The Development of Gaze Following in Monolingual and Bilingual Infants: A Multi-Lab Study

Krista Byers-Heinlein[1]

[1] Concordia University

Author Note

Correspondence concerning this article should be addressed to Krista Byers-Heinlein, Department of Psychology, 7141 Sherbrooke St. West, Montreal, QC, Canada, H2T1V2. E-mail: k.byers@concordia.ca

Abstract

Determining the meanings of words requires language learners to attend to what other people say. However, it behooves a young language learner to simultaneously attend to what other people do, for example, following the direction of their eye gaze. Sensitivity to cues such as eye gaze might be particularly important for bilingual infants, as they encounter less consistency between words and objects than monolinguals, and do not always have access to the same word learning heuristics (e.g. mutual exclusivity). We tested the hypothesis that bilingual experience would lead to a more pronounced ability to follow another's gaze. We used the gaze-following paradigm developed by Senju and Csibra (2008) to test a total of XX 6–9 month-old and XX 12–15 month-old monolingual and bilingual infants, in XX labs located in XX countries. DESCRIBE RESULTS HERE.

Keywords: gaze following, bilingualism, infancy, replication, reproducibility, age-related changes

The Development of Gaze Following in Monolingual and Bilingual Infants: A Multi-Lab Study

Bilingual infants face the remarkable task of acquiring two languages simultaneously. Bilinguals show developmental adaptations to their unique environments, which might support their observed success in learning their two languages (Werker & Byers-Heinlein, 2008). In comparison to monolinguals, bilingual infants show differences in early speech perception (Byers-Heinlein & Fennell, 2014), in word learning (Fennell, Byers-Heinlein, & Werker, 2007; Graf Estes & Hay, 2015; Singh, Fu, Tay, & Golinkoff, 2017), and in acquisition of grammatical structures (Kovács & Mehler, 2009a; Antovich & Graf-Estes, 2017). They show different patterns of looking towards talking faces (Pons, Bosch, & Lewkowicz, 2015), and are more sensitive to facial cues that discriminate speakers of different languages (Weikum et al., 2007; Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012). These differences have been attributed to specific features of bilingual environments that may influence developing cognitive processes. Specifically, the notion that bilingual infants attend to and learn two languages is thought to sharpen their capacity to flexibly switch between their languages (Antovich & Graf-Estes, 2017; Kandahai, Danielson & Werker, 2014; Kovács & Mehler, 2009b) and to acquire the individual properties of two language systems (Kovács & Mehler, 2009a). Moreover, as bilingual infants typically encounter less single-language input than their monolingual peers, new information may be encoded with increased efficiency and detail (Brito & Barr, 2014; Liu & Kager, 2016; Singh et al., 2015). These findings suggest that, before infants begin to produce words in their native language(s), immersion in a bilingual environment modifies the development of some aspects of infants' perception and learning.

More intriguingly, bilingualism also appears to impact abilities that do not directly involve language. For example, relative to monolinguals, bilingual infants are more likely to inhibit recently learned information (Kovács & Mehler, 2009b), generalize across visual features when categorizing objects (Brito & Barr, 2014), and encode and retrieve visual information (Singh et al., 2015). Here, we ask whether bilingual infants also show enhanced sensitivity to non-linguistic social information, a question that has thus far received very little attention. In an international, multi-site study, we investigated whether the ability to follow a social partner's eye gaze follows the same developmental trajectory in monolingual and bilingual infants. REPORT MAIN RESULT HERE.

The development of gaze following

Infants show an early-emerging sensitivity to a social partner's eye gaze. In a primitive form, very young infants are sensitive to the direction of a speaker's gaze, attending to visual targets more rapidly when they are cued by an adult's gaze (Farroni, Massaccesi, Pivodori, & Johnson, 2004; Reid, Striano, Kaufman, & Johnson, 2004). Throughout the first year and a half of life, infants refine their interpretation of eye and head movements: they distinguish between head-turns with open versus closed eyes (Brooks & Meltzoff, 2005), become able to follow changes in gaze unaccompanied by a head turn (Corkum & Moore, 1995; Moore & Corkum, 1998), and attend to whether another's gaze is obscured from view by a physical barrier (Meltzoff & Brooks, 2007). In sum, over the course of infancy, infants progress from attending to the direction of the eyes, to engaging in gaze following in a more selective fashion, to true gaze following where the actions of a social partner are interpreted as intentional and informative (Brooks & Meltzoff, 2014; Frischen, Bayliss, & Tipper, 2007; Moore, 2008).

A number of recent studies have highlighted the situations that most reliably elicit gaze following in infancy. As an example, Senju and Csibra (2008) investigated gaze-following abilities of 6-month-old infants. This age is of particular interest as it corresponds to the onset of word comprehension (Bergelson & Swingley, 2012; Fenson et al., 2007). An adult model sat in between two toys, one located to her left and one to her right. Infants were tested in one of two conditions. In the eye gaze condition (labelled the

Eye Contact condition by Senju and Csibra), the model made an eye gaze towards the infant and then directed her gaze at one of the toys. In the no eye gaze condition (labelled the No Eye Contact condition by Senju and Csibra), the model initially looked down instead of towards the infant, and a superimposed animation drew the infant's attention to her head. Results revealed that infants followed the model's gaze in the eye gaze condition, but not in the no eye gaze condition. In a replication and extension of Senju and Csibra's (2008) paradigm, Szufnarowska, Rohlfing, Fawcett and Gredebäck (2014) demonstrated that 6-month-old infants responsively followed an adult's gaze similarly when it was cued by attention-grabbing behaviors without eye gaze, such as shivering or nodding. This suggests that the ability for eye gaze to elicit gaze following behavior may be partially related to its attentional draw.

Several studies have used the paradigm developed by Senju and Csibra (2008) to explore how infants' individual experiences with gaze affect their gaze-following abilities. For example, one study investigated gaze following in sighted infants of blind parents (Senju et al., 2013). These infants showed a similar ability to follow the gaze of a sighted social partner as infants of sighted parents, despite having less experience with gaze behaviors. Another study looked at gaze following in infants at risk for communicative impairments (Bedford et al., 2012). Although both at-risk and low-risk infants were equally likely to follow an adult's gaze, at-risk 13-month-olds spent less total time looking at objects to which an adult's gaze was directed. This suggested that they might have been less able to make use of gaze as a socially relevant cue than typically developing infants. Together, these studies suggest that gaze following is an ability that develops across varied developmental circumstances, although the results from at-risk infants show that the use of gaze information can differ across populations. Importantly, these studies provide support for the use of Senju and Csibra's (2008) task, which has elicited gaze following across studies and populations.

Gaze following in bilinguals

One group of infants that might differ in the development of gaze-following abilities is bilingual infants, although no study to date has specifically tested this group. There are several reasons to posit that bilinguals may demonstrate increased attention to gaze patterns of social partners. One reason is that gaze following is not only an important social skill, but it also contributes to early language learning. Language is a highly social system of communication. Speakers often look towards their intended referent. Thus the ability to follow a conversational partner's gaze can guide children in correctly mapping words to objects, and help to resolve the problem of referential ambiguity (Baldwin, 1995; Brooks & Meltzoff, 2002; Tomasello, 2003; Woodward, 2003). Many theories of language acquisition emphasize the influence of social cues in the search for meaning, proposing that infants' sensitivity to social cues scaffolds accurate and efficient vocabulary development (Baldwin, 1995; Bloom, 2000; Hollich, Hirsh-Pasek, & Golinkoff, 2000; Mundy, Sullivan, & Mastergeorge, 2009; Tomasello, 2003). There is substantial empirical support for this theoretical stance: infant gaze following is both concurrently and predictively related to word learning (e.g. Brooks & Meltzoff, 2005; 2008; Carpenter, Nagell, & Tomasello, 1998; Morales et al., 2000; Mundy et al., 2007; Paulus & Fikkert, 2014; Tenenbaum et al., 2015).

