A Multilab Study of Bilingual Infants: Exploring the Preference for Infant-Directed Speech

Krista Byers-Heinlein[1], Angeline Sin Mei Tsui[2], Christina Bergmann[3], Alexis K. Black[4], Anna Brown[5], Maria Julia Carbajal[6], Samantha Durrant[5], Christopher T. Fennell[7], Anne-Caroline Fiévet[6], Michael C. Frank[2], Anja Gampe[8], Judit Gervain[9,10], Nayeli Gonzalez-Gomez[11], J. Kiley Hamlin[12], Naomi Havron[6], Mikołaj Hernik[13], Shila Kerr[14], Hilary Killam[1], Kelsey Klassen[15], Jessica E. Kosie[16], Ágnes Melinda Kovács[17], Casey Lew-Williams[16], Liquan Liu[18], Nivedita Mani[19,20], Caterina Marino[9], Meghan Mastroberardino[1], Victoria Mateu[21], Claire Noble[5], Adriel John Orena[14], Linda Polka[14], Christine E. Potter[16], Melanie S. Schreiner[19], Leher Singh[22], Melanie Soderstrom[15], Megha Sundara[23], Connor Waddell[18], Janet F. Werker[12], and Stephanie Wermelinger[8]

[1]Department of Psychology, Concordia University; [2]Department of Psychology, Stanford University; [3]Language Development Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; [4]School of Audiology & Speech Sciences, University of British Columbia; [5]School of Psychology, University of Liverpool; [6]Département d'Études cognitives, École normale supérieure, École des hautes études en sciences sociales, Centre national de la recherche scientifique (CNRS), Université Paris Sciences et Lettres; [7]School of Psychology, University of Ottawa; [8]Department of Psychology, University of Zurich; [9]Integrative Neuroscience and Cognition Center (INCC), CNRS and University of Paris; [10]Department of Developmental and Social Psychology, University of Padua; [11]Department of Psychology, Oxford Brookes University; [12]Department of Psychology, University of British Columbia; [13]Department of Psychology, The University of Tromsø – The Arctic University of Norway; [14]School of Communication Sciences and Disorders, McGill University; [15]Department of Psychology, University of Manitoba; [16]Department of Psychology, Princeton University; [17]Department of Cognitive Science, Central European University; [18]School of Social Sciences and Psychology, Western Sydney University; [19]George-Elias-Müller Institute for Psychology, University of Göttingen; [20]Leibniz Science Campus Primate Cognition, Göttingen, Germany; [21]Department of Spanish and Portuguese, University of California, Los Angeles; [22]Department of Psychology, National University of Singapore; and [23]Department of Linguistics, University of California, Los Angeles

Abstract

From the earliest months of life, infants prefer listening to and learn better from infant-directed speech (IDS) compared with adult-directed speech (ADS). Yet IDS differs within communities, across languages, and across cultures, both in form and in prevalence. This large-scale, multisite study used the diversity of bilingual infant experiences to explore the impact of different types of linguistic experience on infants' IDS preference. As part of the multilab ManyBabies 1 project, we compared preference for North American English (NAE) IDS in lab-matched samples of 333 bilingual and 384 monolingual infants tested in 17 labs in seven countries. The tested infants were in two age groups: 6 to 9 months

Corresponding Author:
Krista Byers-Heinlein, Department of Psychology, Concordia University
E-mail: k.byers@concordia.ca

and 12 to 15 months. We found that bilingual and monolingual infants both preferred IDS to ADS, and the two groups did not differ in terms of the overall magnitude of this preference. However, among bilingual infants who were acquiring NAE as a native language, greater exposure to NAE was associated with a stronger IDS preference. These findings extend the previous finding from ManyBabies 1 that monolinguals learning NAE as a native language showed a stronger IDS preference than infants unexposed to NAE. Together, our findings indicate that IDS preference likely makes similar contributions to monolingual and bilingual development, and that infants are exquisitely sensitive to the nature and frequency of different types of language input in their early environments.

When caregivers interact with their infants, their speech often takes on specific, distinguishing features in a speech register known as infant-directed speech (IDS; Fernald et al., 1989). IDS is produced by caregivers of most (although not all) linguistic and cultural backgrounds and is typically characterized by a slow, melodic, high-pitched, and exaggerated cadence (Farran et al., 2016; Fernald et al., 1989; Kitamura et al., 2001; Pye, 1986; Shute & Wheldall, 1999). From early in life, infants tune their attention to IDS, preferring to listen to IDS over adult-directed speech (ADS) at birth (Cooper & Aslin, 1990), as well as later in infancy (Cooper & Aslin, 1994; Cooper et al., 1997; Fernald, 1985; Hayashi et al., 2001; Kitamura & Lam, 2009; Newman & Hussain, 2006; Pegg et al., 1992; Santesso et al., 2007; Singh et al., 2002; Werker & McLeod, 1989).

Infants' preference for IDS may play a useful role in early language learning. For example, infants are better able to discriminate speech sounds in IDS than in ADS (Karzon, 1985; Trainor & Desjardins, 2002), more efficiently segment words from continuous speech in an IDS register (Thiessen et al., 2005), demonstrate better long-term memory for words spoken in IDS (Singh et al., 2009), and learn new words more effectively from IDS (Graf Estes & Hurley, 2013; Ma et al., 2011; but see Schreiner et al., 2016).

Although most studies have confirmed a general, early preference for IDS, to date there has been very little research aimed at understanding how different linguistic experiences affect infants' preferences. For instance, although the use of IDS has been demonstrated in a large number of cultures (see the sources cited above), the vast majority of the research on infants' IDS preferences has been conducted in North America, using English speech typically directed at infants learning North American English (NAE; Dunst et al., 2012). Most critically, past work has been limited to a particular kind of linguistic (and cultural) experience: that of the monolingual infant. Here, we present a large-scale, multisite,

preregistered study on bilingual infants, a population that is particularly suited for exploring the relationship between language experience and IDS preference. Moreover, this research provides important insight into the early development of bilingual infants, a large but understudied population.

Does Experience Tune Infants' Preference for IDS?

What role might experience play in tuning infants' attention to IDS? We aggregated results from a published meta-analysis (Dunst et al., 2012) with additional community-contributed data available through the Meta-Lab database (http://metalab.stanford.edu) to examine their combined results. When all 62 studies were considered, we found a moderately sized average effect, Cohen's $d = 0.64$, showing a preference for IDS compared with ADS. Focusing on the 22 studies most similar to ours (testing IDS preference in a laboratory using looking times of typically developing infants ages 3 through 15 months, with stimuli consisting of naturally produced English-spoken IDS from an unfamiliar female speaker), we found a slightly smaller effect size, $d = 0.60$. Although this meta-analysis focused on infants in the first year of life, other studies of infants aged 18 to 21 months have also found a preference for IDS over ADS (Glenn & Cunningham, 1983; Robertson et al., 2013). There is some evidence that older infants show a greater preference for IDS than younger infants (Dunst et al., 2012), although an age effect was not found in the subsample of 22 studies mentioned above. More evidence is needed to explore the possibility that increased language experience as children grow enhances their preference for IDS.

Another experimental variable that would be important in understanding the role of experience in the preference for IDS is whether the speech stimuli are presented in a native or nonnative language. Numerous studies in early perception have found different developmental trajectories

for perception of native versus nonnative stimuli (e.g., discriminating human faces vs. discriminating monkey faces in Lewkowicz & Ghazanfar, 2006; discriminating native vs. discriminating nonnative speech-sound categories in Maurer & Werker, 2014; segmenting word forms from fluent speech in Polka & Sundara, 2012). Generally, whereas infants show increasing proficiency in discriminating the types of faces and sounds that are present in their environment, they lose sensitivity to the differences between nonnative stimuli over time. This general pattern might lead one to predict that infants will initially be sensitive to differences between IDS and ADS in both the native and the nonnative languages, but that this initial cross-linguistic sensitivity will decline with age. In other words, at some ages, infants' preference for IDS over ADS could be enhanced when they hear their native language. However, to date, researchers have collected very little data relevant to this question. Importantly, this general trend, if it exists, may interact with differences across languages in the production of IDS. The exaggerated IDS of NAE might be either more interesting or less interesting to an infant whose native language is characterized by a less exaggerated form of IDS than to an infant who regularly hears NAE IDS.

Only a handful of IDS-preference studies have explicitly explored infants' preference for IDS in a native versus a nonnative language. Werker et al. (1994) compared 4.5- and 9-month-old English- and Cantonese-learning infants' preference for videos of Cantonese mothers using IDS versus ADS. Both groups showed a preference for IDS; however, the magnitude of the preference was not specifically compared between the two groups. Hayashi et al. (2001) studied Japanese-learning infants' (aged 4–14 months) preference for native (Japanese) and nonnative (English) speech. These infants generally showed a preference for Japanese IDS over Japanese ADS, as well as an increasing preference for Japanese IDS over English IDS. The latter finding shows that infants tune into their native language with increased experience; however, as the study did not measure infants' interest in English ADS, it does not shed light on whether the infants were equally sensitive to the difference between ADS and IDS in the nonnative stimuli, or whether and how such sensitivity might change over time.

Infants growing up bilingual are typically exposed to IDS in two languages. They provide a particularly useful wedge in understanding experiential influences on infants' attention to IDS. Bilingual infants receive less exposure to each of their languages than monolingual infants do, and the exact proportion of exposure to each of their two languages varies from infant to infant. This divided exposure does not appear to slow the overall rate of language acquisition: Bilinguals pass their language milestones, such as the onset of babbling and the production of their first words, on approximately the same schedule as monolingual infants (Werker & Byers-Heinlein, 2008). Nonetheless, children from different language backgrounds receive different types of input and must ultimately acquire different language forms, which can alter some patterns of language acquisition (e.g., Choi & Bowerman, 1991; Slobin, 1985; Tardif, 1996; Tardif et al., 1997; Werker & Tees, 1984). As a consequence, bilingual infants allow researchers to investigate how a given "dose" of experience with a specific language relates to phenomena in language acquisition, while holding infants' age and total experience with language constant (Byers-Heinlein & Fennell, 2014).

Aside from providing the opportunity to study dose effects, research on the preference for IDS in bilingual infants is important for the sake of understanding bilingual development itself. Several lines of research suggest that early exposure to two languages changes some aspects of early development (Byers-Heinlein & Fennell, 2014), including bilinguals' perception of nonnative speech sounds (i.e., sounds that are in neither of their native languages). For example, a number of studies have found that bilinguals maintain sensitivity to nonnative consonant contrasts (García-Sierra et al., 2016; Petitto et al., 2012; Ramírez et al., 2017), nonnative tone contrasts (Graf Estes & Hay, 2015; Liu & Kager, 2017a), and visual differences between languages (i.e., rhythmic and phonetic information available on talkers' faces; Sebastián-Gallés et al., 2012) until a later age than monolinguals. Other studies have suggested that bilinguals' early speech perception is linked to their language dominance (Liu & Kager, 2015; Molnar et al., 2016; Sebastián-Gallés & Bosch, 2002), such that bilinguals' perception more closely matches that of monolinguals in their dominant language than that of monolinguals in their nondominant language. Bilingual infants also demonstrate some cognitive differences from monolinguals that are not specific to language, including faster visual habituation (Singh et al., 2015), better memory generalization (Brito & Barr, 2014; Brito et al., 2015), and greater cognitive flexibility (Kovács & Mehler, 2009a, 2009b). This might reflect an early-emerging difference in information processing between the two groups. Together, these lines of work raise the possibility that preference for IDS over ADS could have a different developmental course for bilingual and monolingual infants, and that bilinguals' distinct course could interact with factors such as language dominance.

Bilinguals' Exposure to and Learning From IDS

Overall, there is very little research on whether bilinguals' experience with IDS is comparable to monolinguals' experience. Some research has compared English monolinguals and English-Spanish bilinguals in the United

States (Ramírez-Esparza et al., 2014, 2017). The researchers reported that bilingual infants around 1 year of age received less exposure to IDS than monolingual infants on average. Moreover, in the bilingual families, input was more evenly distributed across infant- and adult-directed registers. It is difficult to know whether the results reported in these studies generalize to other populations of bilinguals, or whether they are specific to this language community. As acknowledged by the authors, the bilinguals in this study were of a lower socioeconomic status (SES) than the monolinguals, which could have driven differences in the amount of IDS that infants heard. On the other hand, it might be the case that bilingual infants more rapidly lose their preference for the IDS register than do monolinguals, and that caregivers of bilinguals respond to this by reducing the amount of IDS input they provide.

Bilingual infants might also hear IDS that differs prosodically and phonetically from that heard by monolingual infants. Bilingual infants often have bilingual caregivers, and even when they are highly proficient speakers, their speech may vary from that of monolinguals. One study compared vowels produced in the IDS of monolingual English, monolingual French, and balanced French-English bilingual mothers living in Montreal (Danielson et al., 2014). Bilingual mothers' vowels were distinct in the two languages, and the magnitude of the difference between French and English vowels was similar to that shown by monolingual mothers. However, another study showed that, in a word-learning task, 17-month-old French-English bilinguals learned new words better from a bilingual speaker than a monolingual speaker, even though acoustic measurements did not reveal what dimension the infants were attending to (Fennell & Byers-Heinlein, 2014; similar findings were reported in Mattock et al., 2010). Finally, a study of Spanish-Catalan bilingual mothers living in Barcelona found that some were more variable than monolingual mothers in their productions of a difficult Catalan vowel contrast (Bosch & Ramon-Casas, 2011). Thus, not only may bilingual infants differ from monolingual infants in the amount of IDS they hear in a particular language, but also different populations of bilingual infants may vary in how similar the IDS they hear is to monolingual-produced IDS in the same languages. This could, in turn, lead to greater variability across bilinguals in their preference for IDS over ADS when tested with any particular stimulus materials.

