2023-09-12

Children’s Sensitivity to Emotional Prosody in an Unfamiliar Language: The Role of Referential Context

Waly, Yomna Khaled

UNIVERSITY OF CALGARY

Children's Sensitivity to Emotional Prosody in an Unfamiliar Language: The Role of Referential Context

by

Yomna Khaled Waly

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN CLINICAL PSYCHOLOGY

CALGARY, ALBERTA

SEPTEMBER, 2023

Abstract

Research suggests that preschoolers use emotional prosody to make sophisticated judgements, such as using a speaker's tone of voice (happy vs. sad) to identify an intended referent (e.g., a broken vs. intact toy). However, 4-year-olds only make these associations in familiar, but not unfamiliar languages. Here, we sought to understand these discrepant findings by investigating the role of referential cues (i.e., cues that signal the speaker's intended object) in children's use of emotional prosody to resolve referential ambiguity. Results indicated that 4-year-old children did not use emotional prosody to modulate their looking nor pointing to the target object. Instead, they consistently fixated to the negative object whilst pointing to the positive object. Conversely, 8-year-olds looked at and point to the target object during sad-sounding trials. Their looking and pointing to the negative object was attenuated during happy-sounding trials such that they looked and pointed equally to both objects. Taken together, findings suggest that even in the presence of heightened referential cues, use of emotional prosody in an unfamiliar language remains an emergent skill during childhood.

Preface

This thesis is an original, independent, unpublished work by Yomna Waly. It was approved by the University of Calgary's Conjoint Faculties Research Ethics Board for the project entitled "Preschoolers' Attention to Vocal Emotional Cues in Unfamiliar Languages" on October 27th, 2022.

Acknowledgements

I would like to express my deepest gratitude to my supervisor and mentor Dr. Susan Graham. Your guidance and dedicated support have been invaluable to my academic journey thus far. This project would not have been possible without your insights, suggestions, and our many meetings and drafts. It is an extraordinary privilege to be one of your graduate students. I also cannot thank Dr. Craig Chambers enough for the valuable contributions to this project. Your suggestions and assistance have been vital to this project. I would like to thank the members of my supervisory committee, Dr. Kathleen Hughes and Dr. Julia Kam, for their time and guidance at various stages of the project.

I extend my thanks to the past and present members of the Language and Cognitive Development Lab for your incredible kindness and day-to-day support. To the many families who participated in this research, thank you. This research would not have been possible without your time and effort. This research was generously funded by a SSHRC Insight Grant awarded to Dr. Susan Graham and Dr. Craig Chambers.

Finally, my sincerest thanks to my parents and brother for their love, patience, and unconditional support throughout this journey. I am grateful everyday for your many sacrifices, large and small, to make this endeavor a reality. I also thank my fiancé for the unwavering encouragement – I am incredibly appreciative of your constant motivation during the final stages.

Table of Contents

Children's Sensitivity to Emotional Prosody in an Unfamiliar Language: The Role of Referential Context

Spoken communication can convey information beyond that which is expressed through words alone. One critical source of information is a speaker's emotional tone of voice. Indeed, the intended meaning of semantically identical sentences can vary considerably depending on the emotional tone the speaker uses. For example, the utterance "I received my exam results" spoken in a happy or sad tone could reveal the speaker's satisfaction or disappointment with their results. Given that emotions in spoken language can provide key insights into a speaker's meaning, quick and accurate detection of emotional signals is paramount to effective communication. Detection of emotions in speech is not limited to one's native language. Indeed, sensitivity to vocal emotional cues has been observed across languages as early as the preschool years (Ma et al., 2022; Waly et al., under review). The present study investigated 4- to 8-year-old children's sensitivity to emotions in speech when presented in an unfamiliar language to infer a speaker's referential intent.

Emotional signals in speech are collectively referred to as emotional prosody and are conveyed through variations in pitch levels, pitch contours, loudness, and rate of speech (Frick, 1985). Sensitivity to emotional prosody emerges early in infancy and precedes productive language. Kao et al. (2022) demonstrated that infants between 3- and 12-months of age discriminate between different emotional tones of voice and show a preference for happy-sounding prosody. By 7-months, evidence from event-related potential (ERP) data suggested that infants can accurately detect and implicitly associate emotional prosody with facial emotions (Grossman, 2010; Grossman et al., 2006). Beyond recognition, infants use emotional prosody to guide their behaviour (see Walle et al., 2017; Wu et al., 2021 for a review). For example, infants

were more likely to interact positively with a stranger when their parents spoke to the stranger using positive affect when compared to negative affect (Boccia & Campos, 1989). Further, infants use emotional prosody to modulate their interaction with new objects such that they interacted with objects less in response to a fearful tone of voice than a happy tone of voice (Mumme & Fernald, 2003).

Sensitivity to emotional prosody continues to develop through the preschool years. Research suggests preschoolers' detection and use of emotional prosody varies as a function of the information presented in a task (e.g., Aguert et al., 2013). When incongruent lexical and prosodic cues are present, preschoolers tend to rely on semantic information to make inferences. In a seminal study, Morton and Trehub (2001) presented children aged 4 to 10 years with utterances with conflicting linguistic and prosodic cues (e.g., "My dog ran away from home" spoken in a happy tone of voice). The youngest participants primarily used semantic cues to judge the speaker's emotional state. Despite not making explicit associations, preschoolers showed evidence of implicitly detecting emotional prosody as they displayed longer response times when linguistic cues were incongruent than when they were congruent (Morton & Trehub, 2001). When semantic and prosodic information were not in conflict, 3-year-olds were shown to implicitly, through eye-gaze, associate emotional prosody with corresponding emotional faces, while 5-year-olds could do so both implicitly and explicitly, through pointing (Berman et al., 2016).

The ability to detect and use emotional prosody continues to develop during childhood, with some studies suggesting that the detection and use of emotional prosody does not emerge until later childhood. Aguert et al. (2013) examined preschoolers' and children's ability to identify a speaker's emotional state through emotional pseudoutterances (i.e., nonsense syllables

that carry no lexical content). Their findings demonstrated that 5-year-olds did not use emotional prosody to infer emotional states, whereas 9- and 13-year-old participants did. The use of pseudoutterances may have posed an additional challenge for the youngest children. A similar pattern of development was observed by Gil et al. (2016). Five-, seven-, and nine-year-old children were asked to match vocal emotions presented through pseudoutterances (e.g., Hat sundig pron you venzy) to a continuum of facial expressions. Only 9-year-old participants accurately matched vocal and facial expressions of emotion. Another cross-modal matching study by Palama et al. (2022) corroborated these findings using an eye-tracking paradigm. When asked to match affective burst stimuli (i.e., an emotional vocal emission of the vowel /a/) to happy- or angry-looking faces, 5-year-old children did not fixate on the target faces during happy or angry trials. By 8 years of age, children showed a preference for the happy face during happy-sounding trials. Preferential looking towards the angry face did not emerge until 10 years of age. These findings indicate that sensitivity to emotional prosody remains emergent throughout early childhood.