The ability to use gaze information in language learning might be particularly important for bilingual infants. Bilingual infants' experiences are divided between their two languages, and they must learn two labels for each object (one in each language). When a monolingual English-learning infant encounters an object such as an apple, they will consistently hear the word "apple" to refer to that object. However, when a French-English bilingual encounters the same object, they will sometimes hear the English word "apple" and sometimes hear the French word "pomme". For bilinguals, there may be less consistency in object-label correspondences. Unlike monolinguals, they eventually have to map at least two labels to each object (one in each language).

The need to map multiple labels onto the same object may make some of the word learning strategies used by monolingual learners less useful for bilingual learners. Both groups should share basic assumptions about the relationship between words and objects that can support word learning, like the assumption that words refer to whole objects rather than their parts, and that a new word should be extended to other objects of the same kind (Markman, 1990). However, one key assumption that may differ across monolinguals and bilinguals is mutual exclusivity, the assumption that each object has a unique label (Markman & Wachtel, 1988). Mutual exclusivity allows monolinguals to reject objects with a known label as a referent for a novel word. Strict use of such a heuristic would be less useful to bilingual learners, as they must learn two labels for each object. Indeed, evidence from bilingual infants at age 9 months (Byers-Heinlein, 2016) and 17-18 months (Byers-Heinlein & Werker, 2009; 2013; Houston-Price, Caloghiris, & Raviglione, 2010) indicates that bilinguals do not assume that each object has only one label, and are less likely to use word learning heuristics such as mutual exclusivity. If mutual exclusivity is less available to bilingual word learners, then they might need to more strongly rely on other cues to word meaning such as eye gaze.

Another important monolingual-bilingual difference is that bilingual learners receive less input in each language in comparison to monolingual learners. While this might be expected to delay word learning, bilingual infants comprehend and produce their first words on largely the same schedule as monolingual infants (De Houwer, Bornstein, & De Coster, 2006; Petitto et al., 2001). Moreover, when vocabulary in both languages is considered, monolinguals and bilinguals have similar vocabulary sizes (Core, Hoff, Rumiche, & Señor, 2013; Pearson, Fernández, & Oller, 1993). Thus, bilinguals appear to have a similar rate of vocabulary development despite reduced frequency of exposure to particular words (although see Bilson, Yoshida, Tran, Woods, & Hills, 2015, for a different perspective). This could imply that bilinguals are adept at leveraging other sources of information for word learning, such as eye gaze, which could offset the effects of reduced single-language input.

There is some evidence from older children to support the hypothesis that bilingual infants have an enhanced ability to follow a social partner's gaze. For example, when object cues and eye gaze cues to meaning were pitted against one another, 2- to 3-year-old bilinguals weighed eye gaze cues more heavily than monolinguals to identify the referent of a newly learned word (Brojde, Ahmed, & Colunga, 2012). In a similar study, Yow and Markman (2011) demonstrated that 2- to 4-year-old bilingual children made greater use of eye gaze than monolingual peers to locate a hidden object. This effect was observed only under challenging circumstances in which the experimenter was seated at a distance from the referent and closer to a competing distractor object. In a study investigating children's use of eye gaze to map novel words to referents and additionally, to infer the meanings of other words via mutual exclusivity, Yow et al. (2017) found that 4- to 5-year-old bilingual children made greater use of eye gaze to identify word-meaning links that were directly taught as well as those that were identifiable via mutual exclusivity.

Together, these studies provide evidence that preschool-aged bilingual children are more adept than monolinguals at using eye gaze cues in word learning contexts. This raises the possibility that bilinguals might also show enhanced sensitivity to a social partner's eye gaze even earlier in development than monolinguals.

A multi-lab collaborative study

We conducted a multi-lab collaborative study to investigate whether infants' language background can influence the development of gaze following. Multi-lab collaborative studies, which involve data collection across multiple sites to generate a large-scale data set, offer several promises for infant research. This approach allows us to increase the diversity and the size of the sample than can be collected in a single laboratory, protecting against incorrect conclusions due to sampling error. Moreover, comparisons across labs can speak to the generalizability of results. For example, such an approach could clarify whether any observed monolingual-bilingual differences generalize across different samples and could reveal whether any observed effects are likely due to bilingualism per se or could be attributed to other sample characteristics, for example, the specific language or cultural context. Within infant bilingualism research, very few studies have collected data from multiple groups of monolingual and bilingual infants on the same task, and cross-cultural comparisons on infant bilingual development are entirely absent.

There are many methodological challenges faced in conducting research with bilinguals, particularly in infancy, that motivate using a multi-lab collaborative approach. Many of these challenges are inherent to the nature of the population, and make it difficult to know whether and how findings from one population of bilinguals generalize to other populations. First, while the term “bilingual” can be used for any infant who is exposed to two or more languages, bilingual infants vary in the particular language pair they are learning. Some studies have included only groups of homogeneous bilinguals (i.e. infants exposed to the same pair of languages, such as Spanish-Catalan), while others have included heterogeneous bilinguals (i.e. infants exposed to different pairs of languages, having one language in common, for example, English-Japanese, English-Spanish, English-French). Different language combinations could present different language-learning challenges. While our study was not designed to tease apart the role of particular language pairings (although our data do allow us to explore this issue in a preliminary way), it will establish the generalizability of findings across different groups of bilinguals.

Second, given the continuous nature of language exposure, it is challenging to validly and consistently define what makes an infant “monolingual” versus “bilingual”. Specifically, few infants are exposed to their two languages in an exactly equal proportion. Instead, the amount of exposure to each of their languages can vary enormously, and there is not always consensus about how much exposure is necessary to acquire a language. As a result, different studies have defined bilingualism differently: while in some studies 10% exposure to the non-dominant language was enough for infants to be considered bilingual, other studies required at least 40% of exposure (Byers-Heinlein, 2015), although 25% is a commonly-used cutoff. An additional complication is that the onset of exposure to any additional languages is highly variable, and could be as early as birth or anytime thereafter. Published studies differ with respect to whether strict or relaxed inclusion requirements are set for the onset of exposure to different languages. A benefit of this collaborative approach is that there is a consistent definition of exposure across participating laboratories.

Finally, bilingualism cannot be randomly assigned. Thus, even when recruited from the same geographic region, monolingual and bilingual populations often differ systematically in culture or socio-economic status. Such confounds can make it difficult to determine whether bilingualism itself, rather than another correlated variable, drives observed monolingual–bilingual differences. While such factors can be statistically controlled, these confounds can raise issues about the validity of conclusions and the replicability of the results in bilingualism research. In particular, a number of reports have suggested that long-standing beliefs about the cognitive effects of bilingualism may not be as robust as previously assumed (de Bruin, Treccani, & Sala, 2015; Duñabeitia, & Carreiras, 2015; Paap, Johnson, & Sawi, 2015, see also Klein, 2015). Indeed, such issues are of increasing concern in the wider field of psychology, where there are ongoing concerns about the replicability of psychological research in general (see Ioannidis, 2012), and specifically about the cross-cultural replicability of basic psychological phenomena thought to be universal (Henrich, Heine, & Norenzayan, 2010). Concerns about the replicability and generalizability of research findings are particularly acute in the field of infant research, where single-lab studies tend to have small sample sizes, high variability, and use indirect experimental measures (see Frank et al., 2017, for a detailed discussion of these issues). Multi-lab studies can go further than single-lab studies to address many of these issues. Characteristics that are idiosyncratic to a particular sample will average out to some degree in a multi-lab study that includes samples from multiple cultures and language backgrounds. Our approach of comparing gaze following in monolinguals and bilinguals growing up in different contexts, tested across multiple labs, provides important information about the replicability and generalizability of the effects we observe.