Regardless of bilingual infants' specific experience with IDS, evidence suggests that bilinguals might enjoy the same learning benefits from IDS as monolinguals. For example, Ramírez-Esparza et al. (2017) found that greater exposure to IDS predicted larger vocabulary size in both monolingual and bilingual infants. Indeed, an untested possibility is that exposure to IDS might be of particular benefit to bilingual infants. Bilinguals face a more complex learning situation than monolinguals, as they acquire two sets of sounds, words, and grammars simultaneously (Werker & Byers-Heinlein, 2008). This raises the possibility that bilingual infants might have enhanced interest in IDS relative to monolinguals, or that they might maintain a preference for IDS until a later age than monolinguals, much as bilingual infants' perception has an extended sensitivity to nonnative phonetic contrasts.

Replicability in Research With Bilingual Infants

Working with bilingual infant populations engenders unique replicability issues above and beyond those common in the wider field of infant research (e.g., between-lab variability, methodological variation; see Frank et al., 2017). These issues begin with the nature of the population. Our discussion of bilingual infants thus far has used "bilingual" as a blanket term to describe infants growing up hearing two or more languages. However, this usage belies the large variability in groups of infants described as bilingual. First, some studies of bilinguals have included infants from a homogeneous language background (i.e., all infants are exposed to the same language pair; e.g., English-Spanish in Ramírez-Esparza et al., 2017), whereas others have included infants from heterogeneous language backgrounds (i.e., infants are exposed to different language pairs, e.g., English-other, where "other" might be Spanish, French, Mandarin, Punjabi, etc.; e.g., Fennell et al., 2007). Second, some bilinguals learn two typologically closely related languages (e.g., Spanish-Catalan), whereas others learn two distant languages (e.g., English-Mandarin). Third, there is wide variability among bilingual infants in the amount of exposure to each language, which introduces an extra dimension of individual differences relative to studies with monolingual infants. Fourth, studies define bilingualism in different ways, ranging from a liberal criterion of at least 10% exposure to the nondominant language to at least 40% exposure to the nondominant language (Byers-Heinlein, 2015). Finally, bilingual and monolingual populations can be difficult to compare because of cultural, sociological, and SES differences that exist between samples.

All of the above difficulties have resulted in very few findings being replicated across different samples of bilinguals. The limited research that has compared different types of bilingual learners has indicated that the particular language pair being learned by bilingual infants influences perception of both native (Bialystok et al., 2005; Sundara & Scutellaro, 2011) and nonnative (Patihis et al., 2015) speech sounds. In contrast, other studies have not found differences between bilinguals

learning different language pairs in, for example, their ability to apply speech perception skills to a word-learning task (Fennell et al., 2007). Generally, it is not known how replicable most findings are across different groups of bilinguals, or how previously reported effects of bilingualism on learning and perception are impacted by the theoretically interesting moderators discussed above.

Research on bilingual infants also faces many of the same general concerns as other infancy research, such as challenges recruiting sufficient numbers of participants to conduct well-powered studies (Frank et al., 2017). Finding an appropriate bilingual sample further limits the availability of research participants, even in locations with significant bilingual populations. Such issues are particularly relevant given the recent emphasis on replicability and best practices in psychological science (Klein et al., 2014; Open Science Collaboration, 2015; Simmons et al., 2011). Of particular interest is whether bilingual infants as a group show greater variability in their responses than monolingual infants, and how to characterize the variability of responses between the different types of samples of bilinguals that can be recruited by particular labs (e.g., homogeneous vs. heterogeneous samples). Understanding whether variability differs systematically across groups is vital for planning appropriately powered studies.

Description of the Current Study

Here, we report a large-scale, multisite, preregistered study aimed at using data from bilingual infants to understand variability in infants' preference for IDS over ADS. This study, ManyBabies 1 Bilingual, is a companion project to the ManyBabies 1 project, published in a previous issue of this journal (ManyBabies Consortium, 2020). The two studies were conducted in parallel, using the same stimuli and experimental procedure. However, whereas the ManyBabies 1 analyses included all data collected from monolingual infants (including those data from monolinguals reported here), we report on a subset of these data together with additional data from bilingual infants not included in that project's analyses. Our multisite approach gave us precision in estimating the overall effect size of bilingual infants' preference for IDS, while also allowing us to investigate how different types of language experience moderate this effect.

Our primary approach was to compare bilinguals' performance with the performance of monolinguals tested in the same lab. This approach had two notable advantages. First, within each lab, bilinguals shared one of their two languages with monolinguals (the language of the wider community). Second, testing procedures were held constant within each lab. Thus, this approach allowed us to minimize procedural confounds with infants' bilingual status. However, a disadvantage of this approach is that it left out data from monolingual infants tested in other labs (because not all laboratories provided data from bilingual infants), which could potentially have added precision to the measured effects. Thus, we performed additional analyses comparing all bilinguals with all monolinguals within the same age bins, regardless of the lab each had been tested in.

We tested bilinguals in two age windows: 6 to 9 months and 12 to 15 months. The specific age bins selected were based on a preliminary survey of participating laboratories' access to participants of different age ranges. The choice of nonadjacent age bins also increased the chances of observing developmental differences.

All infants were tested using the same stimuli, which consisted of recordings of NAE-accented IDS and ADS. Because of the international nature of this multisite project, these stimuli were native for some infants but nonnative for other infants, both in terms of the language of the stimuli (English) and the variety of IDS (NAE IDS is particularly exaggerated in its IDS characteristics relative to other varieties of IDS; see Soderstrom, 2007, for a review). Moreover, the stimuli were produced by monolingual mothers. Thus, infants' exposure to the type of stimuli used varied from low (monolinguals and bilinguals not exposed to NAE), to moderate (bilinguals learning NAE as one of their two languages), to high (monolinguals learning NAE).

Infants were tested in one of three experimental setups regularly used to test infant auditory preference: central fixation, eye tracking, and head-turn-preference procedure (HPP). The use of a particular setup was the choice of each lab, depending on their equipment and expertise. Labs that tested both monolinguals and bilinguals used the same setup for both groups. In all setups, infants heard a series of trials presenting either IDS or ADS, and their looking time to an unrelated visual stimulus (e.g., a checkerboard) was used as an index of their attention. In the central-fixation setup, infants sat in front of a single screen that displayed a visual stimulus, and their looking time to this visual stimulus while an auditory stimulus was played was coded via button press using a centrally positioned camera. Looking time was recorded similarly in the eye-tracking setup, except that infants' looking was coded automatically using a corneal-reflection eye tracker. In the HPP (see Kemler Nelson et al., 1995), infants sat in the middle of a room facing a central visual stimulus. Their attention was drawn to the left or right side of the room by a visual stimulus while the auditory stimulus played, and the duration of their looking to the visual stimulus was measured via button press using a centrally positioned camera.

Research Questions

We identified three basic research questions to be addressed by this study. Note that it was not always possible to make specific predictions given the very limited

data on infants' cross-language preferences for IDS over ADS, and particularly the absence of data from bilingual infants. We also note that the ManyBabies 1 project, focusing on monolingual infants, addressed other more general questions, such as the average magnitude of the IDS preference, changes in preference over age, and the effects of methodological variation on IDS preference (ManyBabies Consortium, 2020). The main questions we planned to address using data from bilingual infants were

- How does bilingualism affect infants' interest in IDS relative to ADS? As described above, monolingual infants display an early preference for IDS that grows in strength at least through the first year of life. We anticipated that the bilingual experience might result in a different pattern of IDS preference; however, the direction and potential source of any difference were difficult to predict. For example, the more challenging nature of early bilingual environments might induce an even greater preference for IDS over ADS than is found in monolinguals. This enhanced preference could be shown across development, or might be observed only at certain ages. On the other hand, given some evidence that parents of bilingual infants produce relatively less IDS than parents of monolingual infants, it may be that bilinguals show less interest in IDS than monolinguals. We also explored the following questions concerning potential sources for an emerging difference between populations: If an overall difference between monolingual and bilingual infants' preference for IDS is observed, can this be accounted for by systematic differences in SES? Do bilinguals show greater variability in their preference for IDS than monolinguals?
- How does the amount of exposure to NAE IDS affect bilingual infants' listening preferences? Although we expected infants across different language backgrounds to show greater interest in IDS over ADS, we investigated whether this effect was moderated by the amount of exposure to NAE. For monolinguals, this exposure was either 100% (monolingual learners of NAE) or 0% (monolingual learners of other languages). For bilinguals, some infants had 0% exposure to NAE IDS (e.g., bilingual infants learning Spanish and Catalan), whereas others had a range of different exposures (e.g., bilingual infants learning NAE and French). This allowed us to at least partially disentangle dose effects of exposure to NAE IDS from infants' bilingualism. An additional possibility was that infants' exposure to NAE would predict overall

attention to both infant-directed and adult-directed NAE, but have no differential effects on interest in IDS versus ADS. Alternately, it was possible that our stimuli would be equally engaging to infants regardless of their experience with NAE.
- Finally, we had planned to ask how bilingual infants' listening to NAE IDS and ADS is impacted by the particular language pair being learned. We intended to ask this question at both the group and the individual level. At the group level, we planned to investigate whether different patterns of overall preference for IDS and group-level variability would be seen in homogeneous versus heterogeneous samples of bilinguals. However, ultimately, we had insufficient homogeneous samples to address this question. At the individual level, we were interested in how the particular language pair being learned modulated infants' preference for IDS. As we did not know a priori what language pairs would have a sufficient sample size for analysis, this was considered a potential exploratory analysis. Ultimately, given the nature of our main results and the diverse language backgrounds of our final sample, we decided to leave this question open for future investigations.

Disclosures

Preregistration

The accepted Stage 1 version of this manuscript and the analysis plan were preregistered via OSF, at https://osf.io/wtfuq.

Data, materials, and online resources

Study instructions and other details are available at the ManyBabies 1 Bilingual OSF site, https://osf.io/zauhq/, and materials are available via the ManyBabies 1 OSF site, https://osf.io/re95x/.

Labs submitted anonymized data for central analysis that identified participants by code only. Data and analytic code are available at https://github.com/manybabies/mb1b-analysis-public. Video recordings of individual participants were coded and stored locally at each lab, and when possible were uploaded to a central controlled-access data bank accessible to other researchers (https://databrary.org).

Reporting

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in our study.

Ethical approval

This research was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Each lab followed the ethical guidelines and ethics-review-board protocols of their own institution.

Method

Participation details

Our monolingual sample originated from the ManyBabies 1 project (ManyBabies Consortium, 2020). Here we report some basic information about that sample (the reader is referred to the original study for further details) and focus primarily on the bilingual sample.

Time frame. An open call for labs to participate was issued on February 2, 2017. Participant testing began on May 1, 2017. Testing for monolinguals ended on April 30, 2018. Because of the additional difficulty of recruiting bilingual samples, the end date for collection of these data was extended by 4 months, to August 31, 2018. Because of a miscommunication, one lab continued testing beyond this deadline but prior to data analysis, and these data were included in the final sample.

Age distribution. Labs contributing data from bilingual infants were asked to test participants in at least one of the two (but preferably both) age bins: 6- to 9-month-olds (i.e., ages 6 months 1 day through 9 months 0 days) and 12- to 15-month-olds (i.e., ages 12 months 1 day through 15 months 0 days). Labs were asked to aim for a mean age at the center of the bin, with distribution across the entire age window. Some labs chose to test additional infants outside the target age ranges for future exploratory analyses; these infants were excluded from the current study.

Lab participation criterion. Considering the challenges associated with recruiting bilingual infants and the importance of counterbalancing in our experimental design, we asked labs to contribute data from a minimum of 16 infants per age and language group (note that infants who met inclusion criteria for age and language exposure but were ultimately excluded for other reasons counted toward this minimum). We expected that requiring a relatively low minimum number of infants would encourage more labs to contribute a bilingual sample, and under our statistical approach, a larger number of groups was more important than a larger number of individuals (Maas & Hox, 2005). However, labs were encouraged to contribute additional data provided that decisions about when to stop data collection were made ahead of time (e.g., by declaring intended start and end dates before data collection). A sensitivity analysis showed that with a sample of 16 infants

and assuming the average effect size of similar previous studies (Cohen's $d = 0.70$; Dunst et al., 2012; data available through the MetaLab database), individual labs would have 74% power to detect a preference for IDS in a paired-samples t test ($\alpha = .05$, one-tailed). Assuming a smaller effect size of 0.60, a conservative estimate based on our previously mentioned analysis of 22 studies most similar to ours, individual labs' power would be 61%. The moderate statistical power that individual labs would have to detect this effect highlights the importance of our approach to combine data across labs. We note that some labs were unable to recruit their planned minimum sample of 16 bilingual infants who met our inclusion criteria in the time frame available, a point we return to later.

Labs were asked to screen infants ahead of time for inclusion criteria, typically by briefly asking about language exposure over the phone. Despite this screening process, some infants who arrived in the lab for testing fell between the criteria for monolingual and bilingual status based on the comprehensive questionnaire. In such cases, the decision whether to test the infant was left up to individual laboratories' policy, but we asked that data from any babies who entered the testing room be submitted for data processing (even though some such data might be excluded from the main analyses).

Participants

Defining bilingualism. Infants are typically categorized as bilingual on the basis of their parent-reported relative exposure to their languages. However, studies vary considerably in terms of inclusion criteria for the minimum exposure to the nondominant language, which in previous studies has ranged from 10% to 40% of infants' total exposure (Byers-Heinlein, 2015). Some bilingual infants may also have some exposure to a third or fourth language. Finally, infants can vary in terms of when the exposure to their additional languages began, which can be as early as birth or any time thereafter. We aimed to take a middle-of-the-road approach to defining bilingualism, attempting to balance a need for experimental power with a need for interpretable data.