Beyond identification of speakers' emotions, research has demonstrated that preschoolers use emotional prosody to make sophisticated judgements. For example, children use prosodic cues to resolve referential ambiguity – instances where a speaker's words alone do not identify their intended referent or outcome (e.g., Berman et al., 2013a, Berman et al., 2013b; San Juan et al., 2017; Khu et al., 2018). For example, Berman et al. conducted a series of studies in which they examined preschoolers' use of emotional prosody to identify the speaker's intended referent. They presented 3-, 4-, and 5-year-olds with semantically ambiguous utterances (e.g., "Look, look at the duck. Point to the duck.") spoken in a happy, sad, or neutral tone of voice (Berman et al., 2010; Berman et al., 2013a). These utterances were accompanied by a visual

array consisting of two related objects and one unrelated object. The related objects belonged to the same category (e.g., balls) but varied in their association with happiness or sadness (e.g., inflated vs. deflated, respectively). Findings indicated that 3-year-olds did not use emotional prosody to identify the speaker's intended referent, whereas 5-year-olds do so both explicitly (through pointing) and implicitly (through eye gaze). Four-year-old children only made implicit associations by looking at the correct object. Specifically, 4-year-olds' eye gaze shifted during the noun region of the utterance such that they were more likely to fixate on the object that matched the emotional prosody in both happy- and sad-sounding trials.

In light of previous research showing that children can detect prosody early in utterances when matching emotional prosody to faces, the emergence of preferential looking towards the target object after the onset of the noun may point to an integrative challenge. Past findings demonstrated that preschoolers make quicker gaze-shifts to target faces (Berman et al., 2013b) than target objects (Berman et al., 2010) when listening to happy- and sad-sounding prosody. This discrepancy suggests that referential inferencing poses an integrative challenge – it may be difficult for children to simultaneously integrate the semantic content of speech relating to the object with emotional prosody during the earliest moments of speech processing. Thus, the simultaneous presence of emotional prosody and semantic cues referencing the object may present an additional cognitive demand, making the referential-identification task more difficult.

One approach to address this possibility is to remove the need to integrate both the emotional prosody and the linguistic information by presenting utterances in an unfamiliar language. A number of studies have demonstrated that children as young as 4-years-old can draw upon emotional prosody to identify a speaker's emotional state when presented with utterances produced in an unfamiliar language (Morton & Trehub, 2001; Chronaki et al., 2018). For

example, Ma et al. (2022) asked participants to match the emotional utterances with corresponding facial expressions. Findings indicated that preschoolers could identify the emotional content of speech using prosodic cues across languages, with no significant differences between native and unfamiliar languages. Drawing upon this body of research, Waly et al. (under review) used the paradigm described above to examine preschoolers' ability to identify a speaker's intended referent when the emotional utterance was presented in an unfamiliar language. English-speaking 4-year-olds were presented with happy-, sad-, and neutral-sounding utterances recorded in Arabic, which, although phonologically different, carries acoustic and prosodic properties that are recognizable by English-speakers. Children were simultaneously shown a visual array consisting of two juxtaposed versions of an object (e.g., clean vs. dirty duck). Results indicated that 4-year-olds did not draw upon the speaker's emotional prosody to identity the intended referent. Further, the use of emotional prosody to modulate looking to the target object during the noun region as reported by Berman et al., 2010 was not observed when preschoolers heard emotional unfamiliar language utterances. Instead, they continued to look to the negative objects across the entire utterance during both happy and sad trials. In contrast, 4-year-olds successfully matched the exact same Arabic utterances to emotional faces. This pattern suggests that preschoolers can indeed extract emotional prosody from Arabic utterances but do not extend this sensitivity when required to use emotional prosody to identify a speaker's intended referent.

Why do children display such variability on tasks requiring the use of emotional prosody to resolve referential ambiguity? As noted above, children's difficulty on the unfamiliar language task compared to the native language task suggests that the presence of an integrative challenge cannot fully explain differences in performance. Further, children's ability to match happy- and

sad-sounding Arabic utterances to faces but not objects indicated that children are indeed sensitive to prosodic cues in unfamiliar languages yet cannot extend that understanding to identify intended referents. One possible explanation for these discrepant findings is that children did not understand the referential nature of the task when presented with utterances in an unfamiliar language. That is, in the absence of semantic cues in the utterances, the association between the emotional prosody and objects on the screen was not evident.

The present study focused on understanding differences in children's performance across familiar and unfamiliar language tasks through exploring the role of referential cues. It is possible that discrepant findings across English (Berman et al. 2010) and Arabic (Waly et al., under review) iterations of the referential-identification task are driven by the absence of salient referential cues, as unfamiliar language utterances do not explicitly link the emotional prosody and objects through a recognizable noun. The present study addresses the proposal that emphasizing the referential nature of the task contextualizes the use of emotional prosody and highlights the link between the emotional prosody and the objects.

The role of referential cues – signals highlighting the object being referred to – in identifying a speaker's intent is well documented, particularly in in word learning contexts. For instance, although infants do not typically map phototactically illegal words onto objects, this tendency can be overridden when the referential nature of the task is highlighted (e.g., MacKenzie et al., 2014, Namy, 2001). Similarly, the presence of the speaker also contributes to referential understanding. Infants make object–label associations only when taught the object name by a speaker in front of them and not when the object was named by a speaker in an adjacent room (Hung et al., 2015). Moses et al. (2001) also found that children altered their reaction to a novel toy in response to a speaker's affect only when the speaker was in the room

and looking at the object (affect-relevant), but not when the speaker was seated in a different room with no visual access to the toy (affect-irrelevant).

Drawing upon this literature, the present study heightened referential cues related to speaker through the introduction of a speaker who was identified as the speaker who was searching for objects. The speaker was introduced at the outset of the study, and again at the mid-point by showing a photograph of a smiling, waving woman named Sarah. Due to methodological constraints relating to the consistent production of identical emotional utterances, the speaker was not physically present in the room. Instead, children heard the speaker's voice with whom they believed they were playing a game.

