Journal of Experimental Child Psychology 228 (2023) 105606

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Young children's attributes are better conveyed by voices than by faces

Carlos Hernández Blasi [a,*], David F. Bjorklund [b], Sonia Agut [a], Francisco Lozano Nomdedeu [a], Miguel Ángel Martínez [a]

[a] Departamento de Psicología, Universitat Jaume I, 12071 Castellón, Spain
[b] Department of Psychology, Florida Atlantic University, Boca Raton, FL 33431, USA

A R T I C L E   I N F O

Article history:
Received 24 February 2022
Revised 18 November 2022
Available online 17 December 2022

Keywords:
Voices
Faces
Caregiving
Early childhood
Evolution
Evolutionary developmental psychology

A B S T R A C T

The purpose of this study was to explore how young children's vocal and facial cues contribute to conveying to adults important information about children's attributes when presented together. In particular, the study aimed to disentangle whether children's vocal or facial cues, if either, are more dominant when both types of cues are displayed in a contradictory mode. To do this, we assigned 127 college students to one of three between-participants conditions. In the Voices-Only condition, participants listened to four pairs of synthetized voices simulating the voices of 4-5-year-old and 9-10-year-old children verbalizing a neutral-content sentence. Participants needed to indicate which voice was better associated with a series of 14 attributes organized into four trait dimensions (Positive Affect, Negative Affect, Intelligence, and Helpless), potentially meaningful in young child–adult interactions. In the Consistent condition, the same four pairs of voices delivered in the Voices-Only condition were presented jointly with morphed photographs of children's faces of equivalent age. In the Inconsistent condition, the four pairs of voices and faces were paired in a contradictory manner (immature voices with mature faces vs. mature voices with immature faces). Results revealed that vocal cues were more effective than facial cues in conveying young children's attributes to adults and that women were more efficient (i.e., faster) than men in responding to children's cues. These results confirm and extend previous evidence on the relevance of

* Corresponding author.
E-mail address: blasi@uji.es (C. Hernández Blasi).

https://doi.org/10.1016/j.jecp.2022.105606

children's vocal cues to signaling important information about children's attributes and needs during their first 6 years of life.

Introduction

Human faces and voices are special for infants. From birth, babies show a strong and consistent attentional bias toward facial and vocal features that, in turn, becomes steadily and progressively tuned across early development thanks to their extensive and continued exposure to conspecifics' faces and voices (Orena & Werker, 2021). For example, newborns prefer face-like patterns over non-face patterns, direct gaze over averted gaze, and open eyes over closed eyes (e.g., Batki et al., 2000; Farroni et al., 2002; Goren et al., 1975). They also prefer infant-directed speech over adult-directed speech, their mother's voice over the voice of another woman, and familiar voices over unfamiliar voices (e.g., Cooper & Aslin, 1990; DeCasper & Fifer, 1980; Naoi et al., 2012). This should not be surprising; human faces and voices are rich sources of socially relevant information (Belin et al., 2011). This social information can be critical both for infants' survival and for the acquisition of culturally relevant knowledge that infants will need to navigate their world (Bettle & Rosati, 2021; Hernández Blasi, 2020; Thiele et al., 2021).

Conversely, infants' faces and voices are special for their caregivers because these cues can provide caregivers with relevant information about infants' condition, needs, and emotional states. For instance, neonatal crying might signal phenotypic quality to parents (Furlow, 1997; Lummaa et al., 1998; Soltis, 2004), and the different types of infants' crying convey acute information about their needs and emotional states to adults (Wolff, 1969). In addition, infants with rounded heads and cheeks, flat noses, and adult-size eyes, all facial features included in the *baby schema* or *Kindchenschema* described by Lorenz (1943), are typically more attractive to adults, increasing the chances of caregiving behavior (Franklin & Volk, 2018; Glocker et al., 2009; Senese et al., 2013).

The onset of language changes child–caregiver interactions substantially during early childhood—the period of human life following infancy until about 6 years of age. Speech makes easier children's communication of their needs to adults and facilitates social learning in both direct and indirect ways. In this vein, for example, young children who verbalize some forms of magical explanations about a natural phenomenon as described by Piaget (e.g., *animism:* "The sun's not out today because it's mad") evoke in adults and older adolescents more positive affect and helpless impressions, as well as cues that these children are less intelligent, than young children who verbalize more mature adult-like explanations (e.g., "The sun's not out today because the clouds are blocking it") (Bjorklund et al., 2010; Periss et al., 2012). These verbalized cognitive cues, which have been labeled as *supernatural thinking,* are also more informative for adults and older adolescents than facial features of the same children when they are presented together, particularly regarding the appraisal of intelligence and helpless traits (Hernández Blasi & Bjorklund, 2018; Hernández Blasi et al., 2015, 2017).

However, even during this early childhood stage, young children's vocal features by themselves, regardless of speech content, seem to be more informative for adults than young children's facial features when both vocal and facial cues are considered independently. For example, Hernández Blasi and colleagues (2022) reported that 5-year-old children who verbalized content-neutral sentences (e.g., "I like the beach more than the mountains") evoked more positive affect and helpless feelings from adults, and were judged to be less intelligent, than 10-year-old children who verbalized the same content-neutral sentences. In contrast, adults and older adolescents typically experience more positive affect toward young children's faces than older children's faces but feel unable to determine which children are more intelligent or helpless based solely on their facial features (Hernández Blasi & Bjorklund, 2018; Hernández Blasi et al., 2015, 2017).

The current study

The purpose of this study was to explore how young children's vocal and facial cues convey to adults important information about children's attributes when presented together and particularly to disentangle which one, if either, is more dominant when both types of cues are presented in a conflicting manner.

As reviewed earlier, currently available evidence indicates that facial, vocal, and cognitive cues are equally effective in transmitting positive-affect impressions about young children to adults when presented independently (Bjorklund et al., 2010; Hernández Blasi et al., 2022). Cognitive and vocal cues are also highly effective in communicating information about intelligence and helpless traits when considered alone, which is not the case for facial cues (Bjorklund et al., 2010; Hernández Blasi et al., 2015, 2022). In addition, cognitive cues seem to be more effective than facial cues in conveying information about young children's intelligence and helpless traits to adults and older adolescents when both are simultaneously available (Hernández Blasi & Bjorklund, 2018; Hernández Blasi et al., 2015, 2017).