Thus, we asked each participating lab to recruit a group of simultaneous bilingual infants who were exposed to each of two languages between 25% and 75% of the time and who had regular exposure to both languages beginning within the first month of life. There was no restriction as to whether infants were exposed to additional languages; thus, some infants could be considered multilingual (although we continue to use the term bilingual throughout this article). These criteria would include, for example, an infant with 40% English, 40% French, and 20% Spanish exposure, but would exclude an infant with 20% English, 70% French, and

10% Spanish exposure. We also asked labs to recruit a sample of bilingual infants who shared at least one language—the community language being learned by monolinguals tested in the same lab. Labs in bilingual communities (e.g., Barcelona, Ottawa, Montréal, Singapore) were free to decide which community language to select as the shared language. Within this constraint, most labs opted to test heterogeneous groups of bilinguals; for example, among the English-other bilinguals for whom English was the community language, the other language might be French, Spanish, Mandarin, or some other language. Only one lab tested a homogeneous group of bilinguals (in this case, all infants were learning English and Mandarin), although we had expected that more labs would test homogeneous samples, given that both heterogeneous and homogeneous samples are used regularly in research with bilingual infants. Because only one homogeneous sample was tested, we were not able to conduct planned analyses examining the impact of this type of sample on our results. Infants who were tested but did not meet inclusion criteria for the bilingual group (e.g., because they did not hear enough of their nondominant language or did not hear enough of the community language) were excluded from the main analyses, but retained for exploratory analyses when appropriate.

Assessing bilingualism. Each lab was asked to use a detailed day-in-the-life parental interview questionnaire to quantify the percentage of time that infants were exposed to each language. This approach has been shown to predict bilingual children's language outcomes better than a one-off parental estimate (DeAnda et al., 2016). Moreover, recent findings based on day-long home language recordings show that caregivers can reliably estimate their bilingual child's relative exposure to each language (Orena et al., 2020). Labs were also asked to pay special attention to whether infants had exposure to NAE (based on caregivers' report of the variety of English spoken to their infants), and if so, which caregiver (or caregivers) this input came from. As most of the labs contributing bilingual data had extensive expertise in assessing bilingual language background, we encouraged each lab to use whatever measurement instrument was normally used in their lab (details of the assessment instruments, including source references for most measures, are outlined below). When possible, labs conducted the interview in the parents' language of choice and documented whether the parents' preferred language was able to be used.

Although standardization of measurement tools is often desirable, we reasoned that different questions and approaches might be best for eliciting information from parents in different communities and from different cultures. Indeed, many labs reported that their own instruments had undergone considerable refinement over the years as a function of their experience working with the families in their communities. However, in order to maximize the overall sample size and the diversity of bilingual groups tested, we encouraged participation from laboratories without extensive experience testing bilingual infants. Labs that did not have an established procedure were paired with more experienced labs working with similar communities to refine a language assessment procedure. Twelve of the labs administered a structured interview-style questionnaire based on the one developed by Bosch and Sebastián-Gallés (1997, 2001; for examples of the measure, see the online supplementary materials of Byers-Heinlein et al., 2019, and DeAnda et al., 2016), and the remaining five labs administered other questionnaires. We describe each of these approaches in detail below.

The Bosch and Sebastián-Gallés (1997, 2001) questionnaire is typically referred to in the literature as the Language Exposure Questionnaire (LEQ; e.g., Byers-Heinlein et al., 2013) or the Language Exposure Assessment Tool (LEAT; DeAnda et al., 2016). For simplicity, we use the former name here. Administration of the questionnaire takes the form of a parental interview, in which a trained experimenter systematically asks at least one of the infant's primary caregivers detailed questions about the infant's language environment. The interviewer obtains an exposure estimate for each person who is in regular contact with the infant, defined as a minimum contact of once a week. The caregiver gives an estimate of how many hours each of these people speaks to the infant in each language on each of the days of the week (e.g., weekdays and weekends may differ depending on work commitments). Further, the caregiver is asked if the language input from each regular-contact person has been similar across the infant's life history. If not, such as in the case of a caregiver returning to work after parental leave or leaving for an extended stay in another country, an estimate is derived for each different period of the infant's life span. The interviewer also asks the caregiver about the language background of each person with regular contact with the infant (as defined above), asking what languages that person speaks and whether he or she is a native speaker of those languages. The caregiver also gives an estimate of language exposure in the infant's day-care facility, if applicable.

Finally, the caregiver gives a global estimate of the infant's percentage of exposure to the two languages, which includes input from those people in regular contact with the infant and other people with whom the infant has less regular contact (e.g., playgroups, friends of caregivers). This global estimate does not include input from television or radio, as such sources have no known positive impact, and may even have a negative impact on monolingual and bilingual language development in

infancy (see Hudon et al., 2013). The primary estimate of the infant's cumulative percentage of exposure in each language is derived from a time-weighted average of input from the primary individuals in the infant's life. Some labs use the global estimate simply to confirm these percentages. Other labs average the primary and global exposure to take into account all language exposure while still giving more weight to the primary individuals. Also, some labs asked additional questions, for example, about videoconferencing with relatives, caregivers' possible mixing of languages when speaking to the infant, or caregivers' cultural background. Finally, although the original form was pen and paper, there have been adaptations, which include using a form-fillable Excel sheet (DeAnda et al., 2016).

For the other language-exposure measures used by five of the labs, we simply highlight the differences from the LEQ, as there is much overlap among all the instruments used to measure infants' exposure to their languages. Two labs used custom assessment measures they designed themselves. The major difference from the LEQ for the first of these custom measures is that parents estimate the percentage of exposure for each language from primary individuals in the infant's life, rather than the number of hours of exposure per day in each language. The other custom measure, unlike the LEQ, specifies estimates of language exposure in settings where more than one speaker is present by weighting each speaker's language contribution. Two other labs used other measures present in the literature: One used the Multilingual Infant Language Questionnaire (MILQ; Liu & Kager, 2017b), and the other used an assessment measure designed by Cattani et al. (2014). For the MILQ, one major difference from the LEQ is that parents complete the assessment directly using an Excel sheet with clear instructions. The other major difference is that the MILQ is much more detailed: It breaks down language exposure to very specific activities (e.g., car time, book reading, mealtime), asks for more detail about the people in regular contact with the infant (e.g., accented speech, level of talkativeness), and asks for estimates of media exposure (e.g., TV, music). The measure from Cattani et al. focuses on parental exposure and uses Likert scales to determine exposure from each parent. The ratings are converted to percentages, and maternal exposure is weighted more in the final calculation because of data showing that mothers are more verbal than fathers. Finally, one lab did not use a detailed measure, but rather simply asked parents to give an estimate of the percentage exposure to each of the languages their infant was hearing.

For monolinguals, labs either used the same assessment as with bilinguals or minimally checked participants' monolingual status by asking parents a single question: to estimate the percentage of time that their infant was exposed to their native language. Under either approach, if that estimate exceeded 90% exposure to a single language, the infant was considered monolingual.

Demographics. Each lab administered a questionnaire that gathered basic demographic data about the infants, including age, health history, and gestation. Infants' SES was measured via parental report of years of maternal education. To standardize the data across different education systems, where formal schooling may begin at different ages, we counted the number of years of education after kindergarten. For example, in the United States, mothers who had completed high school were considered to have 12 years of education.

Final sample. The final sample of bilinguals who met our infant-level inclusion criteria included 333 infants tested in 17 labs; 148 were 6 to 9 months old, and 185 were 12 to 15 months old (a full account of exclusions is detailed in the Results section). These 17 labs also collected data from 384 monolingual infants who met infant-level inclusion criteria, of whom 181 were 6 to 9 months old and 203 were 12 to 15 months old. Although all analyses required that the data met the infant-level inclusion criteria, some analyses further required that the data met the lab-level inclusion criteria (lab-level inclusion criteria are discussed in the Results section in connection with the specific analyses for which they were implemented). Data from monolingual infants in these age ranges were available from 59 additional labs (6- to 9-month-olds: $n = 574$; 12- to 15-month-olds: $n = 463$) that did not contribute bilingual data. Table 1 provides summary information describing the bilingual infants and lab-matched monolingual samples. The appendix details the gender distributions across subsamples (Table A1) and the language pairs being learned by bilingual infants (Table A2).

Materials

Visual stimuli. Labs using a central-fixation or eye-tracking method presented infants with a brightly colored checkerboard as the main visual stimulus. A video of a laughing baby was used as an attention getter between trials to reorient infants to the screen. Labs using the HPP used the typical visual stimuli employed in their labs, which were sometimes light bulbs (as in the original development of the procedure in the 1980s) or sometimes colorful stimuli presented on LCD screens. All visual stimuli are available via the ManyBabies 1 OSF site at osf.io/re95x/.

Auditory stimuli. Auditory stimuli consisted of seminaturalistic recordings of mothers interacting with their infants (ranging in age from 122 to 250 days) in a laboratory setting. Mothers were asked to talk about a set of objects with their infant and also separately with an experimenter. A set

Table 1. Sample Characteristics by Lab

			Sample size (n)				Average maternal education (years)	Bilinguals' average NAE exposure (%)
Institution	Lab	Method	6- to 9-month-old bilinguals	6- to 9-month-old monolinguals	12- to 15-month-old bilinguals	12- to 15-month-old monolinguals		
Oxford Brookes University	babylabbrookes	Central fixation	17	15	17	16	16.76	0.00
Western Sydney University	babylabkingswood	HPP	9	15	15	15	16.98	0.00
Université Paris Descartes	babylabparisdescartes1	HPP	10	0	1	16	16.33	0.00
Central European University	cdcceu	Eye tracking	0	0	14	13	18.15	0.00
University of Liverpool	lllliv	Eye tracking	7	19	6	15	17.38	0.00
École normale supérieure	lscppsl	Eye tracking	0	0	16	14	16.98	0.00
National University of Singapore	nusinfantlanguagecentre	Eye tracking	26	10	12	10	14.99	0.00
University of Zurich	weltentdeckerzurich	Eye tracking	0	0	28	30	15.17	0.00
University of Göttingen	wsigoettingen	Central fixation	9	31	7	15	16.06	0.00
McGill University	isplabmcgill	HPP	0	0	16	11	18.07	28.63
University of Manitoba	bllumanitoba	HPP	7	26	8	16	15.54	47.16
University of British Columbia (Werker)	infantstudiesubc	HPP	15	20	0	0	16.38	48.88
Concordia University	irlconcordia	Eye tracking	16	17	18	18	16.63	48.98
Princeton University	babylabprinceton	HPP	15	1	0	0	18.00	49.07
University of California, Los Angeles	langlabucla	HPP	0	0	9	3	14.91	53.04
University of Ottawa	ldlottawa	Central fixation	7	17	18	11	17.99	54.79
University of British Columbia (Hamlin)	infantcogubc	Eye tracking	10	11	0	0	16.50	55.69

Note: The labs are listed in order of their samples' exposure to North American English (NAE). Because of lab-level inclusion criteria, cells with fewer than 10 participants were excluded from the meta-analyses, but were included in the mixed-effects regression analyses. Labs that tested only monolingual infants are not listed. HPP = head-turn-preference procedure.

of eight IDS and eight ADS auditory stimuli of 18 s each was created from these recordings. Details regarding the recording and selection process, acoustic details, and ratings from naive adult listeners can be found in the ManyBabies 1 report (ManyBabies Consortium, 2020) and the associated OSF project at osf.io/re95x.

Procedure

Each lab used one of three procedures common in infant studies, according to their own expertise and the experimental setups available in the lab: central fixation (three labs), eye tracking (seven labs), or HPP (seven labs). The testing procedure was identical to that used in the ManyBabies 1 project (ManyBabies Consortium, 2020; deviations from the protocol are also described there), and we describe key aspects here.

Infants sat on a parent's lap or in a high chair, and parents listened to masking music over headphones throughout the study. Infants saw two training trials that presented an unrelated auditory stimulus (piano music), followed by 16 test trials that presented either IDS or ADS. Trials were presented in one of four pseudorandom orders that counterbalanced the order of presentation of the two stimulus types. Note that within each order, specific IDS and ADS clips were presented adjacently in yoked pairs to facilitate analyses. On each trial, the auditory stimulus played until the infant looked away for 2 consecutive seconds (in labs that implemented an infant-controlled procedure) or until the entire stimulus played, up to 18 s (in labs that implemented a fixed-trial-length procedure). The implementation of the procedure depended on the software that was available in each lab. Trials with less than 2 s of looking were excluded from analyses. Attention-grabbing visual stimuli were presented centrally between trials to reorient infants to the task.

The main differences between the setups were the type and position of visual stimuli presented and their onset relative to the auditory stimuli. For central-fixation and eye-tracking procedures, infants saw a checkerboard on a central monitor, and the presentation of the checkerboard coincided with the onset of the auditory stimulus on each trial. For the HPP, the visual stimulus (either flashing light bulbs or a colorful stimulus on a monitor) was presented silently in the center of the room and then on one side until the infants turned their head toward the side stimulus, at which point the auditory stimulus began playing.

The dependent variable was looking time to the visual stimulus during each trial. For eye-tracking setups, looking time was measured automatically via corneal reflection. For central-fixation and HPP setups, looking time was measured by trained human coders who were blind to trial type, according to the lab's standard procedures.

Parents completed questionnaires about participants' demographic and language background either prior to or after the main experiment.

Results

Analysis overview

Data exclusion. Labs were asked to submit all data collected as part of the bilingual study to the analysis team, and this section focuses on exclusions for the bilingual sample. The initial data set contained 501 infants, of which 333 met each of the inclusion criteria, which are detailed below. We note that exclusions were applied sequentially (i.e., percentages reflect exclusions among the remaining sample after previous criteria were applied).