The objectives of the present study are as follows: First, to investigate role of referential cues in children's use of emotional prosody presented in an unfamiliar language to resolve referential ambiguity; and second, to examine the development of this ability between 4- and 8-years of age. Four- and eight-year-olds were presented with happy-, sad-, and neutral-sounding Arabic utterances along side a visual array consisting of two juxtaposed objects (e.g., an inflated vs deflated ball) that served as referents in following with Berman et al. (2010) and Waly et al. (under review). Children were instructed to listen to the utterances and look and point at the image that matches the speaker's tone of voice. Their moment-by-moment sensitivity to prosodic cues was assessed using an eye-tracking paradigm. As noted previously, referential cues were highlighted by introducing children to the speaker, Sarah, at the outset, and again at the midpoint of the study. The introduction of the speaker served to highlight the referential nature of the task by explicitly stating that the emotional utterances were spoken by "Sarah" who is searching for her objects. Following each instance of speaker introduction, children heard a series of neutral-sounding cognate trials wherein the noun sounded similar across both English and Arabic. The

use of cognates served to reinforce that utterances do indeed impart useful information regarding the objects viewed on the screen and discourage children from disregarding auditory stimuli.

This study addresses the gap in current understanding of the discrepancies in children's performance on referential identification tasks in familiar (Berman et al., 2010) and unfamiliar (Waly et al., under review) language conditions, by examining the role of referential context. Increasing referential context served to emphasize the goal of the task, that is, to identify the speaker's intended referent. If children are successful at identifying the intended object upon amplification of referential cues, this would suggest that observed differences between native and unfamiliar languages can be ascribed to children not attending to the referential nature of the task. If children are not successful, this may indicate that differences across language conditions are driven by developmental differences in sensitivity to familiar and unfamiliar languages. This will be most evident if 8-year-olds demonstrate success on the task.

It is expected that upon enhancing referential cues, 4-year-olds would match their performance in native language studies (Berman et al., 2010) by fixating on, but not pointing to, objects matching the emotional prosody of utterances at above chance levels. Conversely, 8-year-olds were expected to successfully fixate on and point to the correct objects.

Method

Participants

Participants were recruited from the ChILD research group database, a registry of parent and child contact information obtained through voluntary parental sign-ups. The sample consisted of 73 primarily English-speaking children, including 37 children between 4 years 0 months to 4 years 6 months (18 females; $M = 4.25$, $SD = 0.13$) and 36 children between 8 years 0 months and 8 years 6 months (18 females; $M = 8.28$, $SD = 0.13$). Seven additional 4-year-olds

and two additional 8-year-olds were tested but were not included in the final sample. Exclusionary criteria included any exposure to the unfamiliar language, less than 80% exposure to English (4-year-olds: $n = 4$, 8-year-olds: $n = 2$), not successfully completing the calibration task (4-year-olds: $n = 2$), and not completing the study (4-year-olds: $n = 1$). Parents most frequently reported undergraduate level of highest education (48%), followed by trade school (26%), graduate school (16%), and high school (10%). Parents reported 8-year-old's ethnic background with White most frequently reported (92%). Ethnic background information was not collected for 4-year-old participants. Children received a toy, t-shirt, and certificate for their participation. This study was received approval from the Conjoint Faculties Research Ethics Board (REB 19-1701; Appendix A).

Stimuli

Auditory Stimuli

Auditory stimuli were recorded by a female native speaker of Arabic. They consisted of twelve utterances that all followed the pattern of "Look, look at the [object]. Point to the [object]". They were matched in length and volume. The emotional prosody of utterances was counterbalanced such that participants could be presented with the happy- or sad-sounding version of any given utterance. Each participant heard six happy-sounding and six sad-sounding utterances, presented in a randomized order. A sample of the utterances and corresponding visual array can be seen in Figure 1. A full list of object pairs is available in Appendix B.

An additional six utterances were recorded by the same speaker using neutral prosody and served as cognate trials (See Appendix B). These utterances followed the same pattern of "Look, look at the [object]. Point to the [object]", wherein the objects names were cognates – words that have similar meaning and pronunciation across languages. Cognates were recorded in

a manner that corresponded to the Arabic variation of pronunciation. For example, chocolate was recorded in Arabic as [aʃːokalata].

Visual Stimuli

The speaker was introduced by showing a photograph of a young woman smiling and waving. Following the introduction of the speaker, children were shown a series of critical and filler trials. Critical trials were comprised of visual arrays consisting of two panels depicting objects belonging to the same category but varying in their association with happiness and sadness. The objects were juxtaposed along the dimensions of clean vs. dirty, intact vs. broken, and inflated vs. deflated. The terms “negative object” and “positive object” will be used henceforth to refer to objects associated with the happy and sad prosody respectively. Filler trials also consisted of arrays of two photos of objects. The images, however, were of unrelated items (e.g., a computer and a stuffed dog). Table 1 displays a sample of the visual stimuli arrays and their corresponding utterances. The placement of the target object was counterbalanced for both the critical and filler trials.

Table 1

Sample Speaker Condition Visual Arrays and Associated Auditory Transcription

Visual stimuli	Utterance [in Arabic]	Trial condition
	*"Look at the doll. Point to the doll."* [ʊnzˤʊr ila ad:ʊmja. Aʃir ila ad:ʊmja]	Critical trial (Happy or sad emotional prosody)
	*"Look at the fork. Point to the fork."* [ʊnzˤʊr ila aʃ:awka. Aʃir ila aʃ:awka]	Critical trial (Happy or sad emotional prosody)
	*"Look at the computer. Point to the computer."* [ʊnzˤʊr ila alkʊmbu:tar. Aʃir ila alkʊmbu:tar]	Filler trial (Neutral emotional prosody)
	*"Look at the chocolate. Point to the chocolate."* [ʊnzˤʊr ila aʃ:okalata. Aʃir ila aʃ:okalata]	Filler trial (Neutral emotional prosody)

Apparatus

Children sat facing a 46-inch screen. Their eye gaze was tracked using a Tobii X3-120 eye-tracker placed below the monitor. The Tobii X3-120 allows for 36 degrees of freedom of head movement with an accuracy of 0.4 degrees. Eye gaze was recorded every 8 ms and a fixation was logged if the child looked at the same image for more than 95 ms. Regions of the screen that include each of the objects were pre-defined as areas of interest.

The various orders and trials were managed using E-Prime software and associated Tobii extensions. At the outset of every testing session a calibration task was conducted using the Tobii Studio software. Data were included in the analysis if participants had accurate calibration on at least 3 of the 5 test points. Participants' pointing was recorded using an HD camera placed directly in front of the child. The auditory stimuli were projected from a speaker placed next to the monitor.

Procedure

Parents completed informed consent and demographics forms before beginning the study (See Appendix C for 4-year-old demographics form and Appendix D for 8-year-old demographics form). Children were tested in a quiet lab space seated approximately 65 cm away from the eye-tracker and 140 cm away from the monitor. The calibration procedure was conducted first with a target of at least 3 out of 5 accurate test points. The experimenter then launched the study once a successful calibration was completed.