In the current study, we presented college students with pairs of photos with morphed faces resembling the physical appearance of approximately 4- to 7-year-old and 8- to 10-year-old children taken from Hernández Blasi et al. (2015). Participants were asked to rate children for Positive-Affect, Negative-Affect, Intelligence, and Helpless behavioral dimensions. In one condition (Consistent condition), we presented morphed face photos matched with synthesized voices of children of an equivalent age verbalizing a neutral-content sentence (e.g., "I like the mountains more than the beach"). In a second condition (Inconsistent condition), we presented the same pairs of photos and voices used in the Consistent condition but matched inconsistently; that is, we paired 4- to 7-year-old morphed faces with 8- to 10-year-old synthesized voices versus 8- to 10-year-old morphed faces paired with 4- to 7-year-old synthesized voices. Finally, in a control condition (Voices-Only condition), the pairs of synthetized voices used in both the Consistent and Inconsistent conditions were presented to participants without morphed face photos. In addition to rating the participants' impressions about the children, we also measured their reaction times to make their decisions in the three conditions. We did not include a Faces-Only condition in this experiment because past research has been consistent about adults' and older adolescents' outcomes. However, for comparative purposes, in Table 1 (see Results), we present data obtained from a Faces-Only condition tested in a previous study (Hernández Blasi et al., 2017), designed with the same set of morphed face photos and behavioral dimensions used here.

We hypothesized, first, that in the Consistent condition children with immature voices and faces would evoke more Positive Affect than children with mature voices and faces. Moreover, given that immature facial and vocal cues typically produce a greater Positive-Affect effect on adults than mature facial and vocal cues when considered singly, we expected that the magnitude of Positive Affect evoked in this condition would be significantly higher, and/or the reaction times would be faster, than when each cue was considered separately. In other words, we anticipated that the combination of immature facial and vocal cues would yield stronger and/or faster participants' decision making on Positive-Affect items. Regarding the Intelligence and Helpless trait dimensions, we predicted that in the Consistent condition the results would be similar to those in the Voices-Only condition given that in previous research facial cues have not been particularly informative on these two dimensions for adults, with children with mature faces and voices being deemed as more intelligent and children with immature faces and voices being perceived as more helpless. Finally, regarding the Negative-Affect dimension, we did not expect significant differences between children with mature versus immature facial and vocal cues given that previous research has shown that adults are often reluctant to appraise children in negative terms based on the maturity of faces or voices (e.g., Hernández Blasi et al., 2015, 2022). It should be noted, however, that adults and adolescents do rate hypothetical children expressing some forms of immature cognition—*natural thinking*—higher on negative affect than children expressing the mature forms of such cognition (Bjorklund et al., 2010; Periss et al., 2012).

Slightly different patterns were predicted for the Inconsistent condition. For the Negative-Affect trait dimension, we did not anticipate significant differences in adults' ratings, similar to what was predicted for the Consistent condition and for the same reason (adults' reluctance to appraise children in negative terms). However, for the Intelligence and Helpless dimensions, we predicted that voices would produce a stronger effect relative to faces (i.e., greater selection of the mature/immature voice). This prediction was based on previous research showing that immature voices are more apt to produce differences in ratings for the Intelligence and Helpless dimensions than immature faces (e.g., Hernández Blasi & Bjorklund, 2018; Hernández Blasi et al., 2022). Lastly, predictions for the Positive-Affect dimension in the Inconsistent condition remained open. One possibility was that there would be no significant differences between children with immature faces *and* mature voices and children with mature faces *and* immature voices (producing a chance effect on the overall results). We might expect this outcome based on the findings from previous studies in which both immature faces and immature voices considered alone evoked similar responses on these dimensions in adults (see, e.g., Hernández Blasi et al., 2017, 2022). Alternatively, one feature (face or voice) may be a more potent cue in eliciting Positive Affect than the other, generating more Positive Affect for children with immature faces *and* mature voices (if faces were more potent than voices) or for children with immature voices *and* mature faces (if voices were more potent than faces).

With respect to reaction times, we hypothesized that overall participants' speed of decision making would be slowest in the Inconsistent condition and fastest in the Voices-Only condition, with the Consistent condition falling in between. We hypothesized this rank order because we assumed that participants would be faster to make decisions on the basis of a single kind of cue (Voices-Only condition) than on the basis of two different kinds of cues, particularly when they are contradictory. We also predicted, in agreement with previous research on children's voices (Hernández Blasi et al., 2022), that reaction times for the Negative-Affect dimension trait would be slower than reaction times for the other three traits.

Finally, we examined sex differences. Historically and across cultures, women have been the primary caretakers of children (Hrdy, 1999; Konner, 2010), and previous research has shown that women are more sensitive to differences in the baby schema than men (e.g., Sprengelmeyer et al., 2009), making it possible that women are also more sensitive or receptive to cues of immaturity in young children. However, previous research has produced null, small, or inconsistent findings regarding sex differences using the current paradigm (e.g., Hernández Blasi & Bjorklund, 2018; Periss et al., 2012), and thus we made no predictions concerning possible sex differences for the current experiment.

Method

Participants

The sample consisted of 127 adults (76 female; M_{age} = 19.63 years, *SD* = 2.81, range = 18–40) attending the School of Education at a public urban university in eastern Spain. This sample size was selected to be comparable to other previous studies in the field (e.g., Hernández Blasi & Bjorklund, 2018; Hernández Blasi et al., 2017). A post-hoc power analysis using G*Power software (Faul et al., 2007) revealed that this size was sufficient to detect a medium effect with 99% power (α = .05) for the main comparisons. The participants' socioeconomic backgrounds were mainly middle class, typical of public universities in Spain. Although the participants were in the process of becoming primary education teachers, their formal experiences with children were still relatively infrequent at their current level of training (sophomores) and age. (In Spain, nowadays women have their first child when they are on average about 32 years old; Instituto Nacional de Estadística [INE], 2022) Many participants still lived with their parents when they were tested for this study, even with some siblings at home, but as typically happens with emergent adults around the world (Arnett, 2000), others shared apartments with other students, lived in university dormitories, or cohabitated with a romantic partner. All participants volunteered and were tested individually at the researchers' laboratory after signing an informed consent form. The study was approved by the University Jaume I Research Ethics Committee.