- *Full term*. We defined full term as gestation times greater than or equal to 37 weeks. There were 9 (1.80%) infants who were tested but excluded because they were preterm.
- *No diagnosed developmental disorders*. We excluded infants whose parents reported developmental disorders (e.g., chromosomal abnormalities) or who were diagnosed with hearing impairments. There were 2 (0.41%) infants who were tested but excluded for these reasons. Because of concerns about the accuracy of parent reports, we did not plan exclusions based on parent-reported ear infections unless parents reported medically confirmed hearing loss.
- *Age*. We included infants in two age groups: 6- to 9-month-olds and 12- to 15-month-olds. There were 58 (11.84%) infants who were tested in the paradigm but who fell outside our target ages. Some labs chose to test such infants for future exploratory analyses, knowing they would be excluded from the current project.
- *Bilingualism*. We excluded infants whose language background did not meet our predefined criteria for bilingualism (see above for details). There were 70 (16.20%) infants whose exposure did not meet this criterion. We also excluded seven (1.93%) additional infants who met this criterion, but who were not learning the community language as one of their languages.
- *Session-level errors*. There were 14 (3.94%) participants excluded because of session-level errors. Seven were excluded for equipment error, three for experimenter error, and four for outside interference.
- *Adequate trials for analysis*. A total of 855 (13.98%) trials were excluded because of errors such as fussiness, presentation of an incorrect stimulus, or a single instance of interference by a parent or sibling.

There was one infant who did not have any trials left for analysis once such trials were excluded. Next, we excluded seven (2.06%) infants who did not have at least one IDS-ADS trial pair available for analysis. For infants with at least one good trial pair, we additionally excluded any trial with less than 2 s of looking ($n = 876$ trials; 16.92% of trials), which was set as a trial-level minimum so that infants had heard enough of the stimulus to discriminate IDS from ADS. As infants did not have to complete the entire experiment to be included, this meant that different infants contributed different numbers of trials. On average, infants contributed 15.70 trials to the analysis.

Data-analysis framework. Our primary dependent variable of interest was looking time, which was defined as the time spent fixating on the visual stimulus during test trials. Given evidence that looking times are nonnormally distributed, we log-transformed all looking times prior to statistical analysis in the mixed-effects model (Csibra et al., 2016). We refer to this transformed variable as "log looking time." For the meta-analysis, we analyzed effect sizes computed from raw difference scores, which did not require log transformation. We preregistered a set of analyses to examine whether monolinguals, heterogeneous samples of bilinguals, and homogeneous samples of bilinguals showed different levels of variability. Unexpectedly, only one lab tested a homogeneous sample of bilinguals, so we deviated from our original plan and did not analyze data as a function of whether our bilingual groups were homogeneous versus heterogeneous. For the main analyses, we adopted two complementary data-analytic frameworks parallel to those used in the ManyBabies 1 project (ManyBabies Consortium, 2020): meta-analysis and mixed-effects regression.

Under the meta-analytic framework, data from each sample of infants (e.g., 6- to 9-month-old bilinguals from Lab 1) were characterized by (a) its effect size (Cohen's *d*) and (b) its standard deviation. Effect-size analyses addressed questions about infants' overall preference for IDS, whereas group-based standard deviation analyses addressed questions about whether some groups of infants showed higher variability in their preference than others. Note that meta-analyses of intragroup variability are relatively rare (Nakagawa et al., 2015; Senior et al., 2016). Unfortunately, our preregistration did not account for the eventuality that several labs would contribute very small numbers of infants to certain groups, as each lab had committed to a minimum sample of 16 infants per group. In two cases, a lab contributed data with a single infant in a particular language group, so it was impossible to compute an effect size. Thus, we implemented a lab-level inclusion criterion for the meta-analysis such that each effect size was computed only if the lab

had contributed data from at least 10 infants in that particular language group and age bin. For example, if lab A had contributed data from seven bilingual 6- to 9-month-old infants and 15 monolingual 6- to 9-month-old infants, we computed the effect size for the monolingual group, but not for the bilingual group. This criterion ensured that each effect size computed was based on a reasonable sample size (i.e., a minimum of 10 infants) and also was consistent with the lab-level inclusion criteria in the ManyBabies 1 study. Because this exclusion criterion was not part of the preregistration, we also ran a robustness analysis with a looser minimum of five infants, which yielded very similar findings (analysis code and results can be found in our GitHub repository).

An advantage of the meta-analytic approach is that it is easy to visualize lab-to-lab differences. Further, the meta-analytic framework most closely mirrors the typical single-lab approach for studying monolingual-bilingual differences, which usually compares groups of monolingual and bilingual infants tested within the same lab. We used this approach specifically to test the overall effect of bilingualism on the magnitude of infants' preference for IDS over ADS and the possible interaction of bilingualism with age. We also compared standard deviations for the bilingual group and monolingual group in a meta-analytic approach. This analysis closely followed Nakagawa et al. (2015).

Under the mixed-effects regression model, trial-by-trial data from each infant were submitted for analysis. Further, independent variables of interest could be specified on an infant-by-infant basis. This approach had the advantage of potentially increasing statistical power, as data were analyzed at a more fine-grained level of detail. As with the meta-analytic approach, this analysis tested the effect of bilingualism and its potential interactions with age. We also investigated whether links between bilingualism and IDS preference were mediated by SES. Additionally, this approach allowed us to assess how the amount of exposure to NAE IDS, measured as a continuous percentage, affected infants' listening preferences. Note that for this analysis, unlike for the meta-analysis, we did not need to apply a lab-level inclusion criterion, which maximized our sample size. Thus, data from all infants who met the infant-level criteria were included in this analysis, which resulted in slightly different sample sizes under the meta-analytic and mixed-effects approaches.

Under both frameworks, we used a dual analysis strategy to investigate how infants' IDS preference was related to bilingualism. First, we examined the lab-matched subset of data (i.e., data from labs that contributed both a monolingual sample and a bilingual sample at a particular age). Second, we examined the complete set of data, including data from labs that contributed

both monolingual and bilingual samples, as well as additional data from labs that tested only monolinguals at the ages of interest as part of the larger ManyBabies 1 project.

Confirmatory analyses

Meta-analytic approach. This approach focused on the analysis of group-level data sets. We defined a data set as data from a group of at least 10 infants who were tested in the same lab, were of the same age (either 6 to 9 or 12 to 15 months old), and had the same language background (monolingual or bilingual). For analyses of within-group variability, we compared bilingual infants with monolingual infants.

To estimate an effect size for each data set, we first computed individual infants' preference for IDS over ADS by (a) subtracting looking time to the ADS stimulus from looking time to the IDS stimulus within each yoked trial pair and (b) then computing a mean difference score for each infant. A total of 20.58% of trials in the lab-matched data set and 13.78% of trials in the full data set were missing. Trial pairs that had either one or both trials missing were excluded from the analysis; 34.53% of pairs in the lab-matched data set and 25.41% of pairs in the full data set were excluded for this reason. Note that we expected many infants to have missing data, particularly on later test trials, given the length of the study (16 test trials). Then, for each data set, we calculated the mean of these difference scores (M_d) and its associated standard deviation across participants. Finally, we used the derived means and standard deviation to compute a within-subjects Cohen's *d* using the formula $d_z = M_d/SD$.

In the following meta-analyses, random-effects meta-analysis models with a restricted maximum likelihood estimator were fit with the *metafor* package (Viechtbauer, 2010). To account for the dependence between monolingual and bilingual data sets stemming from the same lab, we added laboratory as a random factor. As part of our preregistered analyses, we planned to include method as a moderator in this analysis if it was found to be a statistically significant moderator in the larger ManyBabies 1 project—which it was (ManyBabies Consortium, 2020). However, only 17 labs contributed data from bilinguals, and we deviated from this plan because of the small number of labs per method (e.g., only three labs used a central-fixation method).

Effect-size-based meta-analysis. Our first set of meta-analyses focused on effect sizes (d_z): how our variables of interest contributed to differences in looking time on IDS versus ADS trials. Recall that we ran the analyses in two ways: The first analysis was restricted to the labs that contributed lab-matched data (lab-matched data set), and the second analysis included all available data (i.e., including data from labs that tested only monolinguals or only bilinguals at the ages of interest; full data set).

We initially fit the following model to examine contributions of age and bilingualism to infants' IDS preference, as well as potential interactions between these variables:

$$d_z \sim 1 + \text{bilingual} + \text{age} + \text{bilingual} * \text{age}$$

Bilingualism was dummy coded (0 = monolingual, 1 = bilingual), and age (a continuous variable) was coded as the average age for each lab's contributed sample for each language group (centered for ease of interpretation).

In the lab-matched data set, we did not find statistically significant effects of age ($d_z = 0.17$, 95% confidence interval [CI] = [−1.01, 1.36], $z = 0.29$, $p = .775$) or bilingualism ($d_z = −0.17$, 95% CI = [−0.44, 0.10], $z = −1.22$, $p = .224$), or a significant interaction between age and bilingualism ($d_z = −0.19$, 95% CI = [−1.84, 1.46], $z = −0.22$, $p = .822$).

Similarly, in the full data set, we did not find significant main effects of age ($d_z = 0.01$, 95% CI = [−0.66, 0.67], $z = 0.02$, $p = .984$) or bilingualism ($d_z = −0.10$, 95% CI = [−0.29, 0.09], $z = −1.03$, $p = .301$), or a significant interaction between age and bilingualism ($d_z = 0.01$, 95% CI = [−0.93, 0.95], $z = 0.03$, $p = .979$).

As bilingualism is the key moderator of research interest in the current article, here we report the effect sizes of monolingual and bilingual infants separately. In the lab-matched data set, the effect size for monolinguals was 0.42 (95% CI = [0.21, 0.63], $z = 3.94$, $p < .001$), and for bilinguals the effect size was 0.24 (95% CI = [0.06, 0.42], $z = 2.64$, $p = .008$). In the full data set, the effect size for monolinguals was 0.36 (95% CI = [0.28, 0.44], $z = 9.15$, $p < .001$), and for bilinguals the effect size was 0.26 (95% CI = [0.09, 0.43], $z = 2.97$, $p = .003$). In sum, numerically, monolinguals showed a stronger preference for IDS than bilinguals, but this tendency was not statistically significant in the effect-size-based meta-analyses. A forest plot for the lab-matched meta-analysis is shown in Figure 1.

Within-group-variability meta-analysis. Our second set of preregistered meta-analyses examined whether the variability in infants' preference for IDS within a sample (within-sample variability) was related to language background (monolingual vs. bilingual). Note that this question of within-sample heterogeneity is different from questions of between-samples heterogeneity that can also be addressed in meta-analysis (see Higgins et al., 2003; Higgins & Thompson, 2002 for approaches to between-groups variability in meta-analysis). Specifically, the within-group-variability meta-analysis approach provides additional insights into how two
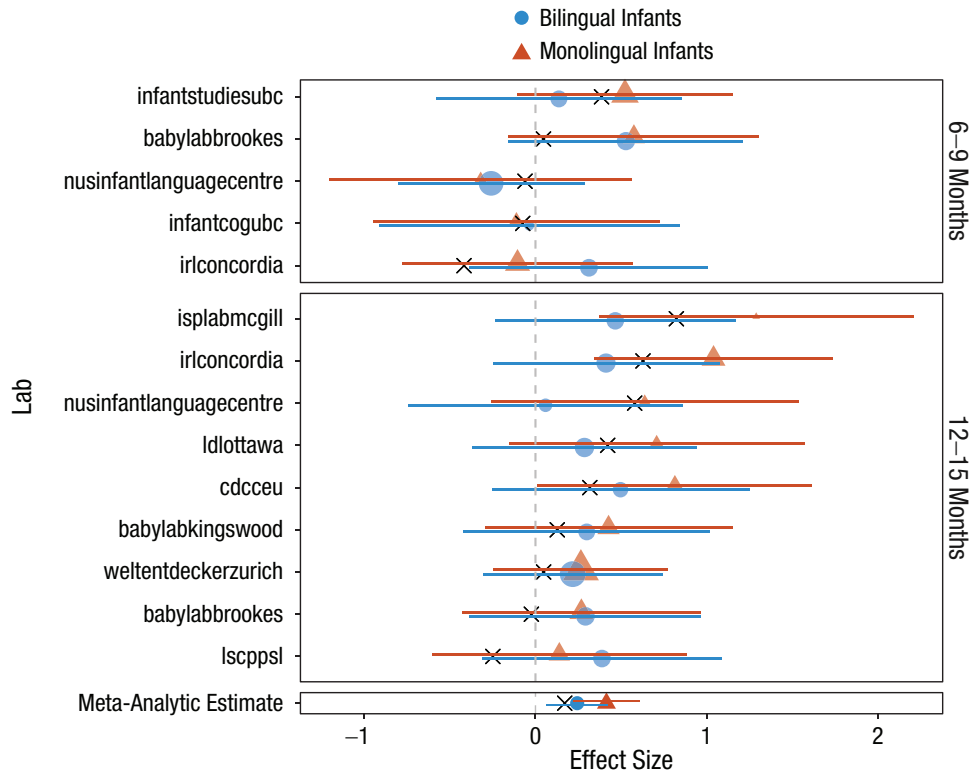
Fig. 1. Forest plot for the lab-matched data set, separated by age group. Standardized effect sizes are shown for each lab; error bars show 95% confidence intervals. Each lab reported two effect sizes: one for the monolingual group and the other one for the bilingual group. The plotted points are scaled by inverse variance (i.e., more precise estimates are denoted by larger shapes). Within each age group, labs are ordered by the difference between the monolingual and bilingual effect sizes, and this effect-size difference is indicated by a black X. The bottom row shows the global meta-analytic estimate.

groups differ in terms of their variances, not merely their mean effect sizes. This approach is useful to investigate whether infants' language backgrounds influence not only the magnitude of infants' IDS preference, but also the variability of their IDS preference. In the following, the standard deviations measure variability of infants' looking-time preference for IDS over ADS in each language group. Again, we report d_z, an effect size that measures the magnitude of infants' preference for IDS over ADS.