Current study

The current study used a multi-lab approach to ask whether monolingual and bilingual infants differ in their basic gaze-following abilities. Data were collected from 11 labs in 8 countries. We tested the hypothesis that the challenging nature of bilingual language-learning environments enhances bilingual infants' attention to the eye gaze of a social partner, even in non-linguistic situations. Our study compared monolingual and bilingual infants aged 6–9 and 12–15 months using the eye gaze stimuli from Senju and Csibra's (2008) study. Note that our study did not include the no eye gaze condition reported in Senju and Csibra's paper, as our interest was in comparing gaze-following behavior in typical situations, across infants from different language backgrounds. On six test trials, infants saw a model look towards the camera, and then direct her head and eyes towards one of two objects located to her left and right. We measured the latency and accuracy of infants' gaze following.

Previous studies have found that infants follow the actor's gaze in this condition at above-chance level by 6 months, but their performance is not always reliable (Senju & Csibra, 2008; Szufnarowska et al., 2014). Moreover, there is evidence for improvement of infants' gaze following in this paradigm from 7 to 13 months (Bedford et al., 2012). We thus expected to see improvement in all infants' gaze following from the younger age to the older age. We also expected that both groups would demonstrate successful gaze following as demonstrated by Senji and Csibra (2008), but that bilingual infants would show faster and more accurate gaze following than monolingual infants. We also suggest that the effects of bilingualism might interact with age. On the one hand, we might observe a stronger effect of bilingualism at 6-9 months if gaze following emerges earlier for bilinguals;

on the other hand we might observe a stronger effect of bilingualism at 12-15 months if this skill emerges at the same age, but is more relied upon by bilingual infants as the demands of language acquisition increase. Both of these findings would reflect interesting and meaningful differences between monolingual and bilingual infants.

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participation Details

Time frame. An open call for labs to participate was issued on March 14, 2017. Participant testing began on July 1, 2017 and ended on August 31, 2018.

Age and language groups. Labs contributed samples from one or both of two possible age bins: 6–9 months (184–274 days) and 12–15 months (366–456 days). Labs were asked to aim for a mean age at the centre of each bin, with distribution across the entire age window. Labs could contribute a monolingual and/or bilingual sample at one or both ages (see below for inclusion criteria for monolingual and bilingual groups).

Lab participation. Considering the challenges associated with recruiting bilingual infants and the importance of counterbalancing in our experimental design, we asked labs to contribute a minimum of 16 healthy, full-term infants per age (6–9, 12–15) and language group (monolingual, bilingual). However, labs were encouraged to contribute data even if they were only able to provide a bin of data for a single age or for a single language group. Further, labs were invited to contribute additional data provided that decisions about when to stop data collection were made without looking at the data, to avoid biasing effect sizes. Labs were asked to screen ahead of time that infants met inclusion criteria. However, it was acknowledged that most labs would end up recruiting infants who did not necessarily meet our pre-defined criteria for bilingualism (detailed below) upon more detailed in-lab language background assessment. In such cases, the decision whether to test the infant was left up to individual laboratories' policies, 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). Eleven labs contributed at least one data bin.

Nine of the 11 participating labs were also participating in two prior multi-lab collaborative studies (ManyBabies 1 study and/or ManyBabies 1 Bilingual study) investigating infants' preference for infant-directed speech (ManyBabies Consortium, 2019a; 2019b). The current study emerged out of the unique opportunity afforded by a significant number of labs with a bilingual population coming together to run the Manybabies 1 Bilingual study, and the desire to make optimal use of these resources. As such, prior to completing the current study, 42.88% of the infants completed the ManyBabies 1 study on the same visit in the lab. Testing infants in two different studies on the same visit is a common practice in many, although not all, infant labs. We note that these two studies adopted different designs (listening preference vs. gaze following), and tracked sensitivity to different sorts of cues (auditory vs. visual). Moreover, the current study (gaze following) presented infants with engaging social stimuli and was short in duration. These features mitigated possible carryover effects.

Power analysis. In their paper, Senju and Csibra (2008) report a comparison against chance of $t(18) = 2.74$ in our target condition (the eye gaze condition of Experiment 2), yielding a calculated effect size of Cohen's $d = 1.29$ for infants of this age for the first look measure. This would necessitate a sample size of only 6 infants to have an 80% chance of detecting a significant difference in a single-sample t-test. With our planned sample size of 16 infants/group per lab, power within each lab to detect this effect will be .94.

However, our primary hypothesis concerned the comparison of monolingual and bilingual infants. Because this is the first study to investigate this question, it is difficult to know what effect size might be expected in this comparison. We thus conducted a sensitivity analysis, setting target power at .8 and alpha at .05. For individual labs to detect a statistically significant difference between monolinguals and bilinguals (n = 16 infants per group) in an independent samples t-test, we would need to observe a large effect size of Cohen's $d = 1.0$. However, collapsing across the labs (projected to be approximately 100 monolinguals and bilinguals per age group for a total sample of 400), we would be able to detect a small to medium effect size of Cohen's $d = .28$ at either age. Conducting multiple regression models with 3–6 predictors (see analytic plan) with the data from all labs across both age groups, we would be able to detect statistically significant contribution(s) from between one (e.g. bilingualism) and three (e.g., bilingualism, age, and their interaction) predictors with a small effect size in the range of Cohen's $f^2 = .019$–$.028$. Thus, we felt confident that our design would have sufficient statistical power to detect a difference between monolinguals and bilinguals that was small to medium in magnitude.

Ethics. The present study was conducted according to the Declaration of Helsinki guidelines, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Institutional Review Board at the institutions where data was collected. Each lab followed the ethical guidelines and ethics review board protocols of their own institutions. Labs submitted anonymized data for central analysis that identified participants by code only. Data from individual participants were coded and stored locally at each lab, and, where possible, were uploaded to a central controlled-access databank accessible to other researchers.

Participants

Classification of participants into language groups. As in previous studies, infants were categorized as bilingual or monolingual according to parent estimates of language input to their child. Infants were classified as monolingual if they heard the community language at least 90% of the time. There is some variation across studies in how much exposure to the non-dominant language is typically required for infants to be classified as bilingual, with a range of values from 10% to 40% (Byers-Heinlein, 2015). A widely accepted criterion is a range of a minimum exposure estimate of 25% and maximum exposure of 75% to each language, which served as a recruitment guideline for the present study. Thus, our bilingual sample included infants who heard their community language (e.g., the language learned by most monolinguals in their community) at least 25% of the time and an additional language at least 25% of the time. Infants with exposure to a third or fourth language were included as long as they met this criterion. We also asked labs to limit their sample to simultaneous bilingual infants, who heard both languages regularly from within the first month of life. Infants who did not meet inclusion criteria for either group (for example, an infant with 85% exposure to one language, and 15% exposure to another, or who began learning a second language at age 6 months) could be tested if they inadvertently arrived in the lab, according to each lab's policy. However, their data were not included in the main sample, but were retained for further exploratory analysis. Each laboratory was asked to recruit a sample of bilingual infants who received exposure to the community language as one of their languages and to recruit monolingual infants exposed to the community language. As a result, some samples consisted of heterogenous bilinguals and others of homogenous bilinguals.

Each laboratory was asked to administer their own adaptation of a day-in-the-life parental interview asking about proportionate exposure to each language, which were typically based on the approach developed by Bosch and Sebastián-Gallés (2001). As laboratories often customize questionnaires to suit their local environment, it was concluded that each laboratory would be best able to decide on the variation of the language exposure tool that was optimal for their participant population. As some participating laboratories had not collected bilingual data prior to the study, these laboratories were paired with laboratories more experienced in infant bilingualism research to receive support and guidance in selecting or adapting a suitable language exposure questionnaire.

Although adapted for their language environment by each lab, there is consistency in the information sought from different versions of the language exposure questionnaire. Specifically, each adaptation walks parents through a "day-in-the-life" of their infant, asking about routines, caregivers, and the languages that they speak. An interviewer notes how much each language is spoken to the child during weekdays, weekends, and at different points of the infants' life from caregivers. Indirect exposure through media such as television and radio, as well as overhead speech, are typically excluded (Byers-Heinlein, 2015). Together, this information is used to calculate the total percentage that the infant is directly exposed to each language.

Demographics. Each lab administered a questionnaire that gathered basic demographic data about infants, including age, health history, gestation, etc.