Design

Participants were assigned to one of three between-participants conditions (Voices-Only, Consistent, or Inconsistent) presented through a computer-delivered questionnaire. Two different versions of the questionnaire were created for each condition. For every condition, participants needed to make a series of decisions about four stimulus pairs. Each stimulus pair displayed two hypothetical children, and participants needed to decide which one better illustrated each of a series of 14 traits (e.g., nice, intelligent, helpless, sneaky). For every condition, two hypothetical pairs of boys and two hypothetical pairs of girls were presented.

In the Voices-Only condition (n = 42), each of the four stimulus pairs consisted of a synthetized immature voice and a mature voice of a child verbalizing a neutral-content sentence (e.g., "I like the mountains more than the beach"). Following the methodology reported in detail in Hernández Blasi et al. (2022), the immature-voice version simulated the voice of an approximately 4- or 5-year-old child, whereas the mature-voice version simulated the voice of an approximately 9- or 10-year-old child. On average, immature voices were spoken at a pitch of 287.68 Hz (SD = 3.85; as expressed by fundamental frequency) for 3.50 s (SD = 0.08), whereas mature voices were spoken at a pitch of 232.20 Hz (SD = 4.12; as expressed by fundamental frequency) for 2.54 s (SD = 0.15). These acoustic parameters (pitch and duration) reflected the acoustic parameters obtained from a sample of voices of 42 4- to 11-year-old children reported in Hernández Blasi et al. (2022) for these two age groups. Overall, these parameters were consistent with previous literature (see, e.g., Berger et al., 2019; Nip & Green, 2013; Trollinger, 2003). Indeed, in Hernández Blasi et al. (2022), on average, 88% of the samples of voices used for the Natural-Voices condition and 80% of the samples of voices used for the Simulated-Voices condition were positively identified by four external referees regarding the real and intended ages of the children. All voices were equalized to about 72 dB volume, which is within the typical range for spoken voices in Spain (65–75 dB) (Bustos, 2012).

In the Consistent condition (n = 44), the same four voice pairs used in the Voices-Only condition were matched with morphed photographs of children's faces of equivalent age. That is, each immature voice, reflecting a voice of a preschool-aged child, was associated with a morphed photograph of a preschool (immature) child's face, whereas each mature voice, reflecting a voice of a school-aged child, was matched with a morphed photograph of a school-aged (mature) child's face. These morphed faces were taken from a previous study where we manipulated eight facial features of the portraits of 26 children by using the Face Filter Studio 2 face-morphing software to resemble the faces of an approximately 4- to 7-year-old child and an approximately 8- to 10-year-old child. A more detailed description of the procedures for creating and selecting children's face photographs is reported in Hernández Blasi et al. (2015). Therefore, participants in this condition needed to make decisions about which of the two children involved in each stimulus pair (the immature child: immature voice + immature face, or the mature child: mature voice + mature face) would illustrate better each of the 14 traits.

Conversely, in the Inconsistent condition (n = 41), the same four stimulus pairs used in the Voices-Only condition were *inconsistently* matched with the morphed photographs of children's faces. That is, the immature children's voices were associated with the mature children's faces, whereas the mature children's voices were matched with the immature children's faces. In this way, participants needed to make their decisions on the trait attribution task evaluating the children with school-aged faces speaking with a typical preschool-aged voice and the children with preschool-aged faces speaking with a typical school-aged voice.

The 14 traits or short descriptions of children, which were sequentially presented to the participants after each stimulus pair display, correspond to a wide range of characteristics that could be potentially meaningful in adult–young child interactions. According to principal component analysis performed in our previous studies (e.g., Hernández Blasi et al., 2017; Periss et al., 2012), we organized these traits into four groups: Positive Affect (cute, friendly, nice, and likeable), Negative Affect (sneaky, likely to lie, feel more irritated with, and feel more angry with), Intelligence (smart and intelligent), and Helpless (helpless, feel more protective toward, and feel like helping). One trait (curious) did not load highly onto any factor and was not included in later analyses.

The design and delivery of the three conditions was made through E-Prime 2.0 professional software. This software also made it possible to counterbalance the order of presentation of (a) the four stimulus pairs within each set, (b) the first item within every stimulus pair, and (c) the 14 traits or short descriptions of children.

Procedure

The experiment was delivered individually via a computer in a university laboratory. Participants were assigned by the experimenter to one of the two versions designed for each of the three between-participants conditions. Participants sat before an Acer V193 HQV LCD 18.5-inch wide monitor and were provided with SHURE SRH440 adjustable headphones to listen to the auditory stimuli. Birthdate, sex, and university degree were the only personal information collected.

Before beginning the experiment, participants read some preliminary information about the experiment on the computer screen and how to use the computer keyboard. They were told that they were to press the key with a yellow sticker on it (over the Z of a QWERTY keyboard) to select stimuli located on the left side of the screen, whereas they were to press the key with a green sticker on it (over the M of the keyboard) to select the stimuli located on the right side of the screen. Participants were also told that they would have the chance to listen to the auditory stimuli of either child as many times as they wanted after both children's voices and faces had been presented and a question had appeared at the bottom of the screen by pressing any of the two red stickers placed over the number 1 (to listen to the child located on the left side of the screen) and the number 0 (to listen to the child located on the right side of the screen) on the keyboard. The onset for measuring each reaction time started when the trait question was displayed to the participants at the bottom of the screen and concluded when the participants made a selection on the keyboard of one of the two children. Reaction times were not measured if and when participants listened again to one of the children's voices. Finally, participants were informed that for each screen (a) there were no correct or incorrect responses, (b) it was not possible to leave questions unanswered, and (c) it was advisable not to overthink their responses. After reading this information, participants were given two practice trials to familiarize them with the experimental procedure.