Our preregistered plan was to follow Nakagawa et al. (2015) and Senior et al. (2016), and we further elaborate on this plan here. According to Nakagawa et al., there are two approaches to run within-group-variability meta-analysis: One approach uses *lnCVR*, the natural logarithm of the ratio between the coefficients of variation, to compare the variability of two groups; a second approach enters *lnSD* (the natural logarithm of standard deviations) and $ln\bar{X}$ (the log mean) into a mixed-effects model. When data meet the assumption that the standard deviation is proportional to the mean (i.e., the two are correlated), the first approach should be used, and otherwise, the second approach should be used. Our data did not meet the necessary assumption; therefore, we

used the second, mixed-effects approach. In the following metaregression model, the natural logarithm of the standard deviation (*lnSD*) of IDS preference in each language group is the dependent variable:

$$lnSD \sim 1 + \text{bilingual} + ln\left(|d_z|\right) + \left(\text{bilingual}|\text{lab}\right),$$

where $|d_z|$ is the absolute value of d_z because we needed to ensure that values entered into the logarithm were positive, and bilingual is the binary dummy variable that indicates whether the language group is monolingual or bilingual. Further, we entered a random intercept and a random slope for bilingualism, which were allowed to vary by lab. We note that this log transformation is entirely unrelated to the log transformation of raw looking times used in the linear mixed-effects models.

In the lab-matched data set, we did not find statistically significant evidence for bilingualism as a moderator of the difference between the two language groups' standard deviations ($d_z = -0.08$, $p = .235$). We also did not find significant evidence for such an effect in the full data set ($d_z = 0.02$, $p = .698$). In short, we did not find

support for the possibility that bilingual infants would show larger within-group variability than monolingual infants.

Mixed-effects approach. Mixed-effects regression allows variables of interest to be specified on a trial-by-trial and infant-by-infant basis. We had anticipated that we would be able to include additional data from labs that aimed to test homogeneous samples (i.e., because we could include infants from these labs who were not learning the homogeneous sample's language pair), but in practice this did not happen, as only one lab contributed a homogeneous data set, and that lab did not test additional infants. We were able to include data from all valid trials, rather than excluding data from yoked pairs with a missing data point as was necessary for the meta-analysis. As under the meta-analytic approach, we ran the models twice, once including only data from labs that contributed lab-matched samples of monolinguals and bilinguals and once including all available data from 6- to 9-month-olds and 12- to 15-month-olds.

The mixed-effects model was specified as follows:

$$\text{DV} \sim \text{IV}_1 + \text{IV}_2 + \ldots + (\ldots|\text{subject}) + (\ldots|\text{item}) + (\ldots|\text{lab})$$

The goal of this framework was to examine effects of the independent variables (IV) on the dependent variable (DV), while controlling for variation in both the DV (random intercepts) and the relationship of the IV to the DV (random slopes) based on relevant grouping units (subjects, items, and labs). Following recent recommendations (Barr et al., 2013), we planned to initially fit a maximal random-effects structure, such that all random effects appropriate for our design were included in the model. However, we also recognized that such a large random-effects structure might be overly complex given our data and would be unlikely to converge. After reviewer feedback during Stage 1 of the Registered Report review process, we preregistered a plan to use a parsimonious mixed-models approach for pruning the random effects (Bates et al., 2018; Matuschek et al., 2017). However, we found that it was computationally difficult to first fit complex models (i.e., our models had multiple interactions and cross-levels grouping) under the maximal random-effects structure and then prune the models using a parsimonious mixed-models approach. Further, we note that this was not the approach used in ManyBabies 1 (ManyBabies Consortium, 2020), which would make a direct comparison between ManyBabies 1 and the current study difficult. Therefore, following ManyBabies 1, we fitted and pruned the models we present in this section using the maximal random-effects structure only (Barr et al., 2013). We fitted all models using the *lme4* package (Bates et al., 2015) and computed *p* values using the *lmerTest* package (Kuznetsova

et al., 2016). All steps of the pruning process we followed are detailed in the analytic code on our GitHub repository. Following a reviewer's suggestion during Stage 2 review, we checked our models for potential issues with multicollinearity by examining variance inflation factors (VIFs) for each model. Variables that have VIF values exceeding 10 are regarded as violating the multicollinearity assumption (Curto & Pinto, 2011). None of our models violated this assumption. Below is a description of our variables for the mixed-effects models:

- log looking time: the dependent variable; log-transformed looking time in seconds
- trial type: a dummy-coded variable with two levels, with ADS trials as the baseline, such that positive effects of trial type indicate longer looking to IDS
- bilingual: a dummy-coded variable with two levels, with monolingual as the baseline, such that positive effects of bilingualism reflect longer looking by bilinguals
- language: a dummy-coded variable with two levels, with NAE learners as the baseline; NAE learners (i.e., infants learning NAE as a native language) were defined as monolinguals with at least 90% exposure to NAE and bilinguals with at least 25% exposure to NAE
- nae exposure: a continuous variable for the percentage of time infants heard NAE
- method: a dummy-coded variable to control for effects of different experimental setups, with central fixation as the reference level
- age: age in days, centered for interpretability of main effects
- trial number: the number of the trial pair, recoded such that the first trial pair was 0
- ses: the number of years of maternal education, centered for ease of interpretation

Note that in this analysis plan, we used a concise format for model specification, which is the form used in R. Thus, lower-order effects subsumed by interactions were modeled even though they were not explicitly written. For example, the interaction of trial type and trial number also assumes a global intercept, a main effect of trial type, and a main effect of trial number.

Homogeneity of variance. We preregistered a Levene's test to examine whether monolinguals and bilinguals showed different amounts of variance in their IDS preference. Our analysis focused on the residual variance for monolinguals and bilinguals in the main linear mixed-effects models, in order to partition out variance associated with other factors (e.g., age, method). The Levene's

test revealed a statistically significant difference in variance between monolinguals and bilinguals for the full data set ($p = .02$) but not the lab-matched data set ($p = .68$). We note that the difference in residual variance between the monolingual (variance = 0.24) and bilingual (variance = 0.25) language groups was small, which suggests that the statistically significant Levene's test for the full data set was mainly driven by its larger sample size, rather than by a meaningful difference between monolinguals and bilinguals.

Effects of bilingualism on IDS preference. We planned a mixed-effects model that was based on the structure of the final model fitted for the ManyBabies project, including bilingualism as an additional moderator. Note that because data collection for the two projects was simultaneous, we did not know prior to registration what the final model structure for the monolingual-only sample would be (it was expected that pruning of this model would be necessary in the case of nonconvergence). The original model proposed for the monolingual-only sample was designed to include simple effects of trial type, method, language (infants exposed vs. not exposed to NAE IDS), age, and trial number, capturing the basic effect of each parameter on looking time (e.g., longer looking times for IDS, shorter looking times on later trials). Additionally, the model included two-way interactions of trial type with method and with trial number, a two-way interaction of age with trial number, and two- and three-way interactions of trial type, age, and language (see ManyBabies Consortium, 2020, for full justification). This model was specified to minimize higher-order interactions while preserving theoretically important interactions. Note that to reduce model complexity, we treated both developmental effects and trial effects linearly. The planned initial model was

log looking time ~ trial type * method + trial type * trial number + age * trial number + trial type * age * language + (trial type * trial number|participant) + (trial type * age|lab) + (method + age * language|item)

Our analysis plan specified that we would add bilingualism to the fixed effects of the final pruned model fitted to the monolingual sample. For higher-order interactions in the model, we ensured that we had at least 20 infants per group. For example, we did not include a three-way interaction of bilingualism, language, and age because we had fewer than twenty 6- to 9-month-old bilinguals who were not exposed to NAE.

In our preregistration, we were uncertain as to whether our sample size would support a model with a four-way interaction of trial type, age, bilingual status, and language. Given our final sample size, we elected

to fit our main model without including the four-way interaction effect. In our main model, we included two fixed three-way interactions—(a) the interaction of bilingualism, age, and trial type and (b) the interaction of language, age, and trial type—as well as other subsumed lower-order interactions.

Regardless of our fixed-effects structure, the model included the random slope for the effect of bilingualism on lab and item, as well as appropriate interactions with other random factors. Our initial unpruned model was

log looking time ~ trial type * method + trial type * trial number + age * trial number + trial type * age * language + trial type * age * bilingual + (trial type * trial number|participant) + (trial type * age * bilingual|lab) + (method + age * language + age * bilingual|item)

After pruning of random effects for nonconvergence and singularity, the final models for the lab-matched data set and full data set were different. The following was the final model for the lab-matched data set:

log looking time ~ trial type * method + trial type * trial number + age * trial number + trial type * age * language + trial type * age * bilingual + (1|participant) + (bilingual|lab) + (1|item)

In contrast, the final model for the full data set was

log looking time ~ trial type * method + trial type * trial number + age * trial number + trial type * age * language + trial type * age * bilingual + (1|participant) + (1|lab) + (1|item)

Overall, the mixed-level analyses in the lab-matched and the full data sets yielded similar results (Tables 2 and 3). More coefficients were statistically significant in the full data set, likely because of the larger sample size. Thus, in the following, we focus on the results of the mixed-level model for the full data set. We found that infants showed a preference for IDS, as indicated by a positive coefficient on the IDS predictor (reflecting greater looking times to IDS stimuli). We did not find an effect of bilingualism on IDS preference or any interaction of bilingualism and other moderators. This finding is consistent with the results of our meta-analysis above.

Surprisingly, the fitted model did not show an interaction between infants' IDS preference and the method used in the lab, a result that is different from the results

Table 2. Linear Mixed-Effects Model Testing the Effect of Bilingualism on Log Looking Time in the Lab-Matched Data Set (*n* = 717)

Predictor	Estimate	SE	t	p
Intercept	1.930	0.074	26.000	< .001
IDS	0.093	0.047	2.000	.050
HPP	0.103	0.092	1.110	.283
Central fixation	0.113	0.103	1.090	.288
Age	−0.027	0.008	−3.410	.001
Trial number	−0.036	0.003	−13.900	< .001
Language	−0.059	0.075	−0.792	.435
Bilingual	0.0003	0.034	0.008	.994
IDS * HPP	0.016	0.029	0.566	.571
IDS * Central Fixation	0.004	0.031	0.124	.901
Age * Trial Number	0.001	0.0004	2.270	.023
IDS * Trial Number	0.001	0.004	0.174	.862
IDS * Age	0.013	0.006	2.180	.029
IDS * Language	0.051	0.026	1.950	.052
Age * Language	0.007	0.010	0.646	.519
IDS * Bilingual	−0.012	0.024	−0.522	.602
Age * Bilingual	−0.006	0.009	−0.671	.503
IDS * Age * Language	0.016	0.008	1.860	.063
IDS * Age * Bilingual	−0.009	0.008	−1.210	.227

Note: The proportion of variance accounted for by the fixed effects only (marginal R^2) was .087, and the proportion of variance accounted for by both the fixed and the random effects (conditional R^2) was .317. IDS = infant-directed speech; HPP = head-turn-preference procedure.

in the ManyBabies 1 project. However, this finding is likely due to smaller sample sizes in the current project, as we restricted the analysis to participants at particular ages. Apart from this, our findings were largely consistent with those of the ManyBabies 1 study. There was a significant and positive two-way interaction between IDS and language suggesting greater IDS preferences for children in NAE contexts than for children in other language contexts. The interaction between IDS and age was also significant and positive, suggesting that older children showed a stronger IDS preference. Finally, we found a marginally significant three-way interaction of IDS, age, and NAE, suggesting that older children in NAE contexts tended to show stronger IDS preference than those in the non-NAE contexts.

Dose effects of exposure to NAE IDS in bilingual infants. In this analysis, we tested whether we could observe a dose-response relationship between infants' exposure to NAE IDS (measured continuously) and their preference for IDS over ADS.

We decided to conduct this analysis including only data from bilinguals. Our reasoning was that bilingualism status and exposure to NAE IDS were confounded, as

monolinguals' exposure to NAE was near either 0% or 100%, whereas bilinguals' NAE experience could be either 0% (because not all bilinguals were learning NAE as one of their two languages) or in the range from 25% to 75%. Because the monolingual sample was larger and their NAE exposure was more extreme, their effects would dominate those of the bilinguals in a merged analysis. Therefore, we reasoned that if there was a dose effect, it should be observable in the bilingual sample alone. Finally, although excluding monolingual infants reduced power overall, we decided that given the relatively large sample of bilingual infants, this disadvantage would be offset by the ease of interpretation afforded by restricting the analysis to bilinguals. On average, bilingual infants in our sample were exposed to 20.17% NAE (range: 0%–75%).

Once again, we based this model on the final pruned monolingual model, replacing the binary measure of exposure to NAE IDS (language) with the continuous measure of exposure (nae exposure), and including a random slope for NAE exposure by item (which was ultimately pruned from the model). After pruning, our model was specified as follows:

Table 3. Linear Mixed-Effects Model Testing the Effect of Bilingualism on Log Looking Time in the Full Data Set (*N* = 1,754)

Predictor	Estimate	SE	t	p
Intercept	1.890	0.047	40.400	< .001
IDS	0.106	0.038	2.770	.009
HPP	0.190	0.058	3.310	.002
Central fixation	0.243	0.054	4.510	< .001
Age	−0.029	0.005	−5.680	< .001
Trial number	−0.037	0.002	−21.200	< .001
Language	0.003	0.048	0.063	.950
Bilingual	−0.006	0.025	−0.234	.815
IDS * HPP	0.029	0.018	1.620	.106
IDS * Central Fixation	−0.020	0.019	−1.060	.291
Age * Trial Number	0.001	0.0003	3.910	< .001
IDS * Trial Number	−0.002	0.002	−0.961	.337
IDS * Age	0.013	0.003	3.800	< .001
IDS * Language	0.038	0.016	2.420	.015
Age * Language	0.002	0.007	0.244	.807
IDS * Bilingual	0.003	0.019	0.142	.887
Age * Bilingual	−0.003	0.008	−0.369	.712
IDS * Age * Language	0.009	0.005	1.960	.051
IDS * Age * Bilingual	−0.007	0.006	−1.110	.265

Note: The proportion of variance accounted for by the fixed effects only (marginal R^2) was .110, and the proportion of variance accounted for by both the fixed and the random effects (conditional R^2) was .361. IDS = infant-directed speech; HPP = head-turn-preference procedure.