First, children were introduced to the speaker. They were shown a photo of a young woman waving while the experimenter introduced the speaker and the task. Participants were informed that the photo is of "Sarah" who speaks Arabic and is searching for her things. They were instructed to look at and point to the objects "Sarah" is searching for.

The speaker was introduced twice: once at the outset of the study and again at the mid-point. This was to remind participants of the purpose of the task. To further emphasize the referential nature of the task, following each introduction of the speaker, children were presented with three neutral cognate trials. These trials included object names that sound similar in English and Arabic, allowing children to identify the noun and associate it with the object on the screen, thus highlighting the association between the auditory and visual stimuli. No corrective feedback was provided.

The testing session lasted approximately 15 minutes. Parents watched the testing session from an adjacent room through a one-way mirror, such that they could see their child, but the child could not see them. Participants and parents were debriefed at the end of the study. Children received a toy, t-shirt, and personalized certificate for their participation.

Results

Eye-tracking Analyses

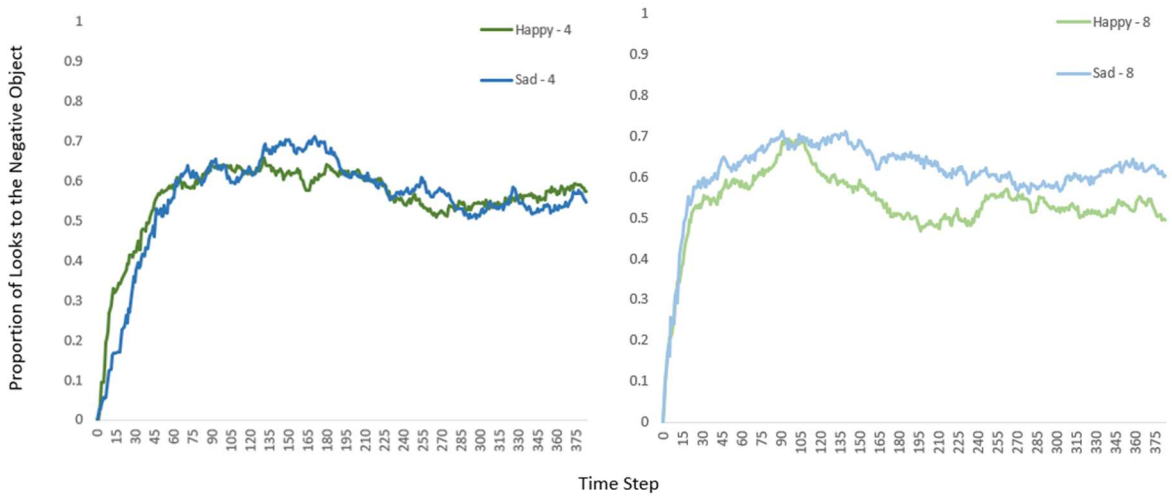
Participants' eye-gaze was tracked in real-time across the unfolding utterances (see Figure 2). Analyses were pre-registered and are available on https://osf.io/uayfv/?view_only=d252c5543c18450f932ab9d397953448. Any deviations from the pre-registered analysis plan are noted.

In keeping with prior research (e.g., Berman et al., 2010), only gaze patterns during the first portion of the utterance, "Look at the [object]", were included in the analysis. The analysis interval began at 200 ms and ended at 3403 ms. This interval was identified by taking the maximum of the first sentence (i.e., "Look at the [object]") across all utterances. The 200 ms offset to the starting time accounts for the typical lag in eye movement in reacting to linguistic information (e.g., Allopena et al., 1998). Only fixations initiated at or after 200 ms were included

in the analysis. This is a departure from the pre-registered analysis plan. Evidence suggests that the margin of 200 ms allows sufficient time for initiation of saccadic eye-gaze shifts in response to a stimulus (Hallett, 1986). The reliance on new fixations recorded after 200 ms of utterance onset was motivated by children making pre-utterance fixations and sustaining these fixations into the analysis window. These fixations are likely driven by visual interest in the images not by the emotional prosody of the utterances. As such, these fixations were excluded from the analysis and all subsequent analyses are exploratory in nature.

Figure 2

Time-course Plot of Children's Fixations to the Negative Object by Age Group and Prosody Type



Note. The analysis interval begins 200 ms after the onset of utterances and only includes looks initiated at or after 200 ms. The window was divided into time steps of approximately 8 second intervals for visualization.

Critical Trials. A linear mixed effects (LME) model was conducted using the statistical program R 4.2.2 (R Core Team, 2022) and the associated lme4 package (Bates et al., 2015). The

lmerTest package (Kuznetsova et al., 2016) was used to assess significance with Satterwaite's approximation for denominator degrees of freedom. The logit transformed proportion scores of looks to the negative object (e.g., deflated ball) served as the dependent variable in the model. Looks to the negative object were selected in keeping with prior research (e.g., Waly et al., under review; Berman et al., 2010). Past research demonstrated that children consistently fixate on the negative object, thus use of negative-object looks as the dependent variable helped establish deviations from baseline behaviour. Logit transformations were employed to allow for analysis of proportion scores using linear models. Fixations were assessed through gaze shifts towards the negative object during the sad-sounding trials, and away from the negative object during the happy-sounding trials. Proportion scores were calculated by summing the number of fixations to the negative object by across the time interval of the analysis. As such, a score of 1 would indicate that a given child fixated exclusively on negative objects across the interval, a score of 0 would indicate exclusive looking to positive objects, and a score of 0.5 would indicate equal fixations on negative and positive objects. The LME model included emotional prosody type (i.e., happy- vs. sad-sounding prosody) and age-group (4- vs. 8-year-olds) as fixed effects. Random effects consisted of a random intercept by participant, a random intercept by named-object, a by-named-object random slope for prosody type, a by-participant random slope for prosody type, and a by-named-object interaction of age-group and prosody type. The LME model converged but generated a singular fit warning. A significant interaction between age-group and prosody type was observed ($\beta = -0.02$, $SE = 0.011$, $t = 2.11$, $p = 0.039$).

