gaze direction (Farroni, Csibra, Simion, & Johnson, 2002), infant-directed speech (Schachner & Hannon, 2011), language or accent (Kinzler, 2013; Kinzler et al., 2007; Kinzler, Dupoux, & Spelke, 2012), and preferences for foods (Liberman et al., 2016). Might infant-directed song be such a cue? In human ancestral history, songs were passed down from person to person rather than learned via music notation, over the internet, or from recordings. Thus, if a new person sings a song to an infant, and if that song is already known to that infant from prior social interactions, the infant may infer that this new person has a social connection to him. On the basis of that inference, he might demonstrate a social preference for that new person, as in-group members are preferable to out-group members.

Whereas the previous two ideas appeal to ultimate-level explanations of human behavior that are reliant on assumptions about human evolution, a third, proximate-level explanation may also inform predictions about infants' musical inferences. This idea is simple: as with language, infants hear music in many social settings in the course of their first months of life, including from their parents, their families, from observations of others singing together, and so on. On this "learning" explanation, infants have had sufficient experience with music in social settings to infer that new individuals who sing to them are likely to be positive social partners. Thus, not only should they demonstrate preferences for
new people who sing to them, but their strongest social preferences should be for those people who sing songs that are already familiar to them from social contexts.

Neither the present experiments nor the proposed follow-ups can disambiguate between these three ideas, but understanding whether and how each contributes to infant music cognition may help to focus other lines of work concerning infants' responses to music in particular, and may elucidate music's ultimate functions and evolutionary origins in general. Elsewhere, I outline a series of predictions (Mehr & Krasnow, 2017) and provide new methods and cross-cultural datasets (Mehr et al., in prep) that will help to do both.

**Implications for education**

Scientific research should require no justification in terms of the future applications of its results but it is nonetheless my hope that the results of this research will be of use to parents, educators, and policymakers, in the service of providing infants and young children with developmentally appropriate musical experiences. I think this could happen in two ways: first, directly, by providing a clear understanding of how music works in infancy; and second, indirectly, by expanding scientific knowledge about what music is and where music comes from, thereby helping to secure music as a topic of serious scientific research in education, cognition, and culture.
As to the first category, this research suggests that infants are likely to engage socially with new people who sing familiar songs to them. Those charged with the care of young infants might thus do well to increase the frequency with which they sing to those infants, potentially matching their repertoires to those of infants’ parents. Parents who wish to broaden the styles of music to which their infants are receptive, and with which their infants are likely to engage, could seek out and perform more varied repertoires. Systematic differences in infant behavior after exposure to live music vs. recorded music suggests that these two forms of music, both of which appear regularly in classrooms (Flohr, 2005), are not interchangeable. Null results described in Chapter 2 demonstrate that the quality of parents’ singing is unrelated to infants’ social preferences; this suggests that infants may not be picky music listeners. Parents without music training thus need not worry that their inexperience with singing is detrimental to their infants’ interest and engagement with music and with those who produce it.

The potential indirect applications of this work are murkier, and I conclude by speculating on them. Research in the psychology of music from the past 150 years falls into two general categories: first, a rich body of work in music perception and psychoacoustics dating to Helmholtz (1885; reviews in Patel, 2008 and Trehub, 2010); and second, a smaller and more recent line of research on extrinsic (i.e., non-musical) effects of music listening and music training, as with
"Mozart Effect" and "Music Makes You Smarter" studies (for review, see Bangerter & Heath, 2004; Hetland & Winner, 2004; Mehr, 2015; Mehr, Schachner, Katz, & Spelke, 2013). Few scholars, however, have focused on music's ultimate functions, and as a result, bona fide theories of music's evolutionary origin have been few and far between (see discussion in Mehr & Krasnow, 2017; Patel, 2008; Pinker, 1997). Basic research in music perception is difficult to apply to education outcomes, and so music education advocates looking to science for ammunition have tended to focus on the potential for music education to engender positive effects on other areas of children's cognitive skills, rather than music education's content, universality, ubiquity, psychological functions, and so on.

I and others have argued that a focus on extrinsic outcomes of music education is a poor advocacy strategy (Mehr, 2013; Winner & Hetland, 2007), and I speculate that such a focus may have the unintended consequences of marginalizing and undercutting public music education, for two reasons. First, by implication, the approach implies that music education is frivolous, in that it must be justified on account of its effects on other areas (as opposed to, say, math education, which is important for its own sake). Similarly, one does not expect extrinsic academic benefits from athletic and physical education, drama, hygiene, home economics, shop class, sex ed, and so on. Second, if and when extrinsic outcomes are not achieved (i.e., if music does not, in fact, improve children's
cognitive abilities in non-musical areas), then music education advocates will be
at pains to demonstrate the value of music education (see discussion in Winner &
Cooper, 2000). Indeed, recent large-dataset studies have all but eviscerated the
possibility that music lessons engender improvements in cognitive abilities
unrelated to music (Elpus, 2013; Mosing, Madison, Pedersen, & Ullén, 2015).

A long-term goal of the research program in which my colleagues and I
are engaged is to shift the basic science of music toward ultimate questions —
including music’s evolutionary origins, universality across cultures, and
psychological functions throughout the lifespan — which may help to better
establish music as a cornerstone of the human experience. If and when that
happens, music education advocates will have a wealth of empirically-derived
and replicable evidence to draw upon, building on prior work in perception and
psychoacoustics, to the benefit of children and families worldwide.
Bibliography


https://www.youtube.com/watch?v=9bZkp7q19f0.

Quam, R. M., Ruiter, D. J. de, Masali, M., Arsuaga, J.-L., Martínez, I., & Moggi-
*Proceedings of the National Academy of Sciences USA* 110, 8847–8851.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially
contingent interactions help toddlers learn language. *Child Development* 85,
956–970.


reveal the structures and functions of human music. *Proceedings of the
National Academy of Sciences USA* 112, 8987–8992.

http://www.theverge.com/2014/12/3/7325819/gangnam-style-broke-
youtube-view-counter


Development* 43, 549–565.