Table 4. Linear Mixed-Effects Model Testing the Effects of Exposure to NAE IDS on Log Looking Time in Bilingual Infants ($n = 333$)

Predictor	Estimate	SE	t	p
Intercept	1.910	0.074	25.900	< .001
IDS	−0.009	0.062	−0.138	.891
HPP	0.088	0.091	0.963	.353
Central fixation	0.168	0.111	1.510	.160
Age	−0.024	0.010	−2.270	.024
Trial number	−0.036	0.004	−10.100	< .001
NAE exposure	−0.001	0.001	−0.565	.575
IDS * HPP	0.054	0.053	1.020	.331
IDS * Central Fixation	0.028	0.060	0.465	.654
Age * Trial Number	0.0002	0.001	0.300	.764
IDS * Trial Number	0.006	0.005	1.150	.251
IDS * Age	0.006	0.008	0.781	.435
IDS * NAE Exposure	0.002	0.001	2.860	.011
Age * NAE Exposure	−0.00005	0.0003	−0.200	.842
IDS * Age * NAE Exposure	0.0002	0.0002	0.891	.373

Note: The proportion of variance accounted for by the fixed effects only (marginal R^2) was .089, and the proportion of variance accounted for by both the fixed and the random effects (conditional R^2) was .318. IDS = infant-directed speech; HPP = head-turn-preference procedure; NAE = North American English.

log looking time ~ trial type * method + trial type * trial number + age * trial number + trial type * age * nae exposure + (1|participant) + (trial type|lab) + (1|item)

Table 4 contains the details of the results in this model. The main effect of infants' exposure to NAE was not significant (β = −0.0007, *SE* = 0.001, *p* = .57). This indicates that bilingual infants who were exposed to more NAE did not pay more attention to the NAE speech stimuli than did those who were exposed to less NAE. However, the interaction between trial type and exposure to NAE was significant (β = 0.002, *SE* = 0.0008, *p* = .011). That is, bilingual infants who were exposed to more NAE showed stronger IDS preferences. This result confirmed a dose-response relationship between infants' exposure to NAE and their preference for IDS over ADS (Fig. 2) even among bilinguals who were learning NAE as one of their native languages.

Socioeconomic status as a moderator of monolingual-bilingual differences. Because SES can vary systematically between monolinguals and bilinguals in the same community, we preregistered an analysis controlling for SES. Given that we did not find an effect of bilingualism on IDS preference, this analysis investigated whether such an effect would be apparent once SES was controlled. Alternately, if we had observed an effect of bilingualism on IDS preference, this analysis would have allowed us to investigate whether it would disappear once SES was controlled.

Thus, this analysis was relevant regardless of whether or not we observed a relationship between IDS preference and bilingualism in the previous model.

First, we computed descriptive statistics for the two groups. Mothers of the bilingual sample had an average of 16.71 years of education (*SD* = 2.47, range = 10–26), those of the lab-matched monolingual sample had an average of 16.33 years of education (*SD* = 2.83, range = 5–28), and those of the full monolingual sample had an average of 16.42 years of education (*SD* = 2.47, range = 8–25).

Our approach was to add SES as a moderator in our final model for bilinguals. We expected that any effects of SES could interact with age. Thus, this model included interactions of trial type, age, and SES as a fixed effect, as well as the corresponding random slope by item. Based on the potential model detailed above for the bilinguals, our expected SES-mediated model was

log looking time ~ trial type * method + trial type * trial number + age * trial number + trial type * age * language + trial type * age * bilingual + trial type * age * ses + (trial type * trial number|participant) + (trial type * age * bilingual|lab) + (method + age * language + age * bilingual + age * ses|item)

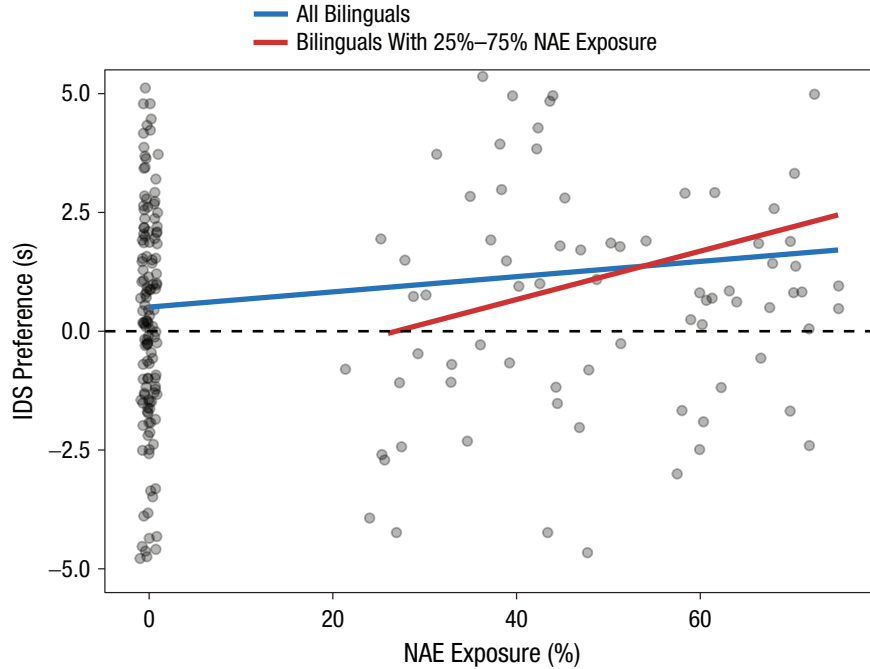After pruning for nonconvergence, our final model specification for the lab-matched data set was

Fig. 2. Scatterplot showing the relationship between bilingual infants' preference for infant-directed speech (IDS) and their percentage of exposure to North American English (NAE). The solid lines show the results of linear regression models, one starting from zero NAE exposure and another focusing on NAE exposure from 25% to 75%. Note that the *y*-axis was truncated to highlight the trend, and some individual points are not plotted.

log looking time ~ trial type * method + trial type * trial number + age * trial number + trial type * age * language + trial type * age * bilingual + trial type * age * ses + (1|participant) + (bilingual|lab)

By contrast, the final model of the full data set was

log looking time ~ trial type * method + trial type * trial number + age * trial number + trial type * age * language + trial type * age * bilingual + trial type * age * ses + (1|participant) + (1|lab) + (1|item)

In general, across the lab-matched and full data sets (Tables 5 and 6), SES did not have a significant effect on infants' looking time, nor did it affect infants' preference for IDS. However, for the lab-matched data set only, we found a statistically significant three-way interaction of IDS, age, and SES. Specifically, 6- to 9-month-olds showed stronger IDS preference when they were from higher-SES families, but older infants, 12- to 15-month-olds, showed similar IDS preference across families with different SES levels. However, this interaction was not observed in the full data set, which raises the possibility that it is spurious, and arose only in the lab-matched data set because that data set is substantially smaller than the full data set.

Exploratory analysis

Our confirmatory analysis showed that bilingual infants with more exposure to NAE showed stronger IDS preference (see Table 4). However, this initial analysis included a number of bilingual infants who were not exposed to NAE at all. This raises the question of whether the relation between NAE exposure and IDS preference was primarily driven by the infants who were not learning NAE. In the following analysis, we reran the preregistered model of the relationship between NAE exposure and IDS preference, this time restricting the model to infants ($n = 135$) who were exposed to NAE between 25% and 75% of the time. After pruning for nonconvergence, the final model was

log looking time ~ trial type * method + trial type * trial number + age * trial number + trial type * age * nae exposure + (1|participant) + (1|lab) + (1|item)

The interaction between IDS preference and NAE exposure was still statistically significant ($\beta = 0.01$, *SE* = 0.002, *p* = .004; see Fig. 2). This result suggested that a dose-response relationship between infants' exposure to NAE and their preference for IDS over ADS was not driven by infants living in non-NAE contexts alone (see Table 7 for details of the model).

Table 5. Linear Mixed-Effects Model Examining Socioeconomic Status (SES) as a Moderator of Monolingual-Bilingual Differences in Log Looking Time in the Lab-Matched Data Set (*n* = 717)

Predictor	Estimate	*SE*	*t*	*p*
Intercept	1.910	0.066	28.800	< .001
IDS	0.133	0.033	4.060	< .001
HPP	0.120	0.089	1.340	.199
Central fixation	0.094	0.100	0.939	.359
Age	−0.029	0.008	−3.590	< .001
Trial number	−0.033	0.002	−17.200	< .001
Language	−0.089	0.072	−1.240	.225
Bilingual	0.022	0.028	0.795	.427
SES	−0.003	0.005	−0.513	.608
IDS * HPP	0.019	0.030	0.633	.527
IDS * Central Fixation	0.006	0.032	0.201	.841
Age * Trial Number	0.001	0.0004	2.330	.020
IDS * Trial Number	−0.005	0.003	−1.740	.081
IDS * Age	0.012	0.006	1.920	.055
IDS * Language	0.054	0.028	1.960	.050
Age * Language	0.012	0.011	1.130	.260
IDS * Bilingual	−0.018	0.025	−0.734	.463
Age * Bilingual	−0.011	0.009	−1.160	.246
IDS * SES	0.003	0.005	0.770	.441
Age * SES	−0.0002	0.002	−0.147	.883
IDS * Age * Language	0.016	0.009	1.810	.071
IDS * Age * Bilingual	−0.005	0.008	−0.606	.545
IDS * Age * SES	−0.004	0.002	−2.330	.020

Note: The proportion of variance accounted for by the fixed effects only (marginal R^2) was .088, and the proportion of variance accounted for by both the fixed and the random effects (conditional R^2) was .304. IDS = infant-directed speech; HPP = head-turn-preference procedure.

General Discussion

The current study was designed to shed light on the effects of experience on the tuning of infants' preference for IDS compared with ADS. Bilingual infants' language experience is split across input in two different languages, which are being acquired simultaneously. Bilinguals and monolinguals may thus differ in their preference for IDS. To explore this question, we used a collaborative, multilab (*N* = 17 labs) approach to gather a large data set from infants who were either 6 to 9 or 12 to 15 months old and growing up bilingual (*N* = 333 bilingual infants in the final sample) and a lab-matched sample of 384 monolingual infants from the same communities. Data were collected as a companion project to ManyBabies 1 (ManyBabies Consortium, 2020), which was limited to infants growing up monolingual. Overall, we found that bilingualism neither enhanced nor attenuated infants' preference for IDS; the magnitude and developmental trajectory of IDS preference was similar for bilinguals and monolinguals from age 6 to 15 months.

Although bilingual experience did not appear to moderate infants' preference for IDS, we found striking evidence that experience hearing NAE, the language of our stimuli, contributed to the magnitude of bilingual infants' IDS preference. Bilinguals with greater exposure to NAE showed greater IDS preference (when tested in NAE) than those who had less exposure to NAE. This relationship between NAE exposure and IDS preference was also observed in a subsample of bilingual infants who were all acquiring NAE, but who varied in how much they were exposed to NAE relative to their other native language. These results converge with those from the larger ManyBabies 1 study, which found that monolinguals acquiring NAE had a stronger preference for IDS than monolinguals acquiring another language. Our approach provides a more nuanced view of the relationship between NAE exposure and IDS preference and suggests that there is a continuous dose effect of exposure on preference. Together, our findings have a number of implications for bilingual language acquisition during infancy. In the following, we discuss each of our research questions in turn, followed by limitations and implications of our study.

Our first research question asked whether bilingualism affects infants' relative attention to IDS versus ADS.

Table 6. Linear Mixed-Effects Model Examining Socioeconomic Status (SES) as a Moderator of Monolingual-Bilingual Differences in Log Looking Time in the Full Data Set (*N* = 1,754)

Predictor	Estimate	*SE*	*t*	*p*
Intercept	1.930	0.052	37.000	< .001
IDS	0.114	0.041	2.780	.009
HPP	0.189	0.063	2.990	.004
Central fixation	0.202	0.064	3.170	.003
Age	−0.036	0.006	−6.300	< .001
Trial number	−0.037	0.002	−19.500	< .001
Language	−0.018	0.051	−0.363	.718
Bilingual	0.003	0.026	0.109	.913
SES	−0.001	0.004	−0.204	.838
IDS * HPP	0.029	0.020	1.410	.160
IDS * Central Fixation	−0.022	0.021	−1.040	.296
Age * Trial Number	0.001	0.0003	4.280	< .001
IDS * Trial Number	−0.003	0.003	−0.949	.343
IDS * Age	0.011	0.004	2.940	.003
IDS * Language	0.031	0.017	1.800	.072
Age * Language	0.003	0.007	0.443	.657
IDS * Bilingual	−0.007	0.020	−0.336	.737
Age * Bilingual	−0.002	0.008	−0.206	.837
IDS * SES	0.004	0.003	1.220	.222
Age * SES	−0.001	0.001	−0.781	.435
IDS * Age * Language	0.012	0.005	2.230	.026
IDS * Age * Bilingual	−0.004	0.007	−0.597	.550
IDS * Age * SES	−0.001	0.001	−0.599	.549

Note: The proportion of variance accounted for by the fixed effects only (marginal R^2) was .109, and the proportion of variance accounted for by both the fixed and the random effects (conditional R^2) was .349. IDS = infant-directed speech; HPP = head-turn-preference procedure.

One possibility we raised was that the complexity of the bilingual infant's learning experience might lead to greater reliance on IDS, given that IDS may be viewed as an enhanced linguistic signal. However, our data were not consistent with this possibility. In the full data set, monolinguals showed a numerically larger meta-analytic effect size, $d_z = 0.36$ [95% CI = 0.28, 0.44], than bilinguals did, $d_z = 0.26$ [95% CI = 0.09, 0.43], but this difference was not statistically significant in either the meta-analyses or the mixed-effects linear models. Although small differences are still possible, our data generally support the conclusion that bilingual and monolingual infants show a similar preference for IDS over ADS. Specifically, both groups show a preference for IDS at 6 to 9 months of age, and this preference gets stronger by the age of 12 to 15 months.