Final sample. Our final sample of bilinguals included 131 infants tested in 9 labs. 45 were 6–9 months, and 86 were 12–15 months old. Each of these labs also collected data from monolingual infants (N = 149), of whom 30 were 6–9 months, and 119 were 12–15 months. Data from monolingual infants were available from two additional labs (N = 42), who did not contribute bilingual data. A list of monolingual and bilingual populations in each lab are reported in Table 1. In addition, 2 labs registered to participate but failed to collect data from at least 10 included infants, and so their data were not included. Information about all included labs is given in Table 1.

Table 1

Statistics of the included labs. N refers to the number of infants included in the final analysis.

Lab	Age group	Lang group	Mean age (days)	N	Method
babylab-brookes	12-15 mo	bilingual	394	15	eye-tracking
babylab-brookes	12-15 mo	monolingual	415	14	eye-tracking
babylab-brookes	6-9 mo	bilingual	242	8	eye-tracking

babylab-brookes	6-9 mo	monolingual	238	8	eye-tracking
babylab-princeton	12-15 mo	monolingual	421	14	hand-coding
babylab-princeton	6-9 mo	bilingual	239	9	hand-coding
cdc-ceu	12-15 mo	bilingual	420	11	eye-tracking
cdc-ceu	12-15 mo	monolingual	404	10	eye-tracking
elp-georgetown	12-15 mo	bilingual	416	4	eye-tracking
elp-georgetown	12-15 mo	monolingual	425	7	eye-tracking
elp-georgetown	6-9 mo	bilingual	260	4	eye-tracking
elp-georgetown	6-9 mo	monolingual	242	5	eye-tracking
infantlanglab-utk	12-15 mo	monolingual	408	15	hand-coding
infantlanglab-utk	6-9 mo	monolingual	239	13	hand-coding
irl-concordia	12-15 mo	bilingual	403	14	eye-tracking
irl-concordia	12-15 mo	monolingual	399	16	eye-tracking
irl-concordia	6-9 mo	bilingual	235	11	eye-tracking
irl-concordia	6-9 mo	monolingual	214	7	eye-tracking
koku-hamburg	12-15 mo	monolingual	419	9	eye-tracking
koku-hamburg	6-9 mo	monolingual	234	5	eye-tracking
lll-liv	12-15 mo	bilingual	390	7	eye-tracking
lll-liv	12-15 mo	monolingual	400	15	eye-tracking
lll-liv	6-9 mo	bilingual	235	7	eye-tracking
lll-liv	6-9 mo	monolingual	230	8	eye-tracking
nusinfantlanguagecentre	12-15 mo	bilingual	426	4	hand-coding, eye-tracking
nusinfantlanguagecentre	12-15 mo	monolingual	416	6	hand-coding, eye-tracking
nusinfantlanguagecentre	6-9 mo	bilingual	261	6	eye-tracking
nusinfantlanguagecentre	6-9 mo	monolingual	246	2	eye-tracking, hand-coding
upf_barcelona	12-15 mo	bilingual	414	7	eye-tracking

upf_barcelona	12-15 mo	monolingual	404	11	eye-tracking
weltentdecker-zurich	12-15 mo	bilingual	408	24	eye-tracking
weltentdecker-zurich	12-15 mo	monolingual	416	26	eye-tracking

Stimuli

Stimuli consisted of videos of a female actor sitting at a table, directing her gaze to one of two colorful toys. Each video had the following sequence: the video began with the actor looking straight ahead for 1 second. She looked down for two seconds, after which a beep sounded to attract infants' attention prior to the actor directing her gaze to a toy. Upon presentation of the beep, the actor looked up at the camera and, maintaining a neutral expression, she raised her eyebrows. Four seconds into the video, she began to turn her head towards the left or right and gazed towards the toy in her line of sight until the end of the video. There were a total of 24 different videos in this style, using six different pairs of colourful objects. Video presentations were counterbalanced for the side of presentation of the objects and the object at which the actor gazed, and arranged such that there were six test trials per infant. Original movies were in .avi format, exported at a framerate of 25 frames/second. Each movie lasted a total of 10 seconds (250 frames).

Procedure

We replicated the eye gaze condition of Experiment 1 from Senju and Csibra (2008), using the original stimuli provided by the authors. Infants were seated on their parents' laps in a quiet, dimly lit testing booth. Caregivers and infants were seated facing a monitor. The caregiver wore an occluder (e.g. sleep mask or opaque sunglasses) to prevent him/her from viewing events on the monitor. An experimenter controlled the study from an area located out of view of the infant, either in the same or a different room. Infants' eye gaze data were collected automatically via a corneal reflection eye-tracker, or on a digital videotape for later offline coding.

Each infant saw a series of 6 test videos. Infants were assigned to one of four possible trial orders that counterbalanced the direction of the actor's gaze (either LRRLRL or RLLRLR, where L denotes gaze to the toy on the left and R denotes fixation to the toy on the right), as well as which particular toy was located on the actor's left and right. Due to a programming error, one lab presented the same trials in a randomized order instead. Videos were separated by an unrelated attention-grabbing cartoon, which was played between trials until the infant had looked towards it for approximately 1–2 seconds. The experiment lasted approximately 1.5 minutes.

Exclusion Criteria

All data collected for the study (i.e., every infant for whom a data file was generated, regardless of how many trials were completed) were given to the analysis team for confirmatory analyses. Participants were only included in the analysis if they met all of the criteria below. All exclusion rules are applied sequentially, and percentages reflect this sequential application. N.B.: the first three criteria preemptively prevent participation (except in case of erroneously running the experiment with children outside of the inclusion guidelines).

Analysis overview

Data exclusion

Labs were asked to submit all data collected as part of the study to the analysis team. Data were first screened to determine whether labs contributed useable data and whether infants met our inclusion criteria. Note that some infants had more than one reason for exclusion, and exclusion criteria were applied sequentially.

- *Lab reliability.* Data from two of the labs using the hand-coding method were excluded after extensive discussions with the participating laboratories. One lab could not achieve an acceptable level of inter-rater reliability, due to difficulty coding infant eye movements from the available videos. A second lab initially coded the data incorrectly (i.e., coded gaze shift from face to object differently than had been specified), but then had insufficient resources to re-code the data. There were 104 (14.50%) infants who were tested in these labs.

- *Age.* There were 55 (9%) infants who were tested but were out of our target age groups (6-9 months and 12-15 months).

- *Language background.* There were 50 (9%) infants who were tested but did not meet our inclusion criteria for either the monolingual or bilingual group. For example, an infant who heard English 20% of the time and Italian 80% of the time would not meet the criteria as either monolingual (at least 90% exposure to one language) or bilingual (at least 25% exposure to each of two languages).

- *Full-term.* We defined full-term as gestation times greater than or equal to 37 weeks. There were 10 (2%) infants who were tested but did not meet this criterion.

- *No diagnosed developmental disorders.* We excluded data from 1 (0.20%) infant with parent-reported developmental disorders or sensory impairments.

- *Session errors.* There were 25 (5.07%) infants excluded from the analysis due to issues including: 12 for equipment failure, 10 for fussiness, and 3 for parental/external interference.

- *Insufficient face-to-object saccades.* Following Senju and Csibra (2008), and per our pre-registration, we also excluded any infant who did not make at least one gaze shift from face to object during the window of analysis in at least three of the six trials. A further 145 (31.05%) infants were excluded from analyses for this reason.

- *Failure to attend*. We also excluded any trials in which infants did not look at the congruent or incongruent object during the window of analysis. This meant that each infant contributed a different number of trials. An additional 360 trials (23.03%) were excluded from the analyses. This left us with a total number of 1563 valid trials (36.87% of the data before all the screenings) for later analyses: 211 trials for 6-to-9-month-old monolinguals (23.84% of the original data), 714 trials for 12-to-15-month-old monolinguals (56.18% of the original data), 201 trials for 6-to-9-month-old bilinguals (31.60% of the original data), and 437 trials for 12-to-15-month-old bilinguals (53.29% of the original data).