As mentioned in the "Design" section, participants were presented sequentially a series of four stimulus pairs. In the Voices-Only condition, the two stimuli of each pair were auditory; participants listened to (a) the voice of a young child verbalizing a neutral-content sentence (e.g., "I like the mountains more than the beach") and (b) the same sentence verbalized by an older child. Participants then needed to select which of the two children illustrated better a series of 14 traits or short descriptions presented one at a time on different screens. Once participants made a decision on the first stimulus pair, the second stimulus pair was presented and so forth. A detailed description of the full experimental sequence for this condition can be found in Hernández Blasi et al. (2022).

In the Consistent and Inconsistent conditions, participants were presented with photographs of (a) a young child's face while listening to a child verbalizing a neutral-content sentence (the same neutral-content sentences used in the Voices-Only condition) or (b) an older child's face while listening to a child verbalizing the same sentence (the size of photographs was 10 cm width $\times$ 15 cm height). Then, as in the Voices-Only condition, participants needed to make their decisions, one at a time, about which of the two children of each pair fitted better the 14 descriptive traits. In the Consistent condition, the age of each child's face was matched with the age of the child's voice (i.e., a 5-year-old's face paired with a 5-year-old's voice vs. a 10-year-old's face paired with a 10-year-old's voice); in the Inconsistent condition, the maturity of the faces and voices was mismatched (i.e., a 5-year-old's face paired with a 10-year-old's voice vs. a 10-year old's face paired with a 5-year-old's voice). The core experimental sequence for each stimulus pair for these two conditions was identical (see Fig. 1).

Results

Participants' responses were coded as 1 when they selected the child with the immature voice, regardless of the condition, and as 0 when they selected the child with the mature voice. Thus, mean scores significantly greater than 0.5 indicate that participants selected more often the child with the immature voice, whereas mean scores significantly less than 0.5 indicate that participants selected more often the child with the mature voice. Preliminary analyses examining the number of times per trait that participants chose to re-listen to a voice indicated that overall levels of re-listening were

Fig. 1. Core audiovisual sequence presented to participants for the Consistent and Inconsistent conditions. In the first screen, 5 s after the upper instruction appeared, an icon and a child's photograph (e.g., a young child) came into view at the middle-left side of the screen, and then a neutral sentence was uttered (voice could be consistent or not with child's age). In the second screen, 1 s later the second icon and child's photograph (from a different age; in this example, an older child) appeared and the same neutral sentence was uttered (in this example, consistent or inconsistent with an older child's voice). In the third screen, after the participant pressed a key of the keyboard to continue, the first question appeared at the bottom of the screen. Once the first question was answered, 13 more questions with their corresponding adjectives or short statements were delivered sequentially, in random order, at the same screen place. (Photos taken from Hernández Blasi et al., 2015, and reproduced with permission.)

low (M = .06, SD = .15) and produced no significant main or interactive effects involving conditions (Ms = .08, .03, and .08, SDs = .17, .10, and .17, for the Voices-Only, Consistent, and Inconsistent conditions, respectively), and therefore these data were not considered further.

We first applied a series of two-tailed, single-sample t tests to assess whether children with immature or mature voices were selected greater than expected by chance (0.5). Table 1 presents the mean proportion of participants who selected the child with the immature voice by trait dimension (Positive Affect, Negative Affect, Intelligence, or Helpless) and condition (Voices-Only, Consistent, or Inconsistent) as well as whether items for each dimension were selected significantly greater than expected by chance ($p \leq .004$ adjusting for multiple contrasts). For comparison purposes, Table 1 also includes the proportion of participants who selected children on the basis of faces alone (Faces-Only condition) from Hernández Blasi et al. (2017). Table 1 also incorporates the mean reaction time for each condition and trait dimension. As can be seen in the table, the pattern of results was nearly identical for the three conditions; children with immature voices elicited significantly more Positive-Affect and Helpless feelings from participants than children with mature voices, whereas children with mature voices

Table 1
Proportions of participants selecting the child with the immature voice (or immature face in Faces-Only condition) and mean reaction times per trait (in milliseconds) for trait dimension and condition.

	Positive Affect (4 items)	Negative Affect (4 items)	Intelligence (2 items)	Helpless (3 items)
Voices-Only (n = 42)				
Proportion	.73[a] (.21)	.45 (.28)	.21[b] (.24)	.84[a] (.20)
Reaction time	1998.91 (675.85)	2416.85 (988.35)	1942.67 (713.76)	1903.59 (685.19)
Consistent (n = 44)				
Proportion	.62[a] (.25)	.42 (.28)	.18[b] (.22)	.82[a] (.24)
Reaction time	2092.67 (706.83)	2391.15 (828.44)	1997.20 (690.17)	2059.53 (685.63)
Inconsistent (n = 41)				
Proportion	.60[a] (.19)	.44 (.29)	.27[b] (.26)	.78[a] (.25)
Reaction time	2118.86 (620.12)	2337.29 (616.22)	2279.95 (643.52)	2069.80 (701.30)
Faces-Only (n = 23)				
Proportion	.75[a] (.23)	.58 (.28)	.46 (.36)	44 (.21)
Reaction time	2061.33 (562.83)	2429.21 (894.54)	2054.25 (688.67)	2865.75 (901.73)

Note. Standard deviations are in parentheses. Significance set at $p \leq .004$. Faces-Only data taken from Hernández Blasi et al. (2017) for comparative purposes. [In Hernández Blasi et al., 2015, where the sample size for the Faces-Only condition was larger (n = 66), results were comparable: Positive Affect, .68[a] (.21); Negative Affect, .57 (.26); Intelligence, .51 (.26); Helpless, .47 (.29), although reaction times were not taken.]

[a] Selecting an immature child significantly greater than expected by chance.

[b] Selecting a mature child significantly greater than expected by chance.

were deemed to exhibit significantly more Intelligence. Selections of both the mature and immature voices did not vary from chance for the Negative-Affect items for any of the three groups.

To assess further the pattern of performance, we conducted a 2 (Sex: female vs. male) × 3 (Condition: Voices-Only vs. Consistent vs. Inconsistent) × 4 (Trait Dimension: Positive Affect vs. Negative Affect vs. Intelligence vs. Helpless) analysis of variance, with repeated measures on trait dimension, for the proportion of participants who selected children with immature voices. We performed an identical analysis for reaction times.