An additional part of our first research question asked whether bilinguals might show more variability than monolinguals in their IDS preference, beyond any differences in the magnitude of the preference. We reasoned that given their diversity of language experiences, bilingual groups may have a higher heterogeneity in terms of their IDS preference compared with monolingual groups (see also Orena & Polka, 2019, for a recent experiment that demonstrated this pattern). Both monolingual and bilingual groups showed high variability. The magnitude of the observed difference in variability was very small. We conducted three analyses to compare the variability between the monolinguals and bilinguals. Only one of the three variability analyses (i.e., the Levene's test with the full data set) was statistically significant. This effect was mainly driven by the large sample size in the full data set (*N* = 1,754) because the difference in variability between the monolinguals and bilinguals remained negligible. Thus, our results do not support the idea that bilingual infants show meaningfully more variability in their IDS preference than their monolingual peers.

Given that monolinguals and bilinguals can differ systematically in their SES, the third part of our first research question asked whether SES might moderate the effects of bilingualism. Using years of maternal education as a proxy for SES, we found mixed support for SES as a moderator in our data sets. In our smaller lab-matched

Table 7. Linear Mixed-Effects Model Testing the Effects of Exposure to NAE IDS on Log Looking Time Among Bilingual Infants Living in NAE Contexts ($n = 135$)

Predictor	Estimate	SE	t	p
Intercept	1.910	0.168	11.400	< .001
IDS	−0.211	0.132	−1.600	.112
HPP	0.227	0.142	1.600	.180
Central fixation	0.094	0.200	0.472	.663
Age	−0.009	0.035	−0.265	.791
Trial Number	−0.041	0.006	−7.410	< .001
NAE exposure	−0.002	0.002	−0.783	.434
IDS * HPP	0.016	0.063	0.260	.795
IDS * Central Fixation	−0.115	0.081	−1.420	.156
Age * Trial Number	0.001	0.001	1.230	.219
IDS * Trial Number	0.016	0.008	1.990	.048
IDS * Age	0.022	0.030	0.720	.472
IDS * NAE Exposure	0.005	0.002	2.900	.004
Age * NAE Exposure	−0.0004	0.001	−0.653	.515
IDS * Age * NAE Exposure	−0.00003	0.001	0.054	.957

Note: The proportion of variance accounted for by the fixed effects only (marginal R^2) was .119, and the proportion of variance accounted for by both the fixed and the random effects (conditional R^2) was .362. IDS = infant-directed speech; HPP = head-turn-preference procedure; NAE = North American English.

data set, we found a statistically significant interaction of age, SES, and IDS preference: Six- to 9-month-olds from higher-SES families showed stronger IDS preference than those from lower-SES families, whereas 12- to 15-month-olds showed similar IDS preference regardless of SES. The direction of this effect aligns with other research indicating that children from higher-SES families generally receive more language input and/or higher-quality input (e.g., by engaging in conversations with more lexical diversity, complexity, and structural variations) than children from lower-SES families (Fernald et al., 2013; Hart & Risley, 1995; Hoff, 2006; Tal & Arnon, 2018). This could suggest that infants from higher-SES families may show stronger IDS preference early in life because they hear more or higher-quality IDS in their daily lives. Further, this positive impact of SES may be most beneficial to younger infants, whose IDS preference is still developing. However, given that in our larger (full) data set SES was unrelated to IDS preference in either 6- to 9- or 12- to 15-month-olds, this result might be spurious and should be interpreted with caution. Further, it is important to note that our samples (both monolingual and bilingual groups) were mainly from higher-SES families. Indeed, in the lab-matched data set, the mothers of 67.79% of children had earned at least a bachelor's degree. Our samples, therefore, had low variability in infants' SES. Thus, this question would be better tested with future studies that have participants from more diverse SES backgrounds.

Our second research question asked whether and how the amount of exposure to NAE affects bilingual infants' listening preferences. Given that our stimuli were produced in NAE, one possibility was that greater exposure to NAE would be linked to greater attention to NAE IDS relative to NAE ADS. Indeed, ManyBabies 1 (ManyBabies Consortium, 2020), which was conducted concurrently with the current study, found that monolinguals acquiring NAE showed a stronger IDS preference than monolinguals not acquiring NAE. However, in the ManyBabies 1 study, exposure to NAE IDS was a binary variable—the infants either heard only NAE or heard only a different language in their environments. In the current project, bilinguals provided a more nuanced way to address this question, as bilinguals' exposure to NAE varied continuously between 25% and 75% (for infants learning NAE as one of their native languages) or was near 0% (for infants learning two non-NAE native languages). We found clear evidence for a positive dose-response relationship between exposure to NAE and infants' preference for NAE IDS. This evidence—that bilinguals with more exposure to NAE showed a stronger NAE IDS preference—was also present when we focused only on bilinguals who were learning NAE as one of their native languages (i.e., those exposed to NAE 25%–75% of the time). Importantly, we did not find a similar effect of exposure to NAE on infants' overall looking time. This suggests that the effect of NAE exposure on preference for IDS is a meaningful relationship, rather than an

artifact due to the stimuli being presented in NAE. Further studies with stimuli in other languages would be necessary to solidify this conclusion.

Our analyses included both meta-analyses and linear mixed-effects models, which allowed us to compare these two approaches. As our field moves toward more large-scale studies of this type, it will be important to determine appropriate standards for analysis. Our meta-analysis allows for better and more direct comparison with prior meta-analyses (e.g., Dunst et al., 2012). However, an important limitation of this approach is that infants' data are collapsed to a single averaged data point per group, which obscures potentially interesting variability. Moreover, because we could not model trial number directly, this average was based on valid adjacent trial pairs, which resulted in many trials being excluded from the analysis. In contrast, the mixed-effects models analyzed data at the individual-trial level, allowing us to examine how data variability could be explained by moderators at the trial and participant level, which increased statistical power. Our finding of a significant effect of age on IDS preference in the mixed models, but not in the meta-analysis, can be attributed to this difference in statistical power. We believe that each of these complementary approaches has its place, but that the mixed-effects model is preferable because it improves statistical power.

Given that this was the first study to recruit and test bilingual infants at such a large scale and at so many sites, we encountered several challenges (see also Byers-Heinlein et al., 2020, for a fuller discussion of challenges in planning and conducting ManyBabies 1). First, several laboratories were not able to recruit the number of bilingual infants originally planned. All labs committed to collecting data from a minimum of 16 bilingual infants per age group. This ended up being unfeasible for some labs within the time frame available (which was more than a year), in some cases because a large number of participants did not meet our strict criterion for inclusion as bilingual. Our experience undoubtedly highlights the challenges for labs in recruiting bilingual infant samples, and moreover raises questions about how bilingualism should be defined and whether it should be treated as a continuous or categorical variable (Anderson et al., 2018; Bialystok et al., 2018; Incera & McLennan, 2018). Second, we had planned to explore the effect of different language pairs on IDS preference. We had expected that some labs would be able to recruit relatively homogeneous samples of infants (i.e., all learning the same language pair), but in the end only one of 17 labs did so (another lab had planned to recruit a homogeneous sample but deviated from this plan when it appeared unfeasible). Thus, we leave the question of the effect of

language pair on infants' IDS preference an open issue to be followed up in future studies. By and large, we believe that our large-scale approach to data collection may in the future allow for the creation of homogeneous samples of infants tested at different laboratories around the world. Large-scale and multisite bilingual research projects provide researchers with a powerful way to examine how the diversity and variability of bilinguals impact the development of their language skills and their cognitive development.

Overall, our finding that bilinguals and monolinguals show a similar preference for IDS reinforces theoretical views that emphasize the similarities in attentional and learning mechanisms across monolingual and bilingual infants (e.g., Curtin et al., 2011). IDS appears to be a signal that enhances attention in infants from a variety of language backgrounds. Yet bilingual infants appear to be exquisitely fine-tuned to the relative amount of input in each of their languages. It could have been the case that we would observe a threshold effect of language exposure such that any regular exposure to NAE enhanced infants' preference for NAE IDS, marking it is a highly relevant speech signal. Instead, we observed a graded effect such that the magnitude of bilingual infants' preference varied continuously with the amount of exposure to NAE. The current study shows that just as bilingual infants' relative vocabulary size and early grammar skills in each language are linked to the amount of input in that language (Hoff et al., 2012; Place & Hoff, 2011), the amount of language input may also play an important role in other language-acquisition processes. Indeed, an intriguing but untested possibility is that different input-related attentional biases (e.g., IDS preference) across bilinguals' two languages explain important variability in the early development of bilingual children's vocabulary and grammar. Future bilingual work can investigate this possibility to further delineate the interplay among infants' language input, IDS preference, vocabulary, and grammar skills.

To conclude, the findings of the current study provide a more nuanced view of the development of infants' preference for IDS than prior studies have allowed. IDS preference develops along a similar trajectory among infants from both monolingual and bilingual backgrounds. By testing bilingual infants, our study revealed that this IDS preference operates in a dose-response fashion; the amount of exposure to NAE positively moderated infants' (NAE) IDS preference in a continuous way. Our experience highlights the challenges in recruiting and testing bilingual infants, but also reveals the promise of large-scale collaborations for increasing sample sizes, and thus improving the replicability and generalizability of key findings in infant research.

Appendix

Table A1. Number of Male and Female Infants in Each Language Group Who Met Infant-Level Inclusion Criteria, by Lab

Lab	Monolingual females	Monolingual males	Bilingual females	Bilingual males
babylabbrookes	18	12	14	20
babylabkingswood	11	19	9	15
babylabparisdescartes1	7	9	5	6
babylabprinceton	1	0	10	5
bllumanitoba	18	24	9	6
cdcceu	8	5	8	6
infantcogubc	8	3	7	3
infantstudiesubc	8	12	9	6
irlconcordia	15	20	16	18
isplabmcgill	5	6	8	8
langlabucla	1	2	5	4
ldlottawa	16	12	14	11
lllliv	17	16	4	9
lscppsl	7	7	7	9
nusinfantlanguagecentre	8	12	24	14
weltentdeckerzurich	14	16	16	12
wsigoettingen	17	29	5	11

Note: Gender was not recorded for one additional infant from babylabbrookes.

Table A2. Number of Bilingual Infants per Unique Language Pair

Language pair	n
Albanian, non-NAE	1
Albanian, Swiss German	1
Arabic, French	5
Arabic, German	1
Arabic, NAE	2
Arabic, non-NAE	2
Armenian, French	1
Bahasa, non-NAE	1
Belizean creole, NAE	1
Bengali, non-NAE	1
Bosnian, non-NAE	1
Bulgarian, German	1
Cantonese, German	1
Cantonese, NAE	14
Cantonese, non-NAE	2
Dutch, French	1
Farsi, non-NAE	2
Finnish, German	1
Finnish, Swiss German	1
French, Georgian	1
French, German	2
French, Hungarian	2
French, Italian	4
French, Kabyle	1
French, Korean	1

(continued)

Table A2. *(continued)*

Language pair	n
French, Lebanese	1
French, Mandarin	1
French, NAE	64
French, non-NAE	9
French, Persian	1
French, Polish	1
French, Portuguese	2
French, Romanian	1
French, Russian	1
French, Spanish	6
French, Swiss German	5
German, Hungarian	1
German, Kurdish	1
German, Lithuanian	1
German, NAE	5
German, non-NAE	9
German, Polish	2
German, Russian	2
Greek, non-NAE	2
Greek, Swiss German	1
Hebrew, Hungarian	3
Hebrew, NAE	3
Hindi, non-NAE	1
Hungarian, Italian	1
Hungarian, NAE	1
Hungarian, non-NAE	4

(continued)

Table A2. *(continued)*

Language pair	n
Hungarian, Russian	2
Hungarian, Spanish	1
Indonesian, NAE	1
Indonesian, non-NAE	1
Italian, NAE	1
Italian, non-NAE	2
Italian, Swiss German	3
Japanese, non-NAE	3
Khmer, non-NAE	1
Korean, NAE	2
Malayalam, NAE	1
Mandarin, NAE	7
Mandarin, non-NAE	44
NAE, Persian	1
NAE, Polish	1
NAE, Punjabi	3
NAE, Russian	3
NAE Spanish	17
NAE, Swedish	2
NAE, Swiss German	1
NAE, Tagalog	2
NAE, Telugu	1
NAE, Urdu	1

(continued)

Table A2. *(continued)*

Language pair	n
Nepali, non-NAE	1
non-NAE, Patois	1
non-NAE, Polish	7
non-NAE, Portuguese	7
non-NAE, Punjabi	1
non-NAE, Russian	1
non-NAE, Slovenian	1
non-NAE, Spanish	7
non-NAE, Swiss German	5
non-NAE, Tagalog	2
non-NAE, Tamil	1
non-NAE, Turkish	1
non-NAE, Ukrainian	1
non-NAE, Urdu	1
non-NAE, Vietnamese	1
non-NAE, Welsh	2
non-NAE, Wu	1
Portuguese, Swiss German	1
Romansh, Swiss German	1
Serbian, Swiss German	1
Slovenian, Swiss German	1
Spanish, Swiss German	6
Swiss German, Turkish	1

Note: NAE = North American English; non-NAE = variety of English other than NAE.

Transparency

ORCID iDs

Krista Byers-Heinlein https://orcid.org/0000-0002-7040-2510
Maria Julia Carbajal https://orcid.org/0000-0001-8275-1147
Samantha Durrant https://orcid.org/0000-0001-9720-0866
Michael C. Frank https://orcid.org/0000-0002-7551-4378
Anja Gampe https://orcid.org/0000-0001-9812-9694
Judit Gervain https://orcid.org/0000-0002-2125-6369
Nayeli Gonzalez-Gomez https://orcid.org/0000-0003-2012-8290
Naomi Havron https://orcid.org/0000-0001-6429-1546
Jessica E. Kosie https://orcid.org/0000-0002-2390-0963
Liquan Liu https://orcid.org/0000-0001-8671-5098
Melanie Soderstrom https://orcid.org/0000-0003-3212-5775

Prior Versions

Preprint versions of this article are posted on PsyArXiv, at https://psyarxiv.com/sqh9d/.