One lab mistakenly used a preliminary rather than the final version of the experiment. The version used contained the same experimental stimuli and events as the final version with two exceptions: the attention getter to recruit the infant's attention to the screen differed and the aspect ratio of the on-screen stimuli differed slightly. As this version of the experiment was only very slightly different from the final version, these data were retained for analysis.

Areas of interest and data pre-processing.

On eye-tracking setups, following Senju and Csibra (2008), we established three areas of interest (AOIs) on each trial (see Figure 1): the actor's entire face (taking into account the model's head movements) and two areas surrounding each of the two objects (corresponding to the size of the largest object). These rather generous AOIs maximized consistency between eye-tracking coding and human coding. The two object AOIs were labeled as congruent (i.e., the object target of the actor's gaze) and incongruent (i.e., the object that was not the target of the actor's gaze). Pixel coordinates for the AOIs were amended proportionally to each individual lab's screen resolution.

Eye-trackers measured the coordinates of eye gaze, from which the direction and duration of fixations and gaze shifts were calculated. See supplemental materials for details of hardware used in each lab. Most eye-tracking software comes with built-in algorithms to parse fixations and gaze shifts, but these are optimized for adult data and perform suboptimally in noisy infant data (Hessels, Andersson, Hooge, Nyström, & Kemner, 2015; van Renswoude et al., 2017; Wass, Smith, & Johnson, 2013). To overcome this, and to standardize results between labs using different eye-tracking systems, we implemented a common approach using the GazePath tool for fixation and saccade detection, as outlined in Van Renswoude et al. (2017). This approach is optimized for dealing with noisy infant data and individual differences that are expected between infants of different ages.

For labs that did not have an eye-tracker, trained human coders examined videos of infants' faces frame-by-frame to identify fixations and gaze shifts. Fixations were coded for duration and location with respect to the areas of interest (i.e. congruent object, incongruent object, actor, or off-target). Shifts were coded for direction, defined with respect to the horizontal and vertical midlines; i.e. movement could be left, right, down, and/or up. For these labs, a minimum of 25% of participants were double-coded by a second human coder and reliability estimates computed.

Data reliability

Because of the variability across labs in terms of methods and setups, different intrinsic reliability issues emerged regarding data consistency across different eye-tracker setups, between different human coders, and between eye-tracker and manual coding setups. These issues have been addressed in three different ways. First, as described above, all eye-tracking data were processed using the same GazePath tool, which is optimized to account for variability across different ages, populations, and setups (Van Renswoude et al., 2017). Second, all labs using human coding rather than an eye-tracker double coded a minimum of 25% of their data. For 6- to 9-month-olds, frame and shift agreement ranged from 98.27-99.27% and 95.40-99.55%, respectively. For 12- to 15-month-olds, frame and

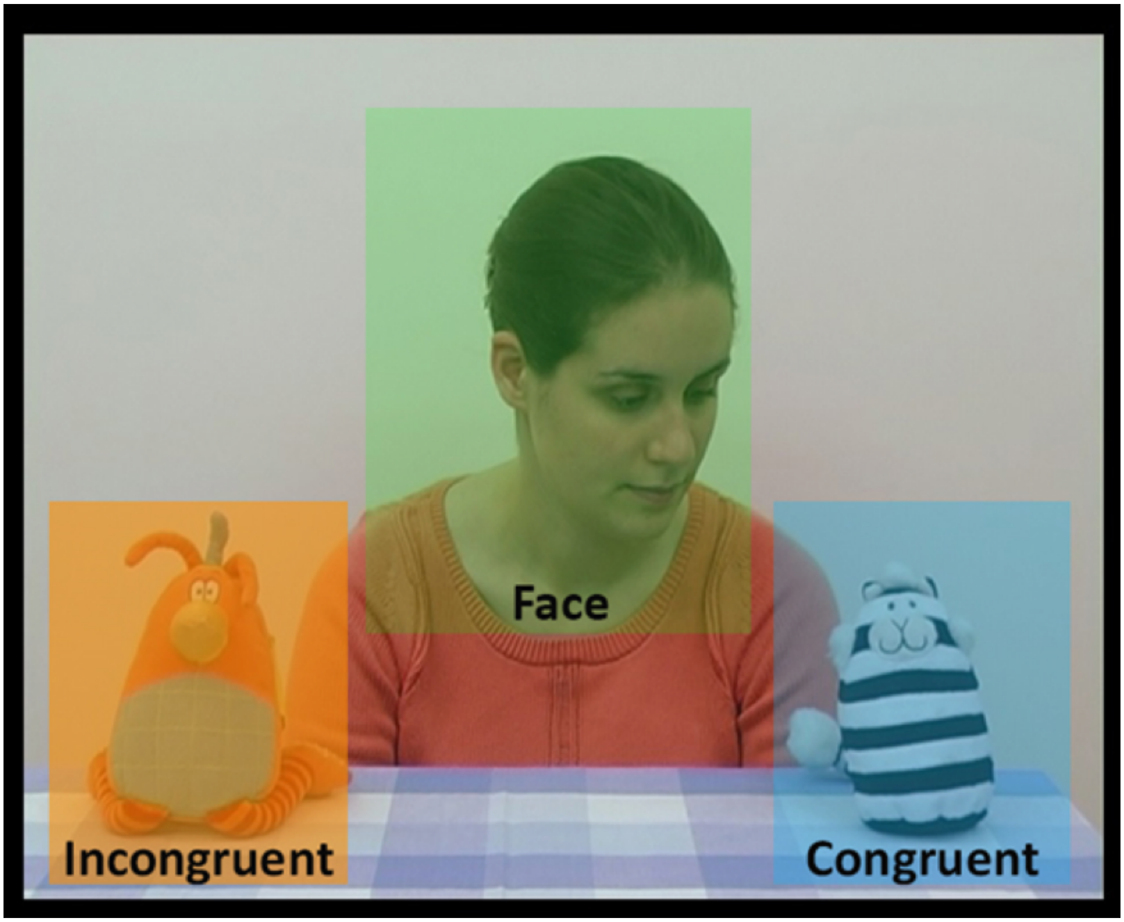

Figure 1. Screenshot of one of the videos presented to infants showing the three areas of interest (AOIs) used: face, congruent object, and incongruent object.

shift agreement ranged from 96.01-99.30% and 90.54-99.63%, respectively. (These numbers do not include the one laboratory described above whose data were excluded due to low inter-rater reliability, which obtained well below 70% agreement due to poor video quality.) One lab had additionally planned to hand-code eye-tracking data to assess the comparability of eye-tracking and human-coded data, but was unable to successfully do so due to unforeseen technical and staffing issues. Overall, offline and eye-tracking-coded data each appeared to have good reliability, although we were not able to assess the comparability of these approaches.

Results

Dependent variables

Following previous studies using this paradigm (Senju & Csibra, 2008; Szufnarowska et al., 2014), we investigated infants' gaze-following abilities via several different approaches. Each approach focused on infants' looking behaviors to the areas of interest starting from the point in time when the model started to turn her head (4 seconds – 100 frames – from the beginning of the trial) to the end of the trial (10 seconds – 250 frames – from the beginning of the trial). We measured four different dependent variables for each infant on each trial. Three measures have been used in previous studies: first look, frequency of looks, and duration of looks (Senju & Csibra, 2008; Szufnarowska et al., 2014). We included an additional measure, latency, as we reasoned that infants' reaction time to follow an actor's gaze might show interesting development over the first two years of life, and might be a potentially sensitive measure. Exploring these four variables in the context of our large sample size can provide insight for future studies about the expected effect sizes for different analytic approaches.

First look. This measured whether the infant shifted their gaze from the face AOI to one of the object AOIs. This yielded a binary variable indicating whether the infant showed a congruent gaze shift towards the actor's target (coded as 1), an incongruent gaze shift towards the other object (coded as 0), or no shift (coded as missing).

Frequency of looks. This yielded two values for each infant: the number of times the infant shifted their gaze from the face AOI to the congruent AOI, and the number of times the infant shifted their gaze from the face AOI to the incongruent AOI.