The analysis of variance on the proportion of participants who selected children with immature voices produced significant main effects of condition, $F(2, 121) = 4.45$, $p = .014$, $\eta_p^2 = .07$ (Voices-Only, $M = .56$ > Consistent, $M = .50$ = Inconsistent, $M = .52$), and trait dimension, $F(2.78, 336) = 115.48$, $p < .001$, $\eta_p^2 = .49$ (Helpless, $M = .81$ > Positive Affect, $M = .64$ > Negative Affect, $M = .44$ > Intelligence, $M = .22$). The main effect of sex and all interactions were not significant, $F\text{s} \leq 1.98$, $p\text{s} \geq .10$, $\eta_p^2\text{s} \leq .03$.

The analysis of variance of the reaction times yielded significant main effects of sex, $F(1, 121) = 5.94$, $p = .016$, $\eta_p^2 = .05$ (men, $M = 2277.46$ ms > women, $M = 2033.19$ ms), and trait dimension, $F(3, 363) = 11.88$, $p < .001$, $\eta_p^2 = .09$ (Negative Affect, $M = 2397.02$ ms > Positive Affect, $M = 2100.35$ ms = Intelligence, $M = 2082.63$ ms = Helpless, $M = 2041.29$ ms). The main effect of condition and all interactions were not significant, $F\text{s} \leq 1.58$, $p\text{s} \geq .15$, $\eta_p^2\text{s} \leq .02$.

Discussion

The main purpose of the current study was to investigate how vocal and facial cues provided by young children inform adults about children's attributes when both types of cues are available. Which type of cue, if either, would be more potent for adults when they were in conflict, that is, when vocal and facial cues provided contradictory information? Our results revealed two new insights. First, overall, vocal cues seem to be more *effective* (potent) than facial cues for influencing adults' impressions about young children. Second, women apparently process more *efficiently* (faster) the information rendered by vocal and facial cues regarding these children than men.

Our initial hypotheses concerning the Consistent condition were confirmed with respect to the Negative-Affect, Intelligence, and Helpless trait dimensions but not for Positive Affect. As predicted, participants rated children with mature voices and faces higher in terms of Intelligence and rated children with immature voices and faces more Helpless than children with mature voices and faces. Given that in previous studies (Hernández Blasi & Bjorklund, 2018; Hernández Blasi et al., 2015) neither adults nor adolescents made clear-cut decisions on these two dimensions based only on the information provided by facial cues, the current data are consistent with the interpretation that vocal cues, rather than facial cues, governed adults' decision making. In addition, as initially predicted and as typically found in other studies (Hernández Blasi et al., 2015, 2017, 2022), no differences were found in ratings of the Negative-Affect items.

With respect to Positive Affect, we found that the redundant information provided by both vocal and facial cues on this trait dimension did not produce an enhanced effect or faster decision making relative to either the Voices-Only condition here or the Faces-Only condition in other studies. In fact, although not significantly different, the magnitude of the means for selecting children with the immature voice was slightly greater in the Voices-Only condition (73%) compared with the Consistent condition (62%). These rates are similar to those reported in previous research, where 68% (Hernández Blasi et al., 2015) and 75% (Hernández Blasi et al., 2017) of participants selected children with immature faces in a Faces-Only condition. That is, the percentage of participants who experienced more positive affect toward children with immature voices *and* faces presented together was no greater than the percentage of participants who experienced the same feelings toward children with immature voices or faces considered independently. Similarly, reaction times were comparable for the Positive-Affect items in the Voice-Only (1999 ms) and Consistent (2093 ms) conditions (and were similar to reaction times for a Faces-Only condition in Hernández Blasi et al., 2017, 2061 ms), demonstrating that multimodal cueing did not facilitate speed of processing relative to single-mode cueing.

For the Inconsistent condition, where the information from vocal and facial cues contradicted each other (i.e., children with immature voices were paired with children with mature faces and vice versa), our expectations were less certain. Overall, children's vocal cues dominated over children's facial cues in participants' decision making for each trait dimension. Compelling evidence of the dominance of voices over faces can be seen in Table 1, where the scores for the Inconsistent condition for each of the four trait dimensions were nearly identical to the scores obtained in both the Voices-Only and Consistent conditions. Moreover, a closer inspection of the scores for the Positive-Affect, Intelligence, and Helpless trait dimensions indicates that the alternative hypothesis (i.e., dominance of children's facial cues) is not tenable for any dimension. For example, if ratings for the Inconsistent condition in Table 1 had been coded on the basis of the immaturity/maturity of the children's faces instead of the immaturity/maturity of children's voices (as done in Table 1), scores for the Positive-Affect, Intelligence, and Helpless traits would have been .40, .73, and .22, respectively. Those scores are completely at odds with the typical profile obtained for these three dimensions in the Faces-Only condition in other studies (e.g., .75, .46, and .44, respectively, in Hernández Blasi et al., 2017), making unlikely the possibility that facial cues had driven participants' decision making in this condition. The only potential exception to this interpretation is the Negative-Affect trait dimension, where the score (.44) is comparable to those in Voices-Only and Faces-Only conditions, where participants have been reluctant to appraise children of any age negatively based on either vocal or facial cues alone.

To our knowledge, this is the first time that the dominance of vocal cues over facial cues on adults' impression formation of young children has been shown. Hitherto, there was extensive evidence in the literature indicating that adults typically and spontaneously attribute a series of traits to adult strangers based exclusively on their facial features (e.g., trustworthiness, honesty, competence, intelligence, aggression, likeability) (e.g., Todorov et al., 2015). We also know that these attributions, which seem to strongly influence people's daily behaviors, tend to be highly consistent in both Western and non-Western cultures (e.g., China; Sutherland et al., 2018), are made relatively quickly (e.g., adults can form these impressions even when faces are presented for as little as 100 ms; Todorov et al., 2009), and seem to emerge as early as 5 years of age (Charlesworth et al., 2019). Such evidence has caused some scholars to suggest that first-impression formation based on faces might have been favored by natural selection (see, e.g., Ewing et al., 2019; Schaller, 2008; Zebrowitz & Zhang, 2011; but see Eggleston et al., 2021, for an alternative view). Nevertheless, evidence of the effects of children's vocal cues on adults' first-impression formation, whether considered alone or jointly with children's facial cues, suggests that facial cues are not the only source of impression formation when dealing with young children and likely are not the dominant one (Hernández Blasi et al., 2022). Rather, our results are consistent with proposals that voices provide more information about people (or at least children) than faces in multimodal presentations and that "aside from conveying semantic information about the spoken message, the other important role of voices is to allow people to infer socially relevant visual information about the speaker, such as information about masculinity/femininity, body size, health, and age" (Smith et al., 2016, p. 7).