References

Anderson, J., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). *Parsimonious mixed models*. arXiv. https://arxiv.org/abs/1506.04967

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

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, *9*(1), 43–61. https://doi.org/10.1207/s1532799xssr0901_4

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2018). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics*, *39*(4), 514–526. https://doi.org/10.1016/j.wocn.2011.02.001

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, *65*(1), 33–69. https://doi.org/10.1016/S0010-0277(97)00040-1

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*(1), 29–49. https://doi.org/10.1207/S15327078IN0201_3

Brito, N., & Barr, R. (2014). Flexible memory retrieval in bilingual 6-month-old infants. *Developmental Psychobiology*, *56*(5), 1156–1163. https://doi.org/10.1002/dev.21188

Brito, N., Sebastián-Gallés, N., & Barr, R. (2015). Differences in language exposure and its effects on memory flexibility in monolingual, bilingual, and trilingual infants. *Bilingualism: Language and Cognition*, *18*(4), 670–682. https://doi.org/10.1017/S1366728914000789

Byers-Heinlein, K. (2015). Methods for studying infant bilingualism. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 133–154). Cambridge University Press. https://doi.org/10.1017/CBO9781107447257

Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., Kominsky, J. F., Kosie, J. E., Lew-Williams, C., Liu, L., Mastroberardino, M., Singh, L., Waddell, C. P. G., Zettersten, M., & Soderstrom, M. (2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology*, *61*(4), 349–363. https://doi.org/10.1037/cap0000216

Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: Insights from bilingual infants. *Developmental Psychobiology*, *56*(2), 274–291. https://doi.org/10.1002/dev.21167

Byers-Heinlein, K., Fennell, C. T., & Werker, J. F. (2013). The development of associative word learning in monolingual and bilingual infants. *Bilingualism: Language and Cognition*, *16*(1), 198–205. https://doi.org/10.1017/S1366728912000417

Byers-Heinlein, K., Schott, E., Gonzalez-Barrero, A. M., Brouillard, M., Dubé, D., Jardak, A., Laoun-Rubenstein, A., Mastroberardino, M., Morin-Lessard, E., Iliaei, S. P., Salama-Siroishka, N., & Tamayo, M. P. (2019). MAPLE: A multilingual approach to parent language estimates. *Bilingualism: Language and Cognition*, *23*(5), 951–957. https://doi.org/10.1017/S1366728919000282

Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language & Communication Disorders*, *49*(6), 649–671. https://doi.org/10.1111/1460-6984.12082

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific

lexicalization patterns. *Cognition*, *41*(1), 83–121. https://doi.org/10.1016/0010-0277(91)90033-Z

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, *20*(4), 477–488. https://doi.org/10.1016/S0163-6383(97)90037-0

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

Cooper, R. P., & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, *65*(6), 1663–1677. https://doi.org/10.2307/1131286

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521–536. https://doi.org/10.1037/dev0000083

Curtin, S., Byers-Heinlein, K., & Werker, J. E. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*(4), 492–504. https://doi.org/10.1016/j.wocn.2010.12.002

Curto, J. D., & Pinto, J. C. (2011). The corrected VIF (CVIF). *Journal of Applied Statistics*, *38*(7), 1499–1507. https://doi.org/10.1080/02664763.2010.505956

Danielson, D. K., Seidl, A., Onishi, K. H., Alamian, G., & Cristia, A. (2014). The acoustic properties of bilingual infant-directed speech. *The Journal of the Acoustical Society of America*, *135*(2), EL95–EL101. https://doi.org/10.1121/1.4862881

DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The Language Exposure Assessment Tool: Quantifying language exposure in infants and children. *Journal of Speech, Language, and Hearing Research*, *59*(6), 1346–1356. https://doi.org/10.1044/2016_JSLHR-L-15-0234

Dunst, C., Gorman, E., & Hamby, D. (2012). Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning*, *5*(1), 1–13.

Farran, L. K., Lee, C.-C., Yoo, H., & Oller, D. K. (2016). Cross-cultural register differences in infant-directed speech: An initial study. *PLOS ONE*, *11*(3), Article e0151518. https://doi.org/10.1371/journal.pone.0151518

Fennell, C. T., & Byers-Heinlein, K. (2014). You sound like Mommy: Bilingual and monolingual infants learn words best from speakers typical of their language environments. *International Journal of Behavioral Development*, *38*(4), 309–316. https://doi.org/10.1177/0165025414530631

Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development*, *78*(5), 1510–1525. https://doi.org/10.1111/j.1467-8624.2007.01080.x

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/S0305000900010679

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., & Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435. https://doi.org/10.1111/infa.12182

García-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2016). Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *International Journal of Psychology*, *110*, 1–17. https://doi.org/10.1016/j.ijpsycho.2016.10.004

Glenn, S. M., & Cunningham, C. C. (1983). What do babies listen to most? A developmental study of auditory preferences in nonhandicapped infants and infants with Down's syndrome. *Developmental Psychology*, *19*(3), 332–337. https://doi.org/10.1037/0012-1649.19.3.332

Graf Estes, K., & Hay, J. F. (2015). Flexibility in bilingual infants' word learning. *Child Development*, *86*(5), 1371–1385. https://doi.org/10.1111/cdev.12392

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824. https://doi.org/10.1111/infa.12006

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes Publishing.

Hayashi, A., Tamekawa, Y., & Kiritani, S. (2001). Developmental change in auditory preferences for speech stimuli in Japanese infants. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1189–1200. https://doi.org/10.1044/1092-4388(2001/092)

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27. https://doi.org/10.1017/S0305000910000759

Hudon, T. M., Fennell, C. T., & Hoftyzer, M. (2013). Quality not quantity of television viewing is associated with bilingual toddlers' vocabulary scores. *Infant Behavior and Development*, *36*(2), 245–254. https://doi.org/10.1016/j.infbeh.2013.01.010

Incera, S., & McLennan, C. T. (2018). The time course of within and between-language interference in bilinguals.

International Journal of Bilingualism, *22*(1), 88–99. https://doi.org/10.1177/1367006916644688

Karzon, R. G. (1985). Discrimination of polysyllabic sequences by one- to four-month-old infants. *Journal of Experimental Child Psychology*, *39*(2), 326–342. https://doi.org/10.1016/0022-0965(85)90044-X

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*(1), 111–116. https://doi.org/10.1016/0163-6383(95)90012-8

Kitamura, C., & Lam, C. (2009). Age-specific preferences for infant-directed affective intent. *Infancy*, *14*(1), 77–100. https://doi.org/10.1080/15250000802569777

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, *24*(4), 372–392. https://doi.org/10.1016/S0163-6383(02)00086-3

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., . . . Nosek, B. A. (2014). Investigating variation in replicability: A "Many Labs" replication project. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Kovács, Á. M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences, USA*, *106*(16), 6556–6560. https://doi.org/10.1073/pnas.0811323106

Kovács, Á. M., & Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*(5940), 611–612. https://doi.org/10.1126/science.1173947

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences, USA*, *103*(17), 6771–6774. https://doi.org/10.1073/pnas.0602027103

Liu, L., & Kager, R. (2015). Bilingual exposure influences infant VOT perception. *Infant Behavior and Development*, *38*, 27–36. https://doi.org/10.1016/j.infbeh.2014.12.004

Liu, L., & Kager, R. (2017a). Is mommy talking to daddy or to me? Exploring parental estimates of child language exposure using the Multilingual Infant Language Questionnaire. *International Journal of Multilingualism*, *14*(4), 366–377. https://doi.org/10.1080/14790718.2016.1216120

Liu, L., & Kager, R. (2017b). Perception of tones by bilingual infants learning non-tone languages. *Bilingualism: Language and Cognition*, *20*(3), 561–575. https://doi.org/10.1017/S1366728916000183

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech.

Language Learning and Development, *7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

The ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Mattock, K., Polka, L., Rvachew, S., & Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science*, *13*(1), 229–243. https://doi.org/10.1111/j.1467-7687.2009.00891.x

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. https://doi.org/10.1002/dev.21177

Molnar, M., Carreiras, M., & Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition*, *152*, 150–159. https://doi.org/10.1016/j.cognition.2016.03.023

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., & Senior, A. M. (2015). Meta-analysis of variation: Ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution*, *6*(2), 143–152. https://doi.org/10.1111/2041-210X.12309

Newman, R. S., & Hussain, I. (2006). Changes in preference for infant-directed speech in low and moderate noise by 4.5- to 13-month-olds. *Infancy*, *10*(1), 61–76. https://doi.org/10.1207/s15327078in1001_4

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Orena, A. J., Byers-Heinlein, K., & Polka, L. (2020). What do bilingual infants actually hear? Evaluating measures of language input to bilingual-learning 10-month-olds. *Developmental Science*, *23*(2), Article e12901. https://doi.org/10.1111/desc.12901

Orena, A. J., & Polka, L. (2019). Monolingual and bilingual infants' word segmentation abilities in an inter-mixed dual-language task. *Infancy*, *24*(5), 718–737. https://doi.org/10.1111/infa.12296

Patihis, L., Oh, J. S., & Mogilner, T. (2015). Phoneme discrimination of an unrelated language: Evidence for a narrow transfer but not a broad-based bilingual advantage. *International Journal of Bilingualism*, *19*(1), 3–16. https://doi.org/10.1177/1367006913476768

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development*, *15*(3), 325–345. https://doi.org/10.1016/0163-6383(92)80003-D

Petitto, L. A., Berens, M. S., Kovelman, I., Dubins, M. H., Jasinska, K., & Shalinsky, M. (2012). The "Perceptual Wedge Hypothesis" as the basis for bilingual babies' phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and Language*, *121*(2), 130–143. https://doi.org/10.1016/j.bandl.2011.05.003

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*, *82*(6), 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy*, *17*(2), 198–232. https://doi.org/10.1111/j.1532-7078.2011.00075.x

Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language*, *13*(1), 85–100. https://doi.org/10.1017/S0305000900000313

Ramírez, N. F., Ramírez, R. R., Clarke, M., Taulu, S., & Kuhl, P. K. (2017). Speech discrimination in 11-month-old bilingual and monolingual infants: A magnetoencephalography study. *Developmental Science*, *20*(1), Article e12427. https://doi.org/10.1111/desc.12427

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891. https://doi.org/10.1111/desc.12172

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). The impact of early social interactions on later language development in Spanish–English bilingual infants. *Child Development*, *88*(4), 1216–1234. https://doi.org/10.1111/cdev.12648

Robertson, S., von Hapsburg, D., & Hay, J. S. (2013). The effect of hearing loss on the perception of infant- and adult-directed speech. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1108–1119. https://doi.org/10.1044/1092-4388(2012/12-0110)

Santesso, D. L., Schmidt, L. A., & Trainor, L. J. (2007). Frontal brain electrical activity (EEG) and heart rate in response to affective infant-directed (ID) speech in 9-month-old infants. *Brain and Cognition*, *65*(1), 14–21. https://doi.org/10.1016/j.bandc.2007.02.008

Schreiner, M. S., Altvater-Mackensen, N., & Mani, N. (2016). Early word segmentation in naturalistic environments: Limited effects of speech register. *Infancy*, *21*(5), 625–647. https://doi.org/10.1111/infa.12133

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science*, *23*(9), 994–999. https://doi.org/10.1177/0956797612436817

Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 974–989. https://doi.org/10.1037/0096-1523.28.4.974

Senior, A. M., Gosby, A. K., Lu, J., Simpson, S. J., & Raubenheimer, D. (2016). Meta-analysis of variance: An illustration comparing the effects of two dietary interventions on variability in weight. *Evolution, Medicine, and Public Health*, *2016*(1), 244–255. https://doi.org/10.1093/emph/eow020

Shute, B., & Wheldall, K. (1999). Fundamental frequency and temporal modifications in the speech of British fathers to their children. *Educational Psychology*, *19*(2), 221–233. https://doi.org/10.1080/0144341990190208

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Singh, L., Fu, C. S. L., Rahman, A. A., Hameed, W. B., Sanmugam, S., Agarwal, P., Jiang., B., Chong, Y. S., Meaney, M. J., Rifkin-Graboi, A., & the GUSTO Research Team. (2015). Back to basics: A bilingual advantage in infant visual habituation. *Child Development*, *86*(1), 294–302. https://doi.org/10.1111/cdev.12271

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, *3*(3), 365–394. https://doi.org/10.1207/S15327078IN0303_5

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, *14*(6), 654–666. https://doi.org/10.1080/15250000903263973

Slobin, D. I. (Ed.). (1985). *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues.* Psychology Press. https://doi.org/10.4324/9780203781890

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*(4), 505–513. https://doi.org/10.1016/j.wocn.2010.08.006

Tal, S., & Arnon, I. (2018). SES effects on the use of variation sets in child-directed speech. *Journal of Child Language*, *45*(6), 1423–1438. https://doi.org/10.1017/S0305000918000223

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, *32*(3), 492–504. https://doi.org/10.1037/0012-1649.32.3.492

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24*(3), 535–565. https://doi.org/10.1017/S030500099700319X

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335–340. https://doi.org/10.3758/BF03196290

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in Cognitive Sciences*, *12*(4), 144–151. https://doi .org/10.1016/j.tics.2008.01.008

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *43*(2), 230–246. https://doi.org/10.1037/ h0084224

Werker, J. F., Pegg, J. E., & McLeod, P. J. (1994). A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior and Development*, *17*(3), 323–333. https://doi.org/10.1016/0163-6383(94)90 012-4

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80 022-3