Duration of looks. This measured the total duration of fixation to the congruent AOI and to the incongruent AOI. Thus, each infant had two values. These values were log-transformed prior to analysis, following recent recommendations for looking time data (Csibra, Hernik, Mascaro, & Tatone, 2016).

Latency. This established infants' reaction times to follow the actor's gaze in milliseconds. On each trial, latency was coded as the latency of the first face-to-object shift, whether it was to such that each infant had a value on each trial for latency to look at both the congruent and incongruent AOIs. As raw latency scores were non-normal, theyscores will be were log-transformed transformed prior to analysis, following the pre-registered analysis plan.

Analysis approach

All planned analyses were pre-registered at osf.io/2ey3k/. Following previous large-scale multi-lab studies with infants (e.g., ManyBabies Consortium, 2019a; 2019b), we used two complementary data analysis frameworks: meta-analysis and mixed-effects regression. Under the meta-analytic framework, we conducted standard analyses within each lab, and then combined these results across labs. An advantage of this approach is that it is easy to understand, and is comparable to results from meta-analyses that gather data from published studies. Under the mixed-effects regression framework, we modeled raw trial-by-trial data from each infant. Because this approach models raw data directly, it can have greater statistical power to detect effects.

Confirmatory Analyses

Meta-analytic framework. Under this framework, we first calculated mean scores for each individual infant on the four dependent variables. For first look, frequency of looks, and total duration of looks, we calculated proportion difference scores for each infant, which subtracted the mean value for congruent trials (c) from the mean for incongruent trials (i), and divided by the total number of trials that contributed to that measure [(c - i)/(c + i1)]. Trials without values for a particular measure were excluded from the calculation. In the case of first looks, this involved taking the number of trials for which the first look was congruent and dividing it by the total number of trials that were coded for a first look. For latency, we limited analysis to only those trials with a congruent first look, and for the meta-analytic model we focused on the mean latency for each infant to look towards the congruent AOI. We then collapsed these for each dataset (i.e., combination of lab, bilingualism status, and age group) to calculate a grand mean (M) and standard deviation across participants (sd) in each dataset. Finally, using the formula dz = M/sd, the derived M and sd were used to compute a within-subject Cohen's d for first look, frequency of looks, and total duration of looks. For latency, we deviated from the preregistered analysis plan. As analysis was limited to latency towards the congruent AOI, it was not ideal to generate a Cohen's d effect size without a comparison between two means. Instead of computing a within-subject Cohen's d, the raw grand mean (M) and standard deviation across participants (sd) were entered into the meta-analytic model for latency. Sampling variance within each lab was calculated based on the formula sd ˆ 2/n.

Random effects meta-analysis models with a restricted maximum-likelihood estimator (REML) were fit with the metafor package (Viechtbauer, 2010). A logistic model was fit for first look, frequency of looks, and total duration of looks as each infant's score was bounded between 0 and 1. A linear model was fit for reaction time. To account for the dependence between mono- and bilingual datasets stemming from the same lab, we included laboratory as a random factor. Bilingualism was dummy coded (0 = monolingual, 1 = bilingual), and age was coded as the average age for each lab's contributed sample for each language group (centered for ease of interpretation).

Our main meta-analytic model for each dependent variable was:

dz ~ 1 + bilingual + age + bilingual * age

First look. We began by examining the relation of the proportion of first looks to bilingualism and age, fitting the main effect model to the 32 separate group means and variances (after aggregating by lab, age, and language group). The meta-analysis on first look yielded a mean effect size of 0.79 (CI = [0.28 - 1.29], $z = 3.07$, $p = .002$) for 6–9 month-old monolingual infants. Age yielded an additional effect of 0.43 (CI = [-0.17 - 1.03], $z = 1.39$, $p = .165$), although the effect was not statistically significant. With an additional bilingualism effect of 0 (CI = [-0.72 - 0.72], $z = 0$, $p = .997$), there was no significant evidence for bilingual infants to perform differently from monolingual infants. Moreover, the interaction between bilingualism and age yielded a very small effect of -0.02 (CI = [-0.91 - 0.88], $z = -0.04$, $p = .970$), suggesting that there was not a reliable difference between bilingual and monolingual infants at either age. A forest plot for this meta-analysis is shown in Figure 2.

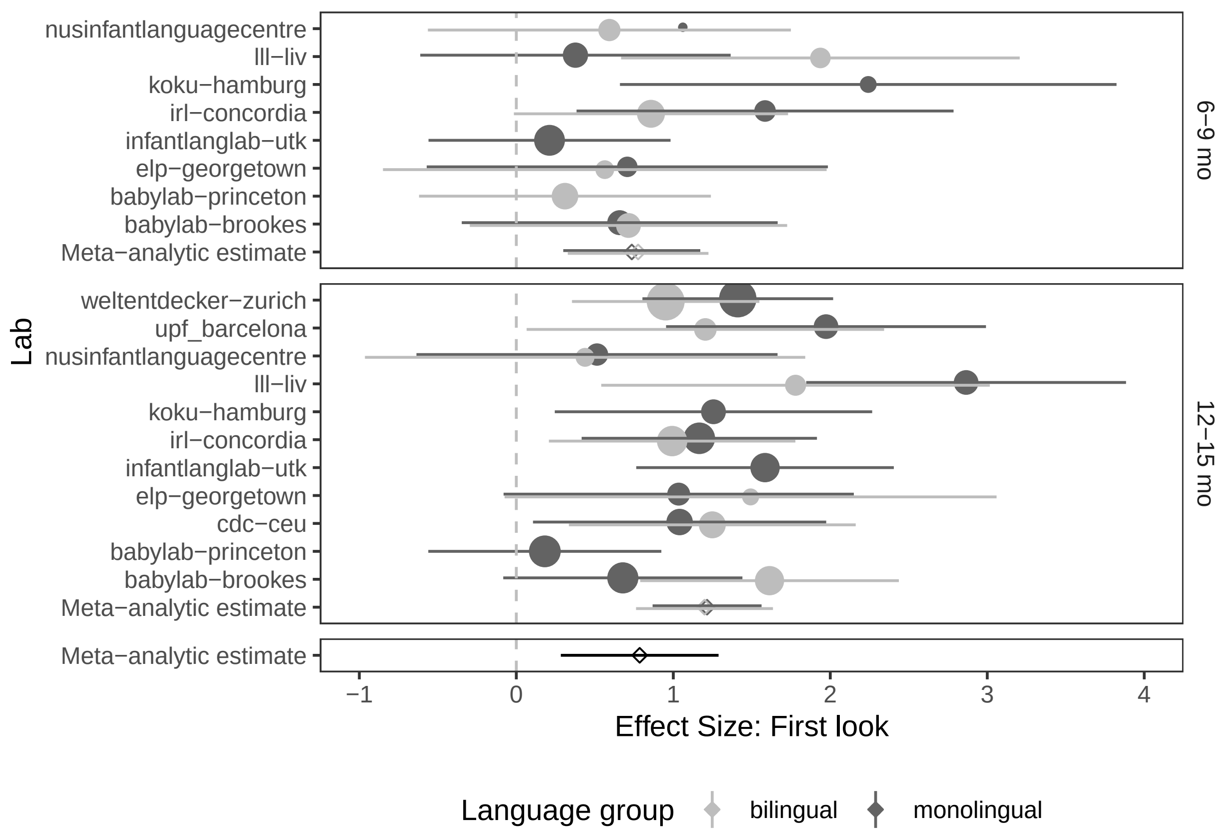

Figure 2. Forest plot for the cross-lab meta-analysis on the proportion of first look.