With regard to our hypotheses about reaction times, our first prediction about Negative-Affect traits being more demanding for participants than traits from the other three dimensions was confirmed. At first glance, this result could be explained by arguing that adults' reluctance to evaluate children negatively might have resulted in more time-consuming decision making. Indeed, this rationale would be consistent with some previous research involving mature and immature children's voices (Hernández Blasi et al., 2022) or children verbalizing supernatural thinking (Hernández Blasi et al., 2017). However, this interpretation would be against evidence indicating that reaction times for the Negative-Affect trait dimension have not always been the slowest ones in similar research. For example, in Hernández Blasi et al. (2017), reaction times for Helpless items were slower than those for Negative-Affect items when participants were presented only with mature and immature children's faces or for children verbalizing different forms of natural thinking. Therefore, it seems that rating children with negative terms is not always a difficult and time-consuming process for adults. Adults' reluctance to rate children in negative terms might be a domain-specific aptitude rather than a domain-general one, depending on the type of cues, context, and/or information given by children.

In contrast, our second hypothesis, predicting that reaction times for the Voices-Only condition would be faster than reaction times for the other two conditions where both voices and faces needed to be processed—with the Inconsistent condition taking more time to be processed than the Consistent condition—was not confirmed. In fact, no significant difference among the three conditions was found in our analyses, suggesting that no condition was more demanding than another in cognitive terms. This finding replicates Hernández Blasi et al.'s (2017) results, where no differences in reaction times were found across conditions that seemingly presented different attentional loads for participants. One explanation for this pattern of results, which is consistent with the failure to find additive effects for selecting the Positive-Affect items in the Consistent condition, is that when two types of cues about the same child are provided (e.g., faces and cognition; faces and voices), participants make a decision of which one to focus on and which one to disregard in advance, such that only one source of information is processed, and the reaction times become quite similar across conditions for each trait dimension.

An unpredicted finding in this study was a significant sex difference for reaction times: overall, women were faster than men in their decision making across trait dimensions and conditions. We should note, however, that the magnitude of the sex difference and effect size found here was small ($p = .016$, $\eta^2_p = .05$), and these sex differences in reaction times were not accompanied by sex differences in performance (scores). That is, women did not react differently to children's cues than men; they simply responded faster. Thus, although women have been shown to be more sensitive to infants' cues of immaturity and need than men, this difference seems not to be as robust when evaluating preschool and school-age children, at least when using the current methodology.

Although the main findings of this study are seemingly clear and compelling, there are at least three potential limitations to this research that need to be further considered. First, the voices and faces used in this investigation did not actually belong to the same children (i.e., they were obtained from different samples of children). Although they were matched thoroughly on the basis of their degree of maturity/immaturity, there is always a chance that participants would be sensitive to the *real* mismatching of voices and faces in the same way that babies have been shown to be highly sensitive to mismatches in intermodal perception, particularly visual identification of speech (see Bahrick & Hollich, 2020, for a review). Second, the faces and voices of children were hypothetical, manipulated, and therefore not real. In such a scenario, one can never be entirely sure to what extent the methodological advantages attained in terms of the internal validity of our study, accomplished through stimulus manipulations, might have potentially affected participants' reactions to children in terms of external validity.[1] In this vein, it is worth noting that in our current study faces were presented in a static manner (photos), as opposed to a dynamic manner (videos), and there is some evidence indicating that this difference might affect adults' impressions about others (e.g., on attractiveness, see Lander, 2008; but see Smith et al., 2016, for different evidence). Likewise, caregivers seem to be especially responsive to infants' moving communications (i.e., behaviors associated temporally with locomotion) (Toyama, 2018; West & Iverson, 2021). Finally, manipulations of children's vocal and facial features are just one of many possible alternatives. For example, our children's voice simulations were based chiefly on duration and pitch, as expressed by fundamental frequency, one of the most informative features of voice, particularly regarding age (Cook, 2002; Titze, 2000). However, as has been indicated in other research (Hernández Blasi et al., 2022), voice information is multidimensional, and other levels of information (e.g., lexical, grammatical, prosody, fluency, timber) likely can be used to simulate the maturity of children's voices, perhaps affecting adults' perceptions in a different way than the one found here. Something similar could be said about our simulations of children's faces.

In sum, this study reveals that overall voices, regardless of the speech content, are more important than faces in determining adults' reactions to young children, with women reacting faster than men. Language is certainly one of the most distinctive attributes of *Homo sapiens* (Laland & Seed, 2021), but our ancestors were already able to process some vocal information long before speech evolved (Belin et al., 2011). Therefore, perhaps it should not be surprising that humans are so proficient at processing and pulling out relevant information from others' voices, including those of young children. Our research shows that in some conditions vocal cues are more informative than facial cues. Indeed,

[1] It should be noted that in Hernández Blasi et al. (2022), results obtained for the Natural-Voices condition and Simulated-Voices condition were fully comparable.

speech is better understood as an integration of vocal *and* facial information, where each type of information can facilitate or interfere with the other, depending on their degree of congruence or incongruence (Belin et al., 2011; see also McGurk & MacDonald, 1976, on the McGurk effect). To date, available research indicates that young children's vocal *and* cognitive cues are highly informative for adults about children's attributes when considered alone (Bjorklund et al., 2010; Hernández Blasi et al., 2022). Previous research and the current study also revealed that both vocal and cognitive cues dominate over facial cues when considered together (Hernández Blasi et al., 2015; this study). However, it is still unknown how young children's vocal and cognitive cues contribute together to convey some relevant information to adults as well as which one, if either, would prevail when providing contradictory information. This is likely one of the research challenges within the field whose undertaking would be worth the effort in the near future.