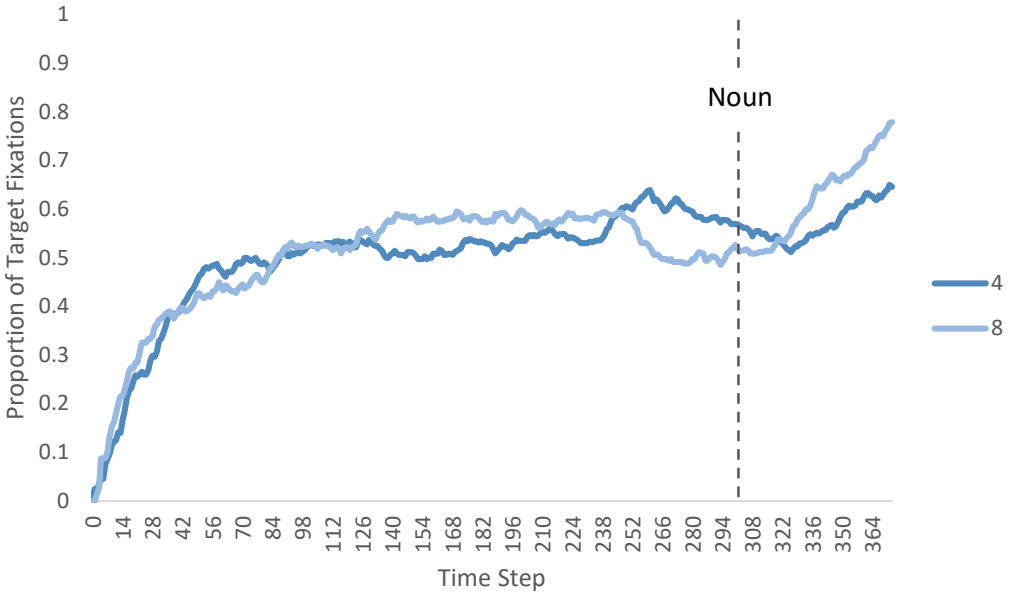
To parse the interaction term, follow-up LME models were conducted for the 4- and 8-year-old age groups separately. As no analysis plans were specified in the event of a significant interaction, the following analyses are not pre-registered and are exploratory. The dependent

variable in the models was the logit transformed proportion score of looking to the negative object. Prosody type served as the fixed effect. Random effects consisted of a random intercept by participant, a random intercept by named-object, and a by-participant random slope of prosody type. Both models converged but displayed singular fit. No significant main effect of prosody type was observed for 4-year-old participants ($\beta = 0.025$, $SE = 0.031$, $t = 0.82$, $p = 0.42$), suggesting their fixations to the negative object did not significantly differ across the happy- and sad-sounding trials. Examination of Figure 2 indicates that 4-year-olds looked more to the negative object regardless of prosody condition. Eight-year-old participants' gaze data, however, displayed a significant main effect of prosody type ($\beta = -0.071$, $SE = 0.034$, $t = 2.06$, $p = 0.047$). Inspection of Figure 2 suggests that 8-year-olds fixate on the negative object during sad-sounding trials and their looks to the negative object are attenuated during happy-sounding trials such that they are fixating equally on the negative and positive objects.

Cognate Trials. Data from the cognate trials was used to establish children's ability to attend to the utterances and extract salient linguistic information. A time-course plot (see Figure 3) was created to describe children's looking behaviour towards the target object over time. Children looked to each of the objects equally prior to hearing the noun. Their gaze shifted towards the negative object 200 ms following the onset of the noun as demarcated by the dashed line in Figure 3. The observed pattern of looking more to the target object after hearing the noun indicates that 4- and 8-year-old participants did indeed attend to the auditory cues and did not disregard the utterances.

Figure 3

Time-course Plot of Children's Fixations to the Target Object by Age Group and Prosody Type



Note. The window of analysis reflects looks initiated 200 ms after the onset of utterances. The window was divided into time steps for visualization.

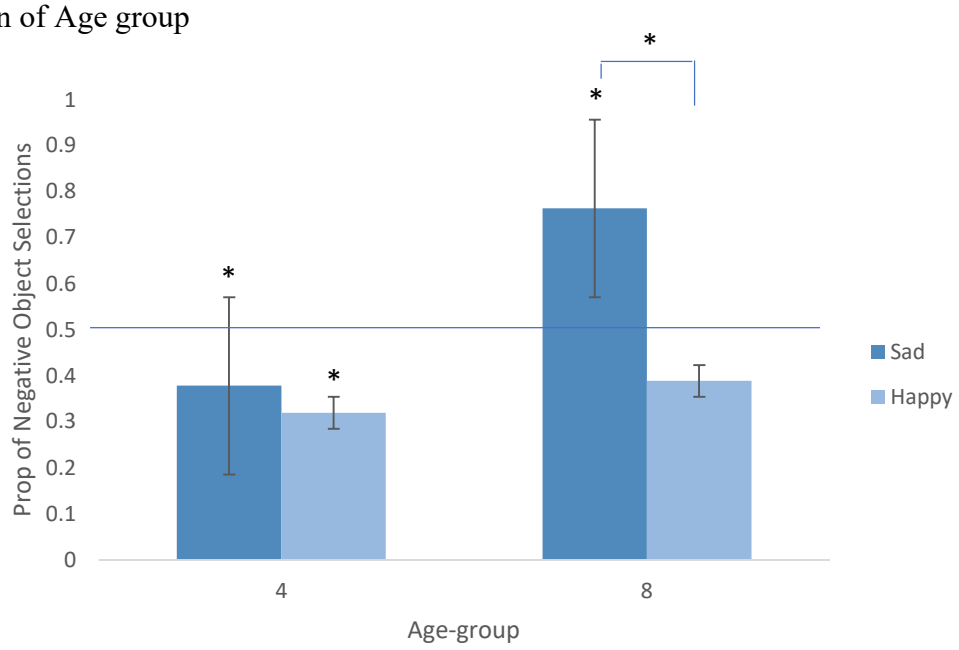
Pointing Analyses

Pointing data was coded by a trained research assistant. Twenty percent of the data was recoded by another research assistant to establish inter-rater reliability. Excellent inter-rater reliability was observed (Cohen's kappa = .88, $p < .001$). The proportion of children's points to the negative object (e.g., deflated ball) was calculated by dividing the number of negative object points by the total number of points. The following analyses were pre-registered prior to data collection and are available at https://osf.io/uayfv/?view_only=d252c5543c18450f932ab9d397953448. Deviations from the analysis plan are noted and pre-registered analyses are reported in Appendix E.

Critical Trials. A series of one-way t-tests were first conducted to determine if the proportion of negative selections differed from chance for each of the happy- and sad-sounding prosody trials (see Figure 4). Results indicated that 4-year-olds'selection of the negative object during happy-sounding trials ($M = 0.32$, $SD = 0.24$) was significantly below chance ($t(36) = 4.50$, $p < .001$). Four-year-olds similarly pointed to the negative object ($M = 0.38$, $SD = 0.27$) at below-chance levels ($t(36) = 2.73$, $p = .01$, $d = 0.45$) during sad-sounding trials, suggesting that they consistently pointed to the happy object across prosody types and did not use emotional prosody to modulate their pointing. Conversely, 8-year-old children made significantly more negative object selections ($M = 0.76$, $SD = 0.33$) during sad-sounding trials than chance ($t(35) = 4.81$, $p < .001$, $d = 0.80$). Eight-year-olds' negative object selections on the happy-sounding trials ($M = 0.39$, $SD = 0.38$) did not differ significantly from chance ($t(35) = 1.76$, $p = .086$, $d = 0.29$).