Frequency of looks. We then investigated the relation of frequency of looks to bilingualism and age group. The overall mean effect size for 6–9 month-old monolingual infants was 0.73 (CI = [0.22 - 1.23], $z = 2.83$, $p = .005$). Age yielded an additional effect of

0.48 (CI = [-0.13 - 1.08], $z = 1.55$, $p = .121$), which was not a statistically significant difference. There was no evidence that bilingual infants would differ from monolingual infants, as the additional effect of bilingualism was 0 (CI = [-0.72 - 0.72], $z = 0$, $p = .998$). Moreover, the interaction between bilingualism and age yielded a very small effect of 0.10 (CI = [-0.80 - 0.99], $z = 0.22$, $p = .829$), implying that no differences were observed between monolingual and bilingual infants at both ages. A forest plot for this meta-analysis is shown in Figure 3.

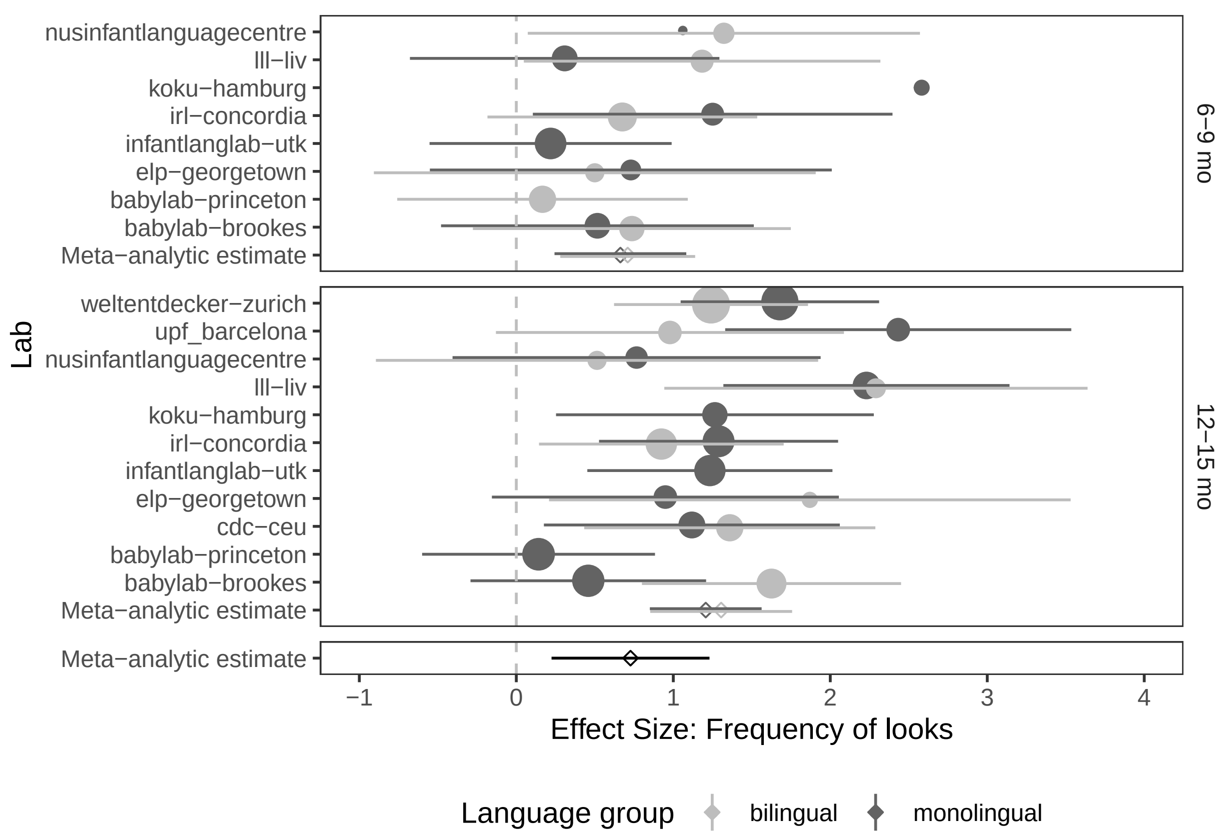

Figure 3. Forest plot for the cross-lab meta-analysis on frequency of looks.

Duration of looks. The cross-lab meta-analysis on duration of looks yielded a mean effect size for 6–9 month-old monolingual infants of 0.32 (CI = [-0.09 - 0.72], $z = 1.53$, $p = .125$). Age yielded a non-significant additional effect of 0.08 (CI = [-0.39 - 0.54], $z = 0.32$, $p = .752$), indicating 12-15 month-old infants performed similarly as 6-9 month-old infants. The additional bilingualism effect of -0.06 (CI = [-0.64 - 0.52], $z = -0.21$, $p = .837$) was also not statistically significant, suggesting that bilingual infants did not look significantly longer compared to monolingual infants. Moreover, the interaction between bilingualism and age yielded a very small effect of 0.08 (CI = [-0.62 - 0.77], $z = 0.22$, $p = .824$), suggesting no evidence of differences between monolingual and bilingual infants across both ages. A forest plot for this meta-analysis is shown in Figure 4.

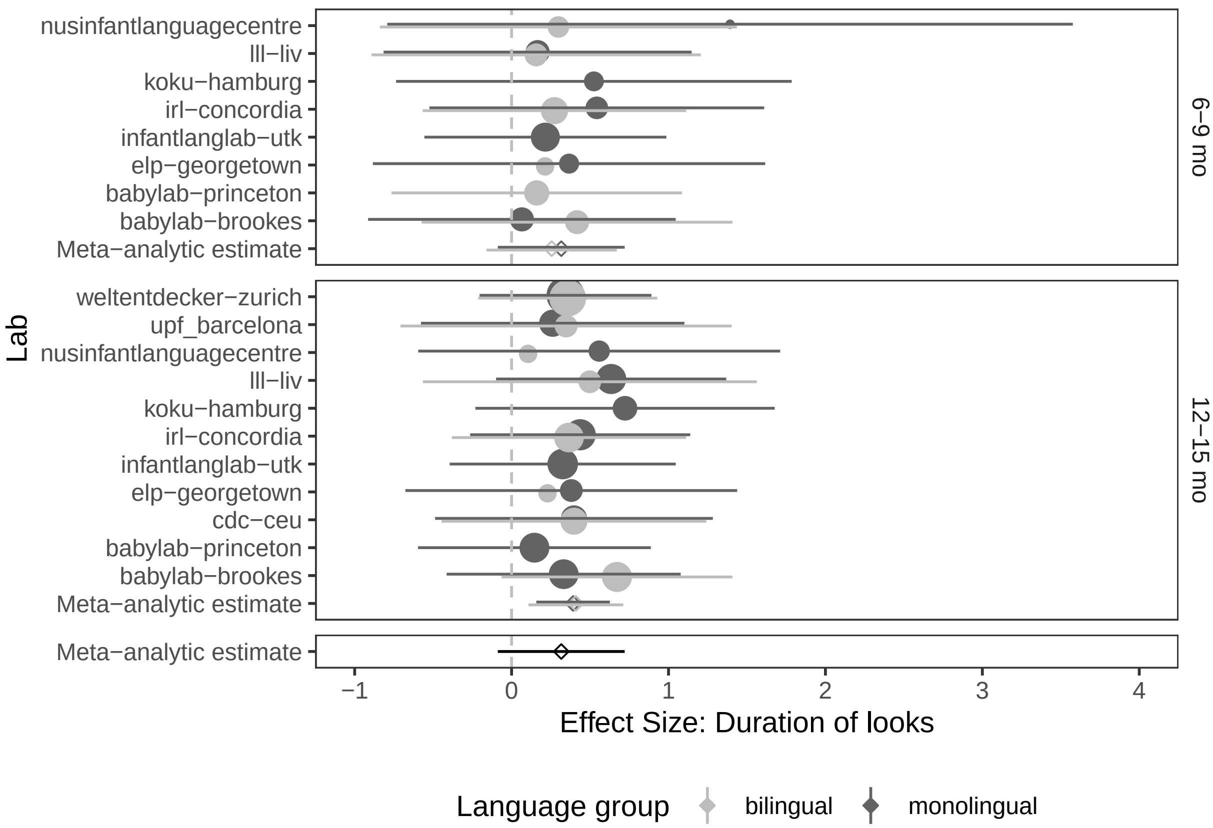

Figure 4. Forest plot for the cross-lab meta-analysis on duration of looks.