More than four decades ago, Trivers (1974) wrote that infants and young children have evolved "psychological weapons" to increase the likelihood that adults, specifically their parents, will care for them, increasing children's chances of survival. One of those "weapons" during infancy is the *Kindchenschema,* with immature physical features promoting caregiving (e.g., Glocker et al., 2009; Lorenz, 1943). The adaptive benefits of immature facial features decline during early childhood (e.g., Luo et al., 2011, 2020), although they do not totally disappear, with immature childhood facial features being associated with higher ratings of positive affect (e.g., Hernández Blasi & Bjorklund, 2018). However, with the advent of early childhood, children's language, as reflected by maturity of the voice (e.g., Hernández Blasi et al., 2022; this study), and expressions of certain types of immature cognition (e.g., Bjorklund et al., 2010; Periss et al., 2012) become salient features promoting caregiving attitudes in adults. From this perspective, children at different levels of ontogeny possess different attributes that natural selection has likely used to evolve into "psychological weapons" to promote caregiving in adults and thus enhance children's chances of both survival and social-learning opportunities in the world that surrounds them.

Acknowledgments

This work was supported by a research grant from the Ministerio de Economía y Competitividad (PSI2015-70071-P, MINECO/FEDER), Spain. We thank Anna Borillo, Antonio Nieto, and the teachers, parents, and children of the *Grans i Menuts* school for their assistance in recording the samples of children's voices and faces.

References

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist, 55*(5), 469–480. https://doi.org/10.1037/0003-066X.55.5.469.

Bahrick, L. E., & Hollich, G. J. (2020). Intermodal perception. In J. B. Benson (Ed.), *Encyclopedia of infant and early childhood development (2nd ed.,* pp. 202–217). Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.23594-3.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development, 23*(2), 223–229. https://doi.org/10.1016/S0163-6383(01)00037-6.

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x.

Berger, T., Peschel, T., Vogel, M., Pietzner, D., Poulain, T., Jurkutat, A., Meuret, S., Engel, C., Kiess, W., & Fuchs, M. (2019). Speaking voice in children and adolescents: Normative data and associations with BMI, Tanner stage, and singing activity. *Journal of Voice, 33*, 580.e21–580.e30. https://doi.org/10.1016/j.jvoice.2018.01.006.

Bettle, R., & Rosati, A. G. (2021). The evolutionary origins of natural pedagogy: Rhesus monkeys show sustained attention following nonsocial cues versus social communicative signals. *Developmental Science, 24*(1). https://doi.org/10.1111/desc.12987. Article e12987.

Bjorklund, D. F., Hernández Blasi, C., & Periss, V. A. (2010). Lorenz revisited: The adaptive nature of cognitive immaturity. *Human Nature, 21*(4), 371–392. https://doi.org/10.1007/s12110-010-9099-8.

Bustos, I. (Ed.) (2012). *La voz: La técnica y la expresión* [Voice: Technique and expression] (2nd ed.). Paidotribo.

Charlesworth, T. E. S., Hudson, S.-K., Cogsdill, E. J., Spelke, E. S., & Banaji, M. R. (2019). Children use targets' facial appearance to guide and predict social behavior. *Developmental Psychology, 55*(7), 1400–1413. https://doi.org/10.1037/dev00 00734.

Cook, N. D. (2002). *Tone of voice and mind: The connections between intonation, emotion, cognition and consciousness*. John Benjamins.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584–1595. https://doi.org/10.2307/1130766.

DeCasper, A., & Fifer, W. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928.

Eggleston, A., Flavell, J. C., Tipper, S. P., Cook, R., & Over, H. (2021). Culturally learned first impressions occur rapidly and automatically and emerge early in development. *Developmental Science, 24*(2). https://doi.org/10.1111/desc.13021. Article e13021.

Ewing, L., Sutherland, C. A. M., & Willis, M. L. (2019). Children show adultlike facial appearance biases when trusting others. *Developmental Psychology, 55*(8), 1694–1701. https://doi.org/10.1037/dev00 00747.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9602–9605. https://doi.org/10.1073/pnas.152159999.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Franklin, P., & Volk, A. (2018). A review of infants' and children's facial cues' influence on adults' perceptions and behaviors. *Evolutionary Behavioral Sciences, 12*(4), 296–321. https://doi.org/10.1037/ebs0000108.

Furlow, F. B. (1997). Human neonatal cry quality as an honest signal of fitness. *Evolution and Human Behavior, 18*(3), 175–193. https://doi.org/10.1016/S1090-5138(97)00006-8.

Glocker, M. L., Langleben, D. D., Ruparel, K., Loughead, J. W., Gur, R. C., & Sachser, N. (2009). Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethology, 115*(3), 257–263. https://doi.org/10.1111/j.1439-0310.2008.01603.x.

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*(4), 544–549. https://doi.org/10.1542/peds.56.4.544.

Hernández Blasi, C. (2020). Evolutionary developmental psychology. In T. K. Shackelford (Ed.), *The SAGE handbook of evolutionary psychology: Integration of evolutionary psychology with other disciplines* (pp. 51–72). Sage. https://doi.org/10.4135/9781529739435.n3.

Hernández Blasi, C., & Bjorklund, D. F. (2018). Adolescents' sensitivity to children's supernatural thinking: A preparation for parenthood? *Psicothema, 30*(2), 201–206. https://doi.org/10.7334/psicothema2017.193.

Hernández Blasi, C., Bjorklund, D. F., Agut, S., Lozano Nomdedeu, F., & Martínez, M. Á. (2022). Voices as cues to children's needs for caregiving. *Human Nature, 33*, 22–42. https://doi.org/10.1007/s12110-021-09418-4.

Hernández Blasi, C., Bjorklund, D. F., & Ruiz Soler, M. (2015). Cognitive cues are more compelling than facial cues in determining adults' reactions towards young children. *Evolutionary Psychology, 13*(2), 511–530. https://doi.org/10.1177/147470491501300212.