Figure 4

Proportions of Children's Negative Object Selections Across Happy- and Sad-sounding Trials as a Function of Age group



Next, a generalized linear mixed effects (GLME) model was conducted using R 4.2.2 (R Core Team, 2022) and lme4 package (Bates et al., 2015). Significance was assessed using the lmerTest package (Kuznetsova et al., 2016) with Satterwaite's approximation for denominator degrees of freedom. The dependent variable in the model was negative object selections, such that a score of 1 would indicate a negative object selection whereas a score of 0 would indicate a positive object selection. The model was specified to use a binomial distribution to fit binary data. Prosody type (i.e., happy- vs. sad-sounding trials) and age-group (i.e., 4- vs. 8-year-olds) served as fixed effects in the model. Random effects included a random intercept by participant, a random intercept by object named in the utterance (e.g., ball, truck), a by-named-object random slope for prosody type, a by-participant random slope for prosody type. A significant interaction between age-group and prosody type was observed ($\beta = -0.48$, $SE = 0.12$, $z = -4.01$, $p < .001$).

Follow-up analyses were conducted to further investigate the nature of the observed interaction. No plans for follow-up were pre-registered, and, as such, the following analyses are exploratory. Two additional models examined differences in sad object selections across 4- and 8-year-olds separately. The fixed effect in the models was prosody type. Random effects included a random intercept by participant, a random intercept by object named in the utterance, a by-participant random slope of prosody type, and a by named-object random slope for prosody type. Both models converged, with the 4-year-old model displaying singular fit. No significant main effect of prosody type was observed for 4-year-olds ($\beta = 0.33$, $SE = 0.22$, $t = 1.49$, $p = .14$), indicating that their pointing behaviour did not differ significantly between happy- and sounding-trials. A significant main effect of prosody type ($\beta = -2.93$, $SE = 0.69$, $z = -4.27$, $p < .001$) was observed for 8-year-old participants, with participants pointing more to the negative object on sad-sounding trials than on happy-sounding trials.

Cognate Trials. Cognate trials were used to evaluate children's attention to utterances and identify a salient, recognizable noun. Both 4-year-olds ($M = 0.76$, $SD = 0.21$; $t(36) = 7.57$, $p < .001$, $d = 1.24$) and 8-year-olds ($M = 0.91$, $SD = 0.13$; $t(35) = 18.14$, $p < .001$, $d = 3.02$) pointed to the correct object at above chance levels. This indicates that children did extract the cognate information from the utterances.

Discussion

The present study investigated 4- and 8-year-olds' use of emotional prosody presented in an unfamiliar language to resolve referential ambiguity when the referential nature of the task was emphasized. Recall that the referential nature of the task was highlighted through the introduction of a speaker at the outset of the study, and again at the mid-point of the study. In both instances, the introduction of the speaker was followed by cognate trials wherein the object labels named in the utterance sound similar across English and Arabic. Cognate trials served to further reinforce the referential nature of the task by highlighting the connection between auditory and visual cues. Results indicated that 4-year-old children did not use emotional prosody to modulate their looking nor pointing to gauge the speaker's referential intent. In contrast, 8-year-olds looked and pointed to the negative object during sad-sounding trials and their looks and points to the negative object were attenuated during happy-sounding trials. Each of these findings are discussed below.

Four-year-olds' eye-gaze demonstrated greater attention to the negative object regardless of emotional prosody type. That is, 4-year-old children consistently fixated on the negative object during both happy- and sad-sounding trials. This attention to negative objects is likely driven by novelty and visual interest as objects associated with sadness (e.g., broken truck, dirty duck) are more complex, thus more visually salient. Numerous past studies have observed this

same tendency to look at negative objects amongst preschoolers (e.g., Berman et al., 2013a, Berman et al., 2010). Recall that Berman et al. (2010) found that 4-year-olds consistently looked to the negative object when listening to happy- and sad-sounding English utterances prior to hearing the noun. Similarly, when English-speaking 4-year-olds were presented with happy- and sad-sounding Arabic utterances, they fixated to the negative object regardless of prosody condition. The current study extends past findings by demonstrating that preschoolers do not use emotional signals presented in an unfamiliar language to modulate their eye-gaze and overcome this attention to the negative object during referential-identification tasks, even when referential cues were highlighted.

Four-year-olds' implicit processing patterns, evident through looking patterns, were distinctly different from their explicit reasoning, demonstrated through pointing. Although 4-year-olds fixated on the negative objects, their pointing behaviour favoured positive objects at above chance levels across happy- and sad-sounding utterances. This tendency to point to positive objects was not observed in previous studies employing the same paradigm in familiar (Berman et al., 2010) or unfamiliar (Waly et al., under review) languages. Previous studies found that preschoolers point to both the positive and negative objects equally at chance levels regardless of prosody type (Waly et al., under review). The tendency to point to the positive object observed in the current study could be explained by the introduction of the speaker and the lack of attention to emotional prosody, such that children infer that the speaker must be seeking a positive (i.e., intact/clean) object. This pattern of responding suggests that 4-year-olds understand the referential nature of the task but do not use emotional prosody to inform their object selections.

The interplay between referential cues and language familiarity may further explain preschoolers' disregard of emotional prosody as a cue to resolve referential ambiguity. The importance of and reliance on referential cues appears to vary depending on children's familiarity with the languages with which they are presented. Hung et al. (2015) examined 3- and 4-year-olds' use of referential cues (i.e., a speaker's pointing and eye gaze) across language contexts. English-Mandarin bilingual preschoolers heard a story, read by a speaker, in one of English only (No Switch), English and Mandarin (Familiar-Switch), English and an unfamiliar language (Unfamiliar-Switch), or English and English-sounding pseudo-sentences (Nonsense-Switch). Children in the Unfamiliar-Switch condition made different inferences about the speaker's referential intent than their counterparts in the Familiar-Switch condition. That is, they were less likely to associate a novel label with a familiar object in the Unfamiliar-Switch condition than the Familiar-Switch condition. Further, at the time of the switch, preschoolers in the Unfamiliar-Switch condition gazed at the speaker longer than children in other conditions. This suggests that children's referential understanding is modulated by language familiarity and may be influenced by social cues.