Hernández Blasi, C., Bjorklund, D. F., & Ruiz Soler, M. (2017). Children's supernatural thinking as a signalling behaviour in early childhood. *British Journal of Psychology, 108*(3), 467–485. https://doi.org/10.1111/bjop.12211.

Hrdy, S. B. (1999). *Mother Nature: A history of mothers, infants, and natural selection*. Pantheon.

Instituto Nacional de Estadística. (2022). *España en cifras 2022* [Spain in figures 2022]. INE. http://www.ine.es/prodyser/espa_cifras.

Konner, M. (2010). *The evolution of childhood: Relationships, emotion, mind*. Harvard University Press.

Laland, K., & Seed, A. (2021). Understanding human cognitive uniqueness. *Annual Review of Psychology, 72*, 689–716. https://doi.org/10.1146/annurev-psych-062220-051256.

Lander, K. (2008). Relating visual and vocal attractiveness for moving and static faces. *Animal Behaviour, 75*(3), 817–822. https://doi.org/10.1016/j.anbehav.2007.07.001.

Lorenz, K. Z. (1943). Die angeboren Formen moglicher Erfahrung [The innate forms of possible experience]. *Zeitschrift fur Tierpsychologie, 5*(2), 233–409. https://doi.org/10.1111/j.1439-0310.1943.tb00655.x.

Lummaa, V., Vuorisalo, T., Barr, R. G., & Lehtonen, L. (1998). Why cry? Adaptive significance of intensive crying in human infants. *Evolution and Human Behavior, 19*(3), 193–202. https://doi.org/10.1016/S1090-5138(98)00014-2.

Luo, L. Z., Li, H., & Lee, K. (2011). Are children's faces really more appealing than those of adults? Testing the baby schema hypothesis beyond infancy. *Journal of Experimental Child Psychology, 110*(1), 115–124. https://doi.org/10.1016/j.jecp.2011.04.002.

Luo, L., Zhang, Q., Wang, J., Lin, Q., Zhao, B., Xu, M., Langley, C., Li, H., & Gao, S. (2020). The baby schema effect in adolescence and its difference from that in adulthood. *Journal of Experimental Child Psychology, 198*. https://doi.org/10.1016/j.jecp.2020.104908. Article 104908.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. https://doi.org/10.1038/264746a0.

Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J., & Shozo, K. (2012). Cerebral responses to infant-directed speech and the effect of talker familiarity. *NeuroImage, 59*(2), 1735–1744. https://doi.org/10.1016/j.neuroimage.2011.07.093.

Nip, I. S., & Green, J. R. (2013). Increases in cognitive and linguistic processing primarily account for increases in speaking rate with age. *Child Development, 84*(4), 1324–1337. https://doi.org/10.1111/cdev.12052.

Orena, A. J., & Werker, J. F. (2021). Infants' mapping of new faces to new voices. *Child Development, 92*(5), e1048–e1060. https://doi.org/10.1111/cdev.13616.

Periss, V., Hernández Blasi, C., & Bjorklund, D. F. (2012). Cognitive "babyness": Developmental differences in the power of young children's supernatural thinking to influence positive and negative affect. *Developmental Psychology, 48*(5), 1203–1214. https://doi.org/10.1037/a0026979.

Schaller, M. (2008). Evolutionary bases of first impressions. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 15–34). Guilford.

Senese, V. P., De Falco, S., Bornstein, M. H., Caria, A., Buffolino, S., & Venutti, P. (2013). Human infant faces provoke implicit positive affective responses in parents and non-parents alike. *PLoS One1, 8*(11). https://doi.org/10.1371/journal.pone.0080379. Article e80379.

Smith, H. M., Dunn, A. K., Baguley, T., & Stacey, P. C. (2016). Concordant cues in faces and voices: Testing the backup signal hypothesis. *Evolutionary Psychology, 14*(1), 1–10. https://doi.org/10.1177/1474704916630317.

Soltis, J. (2004). The signal functions of early infant crying. *Behavioral and Brain Sciences, 27*(4), 443–458. https://doi.org/10.1017/S0140525X0400010X.

Sprengelmeyer, R., Perrett, D. I., Fagan, E. C., Cornwell, R. E., Lobmaier, J. S., Sprengelmeyer, A., Aasheim, H. B. M., Black, I. M., Cameron, L. M., Crow, S., Milne, N., Rhodes, E. C., & Young, A. W. (2009). The cutest little baby face: A hormonal link to sensitivity to cuteness in infant faces. *Psychological Science, 20*(2), 149–154. https://doi.org/10.1111/j.1467-9280.2009.02272.x.

Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin, 44*(4), 521–537. https://doi.org/10.1177/01461 67217 744194.

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2021). Infants' preference for social interactions increases from 7 to 13 months of age. *Child Development, 92*(6), 2577–2594. https://doi.org/10.1111/cdev.13636.

Titze, I. R. (2000). *Principles of voice production (2nd printing)*. National Center for Voice and Speech.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*, 519–545. https://doi.org/10.1146/annurev-psych-113011-143831.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813.

Toyama, N. (2018). Social exchanges with objects across the transition from crawling to walking. *Early Child Development and Care, 190*(7), 1031–1041. https://doi.org/10.1080/03004430.2018.1511550.

Trivers, R. L. (1974). Parent–offspring conflict. *American Zoologist, 14*(1), 249–264. https://doi.org/10.1093/icb/14.1.249.

Trollinger, V. L. (2003). Relationships between pitch-matching accuracy, speech fundamental frequency, speech range, age, and gender in American English-speaking preschool children. *Journal of Research in Music Education, 51*, 78–94. https://doi.org/10.2307/3345650.

West, K. L., & Iverson, J. M. (2021). Communication changes when infants begin to walk. *Developmental Science, 24*(5). https://doi.org/10.1111/desc.13102. Article 13102.

Wolff, P. H. (1969). The natural history of crying and other vocalizations in early infancy. In B. M. Foss (Ed.). *Determinants of infant behavior* (Vol. 4, pp. 81–111). Methuen.

Zebrowitz, L. A., & Zhang, Y. (2011). The origins of first impressions in animal and infant face perception. In J. Decety & J. T. Cacioppo (Eds.), *The Oxford handbook of social neuroscience* (pp. 434–444). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195342161.013.0029.