Although 4-year-olds did not make correct implicit or explicit associations between emotional prosody and objects, results from cognate trials indicated that they were indeed attentive to the auditory stimuli. When the object named in the utterance was identifiable, preschoolers successfully identified the intended object through both looking and pointing. This indicates that, despite not using emotional prosody, 4-year-old children did not disregard the utterances, but were able to extract and use salient semantic information.

Altogether, contrary to hypotheses, 4-year-olds' performance on the referential-identification task did not improve when referential cues were amplified, when compared to the

results of previous research. This suggests that the absence of a salient noun (i.e., a noun recognizable by English-speaking children) in the emotional Arabic utterances may have contributed to 4-year-olds' inability to use emotional prosody to identify the intended referent. Specifically, the presence of a recognizable noun confers referential information that may have highlighted the goal of the task. Recall that when presented with a similar task in a familiar language, 4-year-olds correctly identified the target referent (Berman et al., 2010). In the case of Arabic, rather than allowing children to draw upon emotional prosody from the early points of the utterance, the use of the unfamiliar language may have obscured the goal of the task for English-speaking 4-year-olds.

By 8-years of age, children demonstrated greater success on the referential-identification task. They made more fixations to the negative object as sad-sounding trials unfolded. Further, their looking to the negative object was attenuated during happy-sounding trials such that they looked equally to both positive and negative objects when listening to happy prosody. The same pattern was evident in 8-year-olds' pointing behaviour wherein they correctly pointed to the negative object during sad-sounding trials but pointed to both positive and negative objects equally during happy-sounding trials. This suggests that the ability to modulate looking and pointing behaviour in response to emotional prosody presented in an unfamiliar language in a referential task remains an emergent skill at 8-years of age. The present findings also highlight differences in the detection and use of happy and sad vocal emotions.

The observed asymmetry in the sensitivity to happy- and sad-sounding emotional prosody is congruent with extant literature documenting a negativity bias in children's recognition of vocal emotions. Evidence suggests that children exhibit better performance with sad-sounding prosody than happy-sounding prosody (Berman et al., 2013a; Grossman et al.,

2005). For instance, in a vocal-facial emotion matching task, 5-year-olds only displayed sensitivity to a cross-modal effect during sad-sounding prosody trials (Gil et al., 2016). In another study, when asked to match emotional faces and voices, children demonstrated quicker processing when presented with sad-sounding utterances than happy-sounding ones (Berman et al., 2016). This trend towards greater sensitivity to negatively valenced emotions persists in unfamiliar languages. Chronaki et al. (2018) reported an overall advantage in children's recognition of negative emotions across four languages. Of relevance to the present study, English-speaking children correctly evaluated sad-sounding Arabic prosody 66.0% of the time, compared to only 20.0% success on happy-sounding trials. The increased sensitivity to sad prosody compared to happy prosody in Arabic persists amongst English-speaking adolescents (61.5% and 19.7%, respectively) and adults (84.7% and 35.7%, respectively; Chronaki et al., 2018). These findings highlight that accurate detection and use of happy-sounding prosody can pose a significant challenge, particularly in unfamiliar languages. This is consistent with present findings wherein 8-year-olds demonstrate success at identifying the correct referent on sad- but not happy-sounding trials.

Although 8-year-olds' displayed greater accuracy on the present referential-identification task compared to 4-year-olds, their performance does not match adult levels. Using the same auditory and visual stimuli employed in the current study, Waly et al. (under review) found that adults exhibit a high degree of accuracy across both happy- and sad-sounding trials in a referential identification task. Adults' near-ceiling level performance in that study reinforces the continued development of sensitivity to emotional prosody past the childhood years.

The present study has several limitations. First, it only used one unfamiliar language, Arabic, recorded by a female speaker. As such, findings cannot be generalized to speakers of all

genders and all unfamiliar languages. Second, the proposed study only examined happy- and sad-sounding emotional prosody rather than a full range of emotional expressions. Finally, conducting the study in a laboratory setting may limit its ecological validity. The pre-recorded stimuli and isolated still life images used in the study are not necessarily reflective of the complexity of the cues children encounter in naturalistic settings. Further, while the present study sought to examine the role of referential salience, the study design did not allow the speaker to be physically present in the room during the testing session, further limiting ecological validity.

Future research could extend the present findings through the pursuit of several avenues. One direction could investigate the amplification of referential cues within the utterance by incorporating cognates in critical trials such that the emotional Arabic utterances include a recognizable noun. Additionally, to increase the generalizability of findings, it would be important to replicate the present study using different unfamiliar languages to ensure that present findings are not attributable to Arabic's acoustic cues. Research on cross-linguistic emotion detection demonstrated that English-speaking children, adolescents, and adults recognize emotional prosody less accurately in Arabic than other unfamiliar languages, including Spanish and Chinese (Chronaki et al., 2018). Lastly, it is important to continue to examine the developmental trajectory of sensitivity to emotional prosody in unfamiliar languages across different age groups to better understand the development of sensitivity to vocal emotions between 8-years of age and adulthood.

The present study investigated the role of referential context in referential communication in an unfamiliar language across 4- and 8-years of age. Taken together, results indicated that sensitivity to emotional prosody presented in unfamiliar languages continues past the preschool

years and remains emergent by 8-years of age. Specifically, 4-year-olds were sensitive to referential context but did not use emotional prosody presented in Arabic to modulate their selections of the referent. By 8-years of age, children were accurate on sad trials whilst their looking and pointing behaviour on happy-sounding trials was attenuated by emotional prosody. Findings have shed light on the developmental trajectory of the use of emotional prosody presented in unfamiliar languages and have contributed to current understanding of the role of referential cues in the detection and use of emotional prosody to resolve referential ambiguity. This understanding will inform future research on emotional prosody as it will help establish the impact of referential salience in experimental settings, and, in turn, will help inform future paradigms for studying emotional prosody in preschool populations.

References

Aguert, M., Laval, V., Lacroix, A., Gil, S., & Le Bigot, L. (2013). Inferring emotions from speech prosody: Not so easy at age five. *PLoS One*, *8*(12), e83657.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, *38*(4), 419-439. doi: 10.1006/jmla.1997.2558

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berman, J. M., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, *107*(2), 87-99.

Berman, J. M., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology*, *142*, 391-399. doi: 10.1016/j.jecp.2015.09.014

Berman, J. M., Graham, S. A., & Chambers, C. G. (2013a). Contextual influences on children's use of vocal affect cues during referential interpretation. *The Quarterly Journal of Experimental Psychology*, *66*(4), 705-726. doi: 10.1080/17470218.2012.713367

Berman, J. M., Graham, S. A., Callaway, D., & Chambers, C. G. (2013b). Preschoolers use emotion in speech to map new words to objects. *Child Development*, *84*, 1791-1805.

Boccia, M., & Campos, J. J. (1989). Maternal emotional signals, social referencing, and infants' reactions to strangers. *New Directions for Child Development*.

Chronaki, G., Wigelsworth, M., Pell, M. D., & Kotz, S. A. (2018). The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Scientific Reports*, *8*(1), 1-17.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*(3), 412-429. doi: 10.1037/0033-2909.97.3.412

Gil, S., Hattouti, J., & Laval, V. (2016). How children use emotional prosody: Crossmodal emotional integration?. *Developmental psychology*, *52*(7), 1064.

Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, *28*(2), 219-236.

Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science*, *9*(3), 309-315.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and human performance (pp. 10.1–10.112). New York: John Wiley.

Hung, W., Ferninda, P., & Yow, W. Q. (2015). Bilingual children weigh speaker's referential cues and word-learning heuristics differently in different language contexts when interpreting a speaker's intent. *Frontiers in Psychology*, *6*, 796–796. https://doi.org/10.3389/fpsyg.2015.00796

Kao, C., Sera, M. D., & Zhang, Y. (2022). Emotional speech processing in 3-to 12-month-old infants: Influences of emotion categories and acoustic parameters. *Journal of Speech, Language, and Hearing Research*, 1-14.

Khu, M., Chambers, C., & Graham, S. A. (2018). When you're happy and I know it: Four-year-olds' emotional perspective taking during online language comprehension. *Child Development*, *89*(6), 2264-2281. doi: 10.1111/cdev.12855

Kuznetsova, A., Brockhoff, P. B., & Bojesen, C. H. (2016). Package 'lmerTest': Tests in linear mixed effects models. R package version 3.1-1. https://CRAN.R-project.org/package=lmerTest

Ma, W., Zhou, P., & Thompson, W. F. (2022). Children's decoding of emotional prosody in four languages. *Emotion*.

MacKenzie, H. K., Graham, S. A., Curtin, S., & Archer, S. L. (2014). The flexibility of 12-month-olds' preferences for phonologically appropriate object labels. *Developmental Psychology*, *50*(2), 422.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*(3), 834-843.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*(3), 718-735. doi: 10.1111/1467-8624.00311

Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development*, *74*(1), 221-237.

Namy, L. L. (2001). What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy*, *2*(1), 73-86.

Palama, A., Malsert, J., Grandjean, D., Sander, D., & Gentaz, E. (2022). The cross-modal transfer of emotional information from voices to faces in 5-, 8-and 10-year-old children and adults: An eye-tracking study. *Emotion*, *22*(4), 725.

R Core Team (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

San Juan, V., Chambers, C.G., Berman, J., Humphry, C., & Graham, S.A. (2017). The object of my desire: Five-year-olds rapidly reason about a speaker's desire during referential communication. *Journal of Experimental Child Psychology*, *162*, 101-119. doi: 10.1016/j.jecp.2017.05.003

Walle, E. A., Reschke, P. J., Knothe, J. M. (2017). Social referencing: Defining and delineating a basic process of emotion. *Emotion Review*, *9*(3), 245–252. https://doi.org/10.1177/1754073916669594

Waly, Y., Chambers, C.G., Graham, S.A. (Under Review). Preschoolers' and adults' attention to emotional prosody in an unfamiliar language [Unpublished Manuscript]. University of Calgary.

Wu, Y., Schulz, L. E., Frank, M. C., & Gweon, H. (2021). Emotion as information in early social learning. *Current Directions in Psychological Science*, *30*(6), 468-475.

Appendix A

Certificate of Research Ethics Board Approval

[Removed]

Appendix B

List of object pairs for critical and filler trials

	Positive Object	Negative Object
Critical Trials	Intact doll	Broken doll
	Inflated ball	Deflated ball
	Intact bird	Broken Bird
	Full bowl of candy	Bowl with few candies
	Clean duck	Dirty duck
	Intact egg	Broken egg
	Blooming flower	Wilted flower
	Intact fork	Bent fork
	Full glass	Spilt glass
	Intact pencil	Broken pencil
	Intact plate	Broken plate
	Intact truck	Broken truck
	Target Object	Distractor Object
Cognate Trials	Computer	Puppy
	Chocolate	Bucket
	Dinosaur	Sunglasses
	Puzzle	Bell
	Balloon	Crayons
	Telephone	Backpack

Appendix C

Demographics Questionnaire (4-year-old participants)

1. What is the highest level of education you have completed? _____

   What about the child's other parent/guardian? _____

2. Does your child attend out of home care? Yes No

   If yes, please specify (e.g., daycare, day home, etc.): _____

3. What language(s) are spoken around the child at home? _____

   What percentage of the time are they exposed to the language(s)?

   Language 1: _____ Percentage: _____ %

   Language 2: _____ Percentage: _____ %

4. Does your child have significant exposure to other languages? Yes No

   If yes, please specify? _____ Hours per week? _____

5. Does your child have any exposure to Arabic? Yes No

6. Has your child experienced any difficulties understanding or producing language?

   Yes No

   If yes, please specify: _____

Appendix D

Demographic Information (8-year-old participants)

1. What is the highest level of education you have completed? _____

   What about the child's other parent/ guardian? _____

2. Does your child attend out of home care? Yes No

   If yes, please specify (e.g., daycare, day home, etc.): _____

3. What language(s) are spoken around the child at home? _____

   What percentage of the time are they exposed to the language(s)?

   Language 1: _____ Percentage: _____ %

   Language 2: _____ Percentage: _____ %

4. Does your child have significant exposure to other languages? Yes No

   If yes, please specify? _____ Hours per week? _____

5. Describe your child's ethnic identity. _____

6. Does your child have any exposure to Arabic or Polish? Yes No

7. Has your child experienced any difficulties understanding or producing language?

   Yes No

   If yes, please specify: _____

Appendix E

Analyses reported in the present study relied on new looks to the negative object – i.e., looks initiated 200 ms after the onset of the utterance to allow sufficient time to initiate gaze-shifts. The use of new looks, however, was not outlined in the pre-registration and thus, constituted a deviation from the pre-registered analysis plan. The data reported below follows the pre-specified analysis plan and includes fixations made at the outset of the utterance. The model included the same fixed and random effects structure described in the analyses reported in the results section. These included fixed effects of prosody type and age, a random intercept by participant, a random intercept by named-object, a by-named-object random slope for prosody type, a by-participant random slope for prosody type, and a by-named-object interaction of age-group and prosody type. The model converged but displayed singular fit. Similar to analyses using new looks, a significant interaction between age group and prosody type was observed (β = 0.11, SE = 0.052, t = 2.03, p = 0.046) when non-new looks were used in the analysis.