Latency. The cross-lab meta-analysis on latency towards the congruent object yielded an overall mean of 2,345.76 (CI = [2,056.47 - 2,635.06], $z = 15.89$, $p =< .001$) for 6–9 month-old monolingual infants. With the effect of age, the mean latency decreased significantly by -493.06 (CI = [-835.03 - -151.09], $z = -2.83$, $p = .005$), where 12-15 month-old infants were more rapid than 6-9 month-old infants in terms of fixating their looks towards the congruent object. The effect of bilingualism increased the mean latency by 378.29 (CI = [-26.76 - 783.34], $z = 1.83$, $p = .067$), with a trend of bilingual infants being slower than monolingual infants in the latency of their fixation. The interaction between bilingualism and age suggested a tendency that the effect of bilingualism on latency was different at the two ages (estimate = -437.30, CI = [-930.57 - 55.97], $z = -1.74$, $p = .082$). Pairwise comparisons revealed that at 6-9 months the bilinguals tended to be slower than the monolinguals by -378.29 ($se = 206.66$, $z = -1.83$, $p = .067$). However, at the age of 12-15 months, there were no longer evidence of a difference in latency of their fixation between monolingual and bilingual infants (estimate = 59.01, $se = 143.63$, $z = 0.41$, $p = .681$). A forest plot for this meta-analysis is shown in Figure 5.

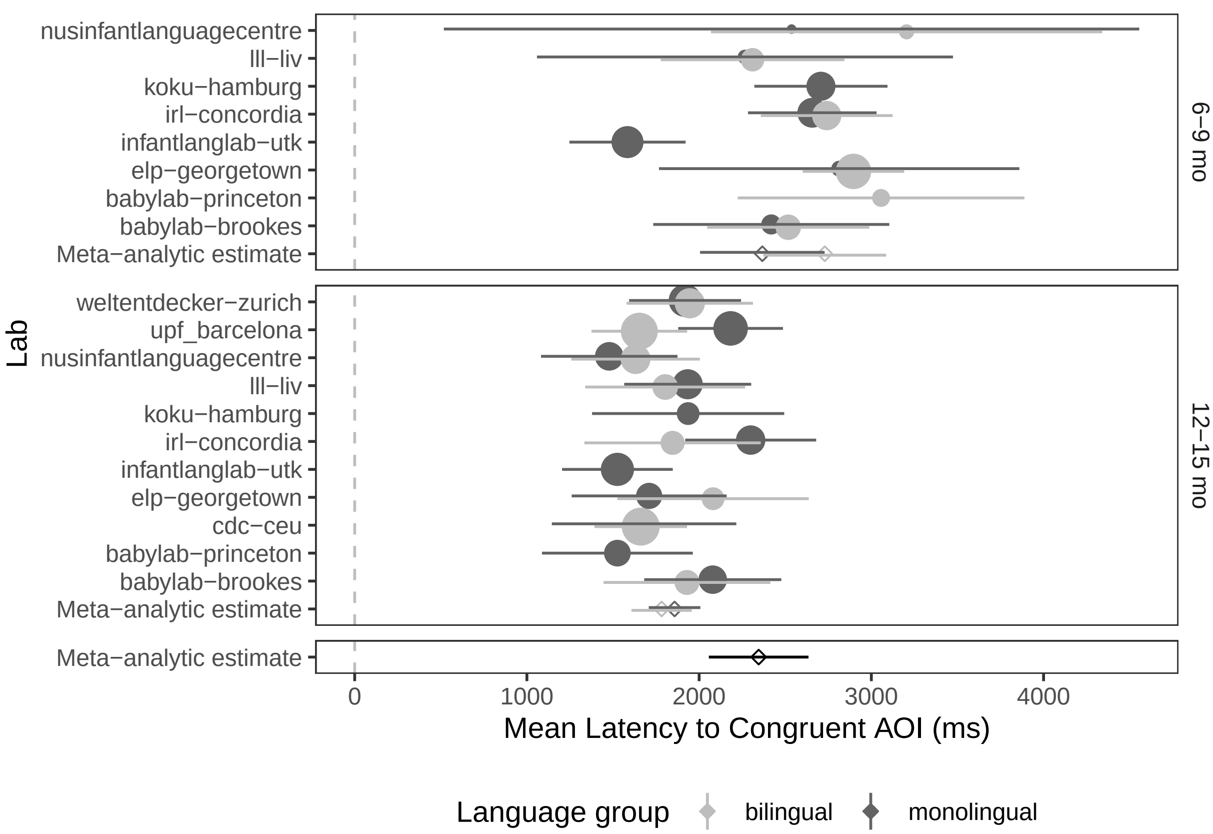

Figure 5. Forest plot for the cross-lab meta-analysis on latency.

Summary of meta-analysis. Overall, our meta-analytic models revealed that infants followed the actor's gaze to the congruent object at above-chance levels measured by first looks and frequency of looks. In general, the measures of first looks and frequency of looks appeared to be more sensitive measures than duration of looks in assessing infants' gaze-following accuracy. Moreover, these two measures revealed a medium effect size for age, although the effect was not statistically significant. This implied that gaze-following development could be modulated by age-related changes, such that gaze-following ability improved with age. However, we observed no evidence of bilingualism as a statistically significant moderator of gaze-following development, suggesting that bilingual infants tend to perform similarly to monolingual infants. On the other hand, the meta-analytic model of latency revealed a significant age effect, where latency of fixation decreases with age. The latency model also found a trend for bilingual infants to be slower than monolinguals in terms of fixating their looks towards the congruent object at a younger age of 6-9 months, but not at an older age of 12-15 months. Together, all these results implied that older infants may show more reliable gaze-following than younger infants.

Mixed-effects regression framework

As opposed to the meta-analytic framework, the mixed-effects regression framework allowed us to model trial-level data from individual infants rather than analyzing averages. Mixed-effects models are described as such because they include both fixed effects and random effects. Our fixed effects modeled the main variables of interest: age and bilingualism. Our random effects accounted for correlations in the data that could arise due to dependency between data from the same infants, lab, and test items. For each model, we planned to initially fit a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013), while anticipating the need for pruning. We aimed to identify a pruned random-effects structure that would be well-supported by our data, while conserving the most theoretically important effects (Matsuchek, Kleigl, Visishth, Baayen, & Bates, 2017).

The approach to pruning random effects was somewhat exploratory, as we did not have a specific hypothesis about the random effects. Note that while the particular random effects structure of the model can affect the estimates of standard errors, in a balanced design it does not affect the estimates of the fixed effects, which were our main interest.

Our dependent variables (DV) were:

- *first_shift*: A binary variable denoting the AOI of the first shift, where 0 is the incongruous object and 1 is the congruous object.
- *latency*: The time interval in milliseconds between the onset of the actor's headturn, and the moment of first fixation on an object AOI.
- *freq_shift*: The number of times in the trial an infant shifted gaze towards the AOI.
- *total_look*: The total duration of fixations towards the AOI during the trial.

Our predictor variables were:

- *bilingual*: A dummy-coded variable where 0 is monolingual, 1 is bilingual.
- *age_days*: The infant's age in days, scaled and centred for ease of interpretation.
- *aoi*: A dummy-coded variable for analysis of freq_shift, total_look, and latency, for which data from both AOIs are reported. Here, 0 denotes the congruous AOI, and 1 denotes the incongruous AOI.

We ran separate models for each DV. We fit all models using the `lme4` package (Bates et al., 2015). For first_shift, we fit a logistic model as this variable is binary at the trial level. The initial model specification was:

first_shift ~ bilingual * age_days + (1|subid) + (bilingual * age_days|lab) + (bilingual * age_days|item)

Table 2

Coefficient estimates from a logistic mixed effects model predicting the probability of making first looks to congruent objects.

	Estimate	*SE*	*z*	*p*
Intercept	0.971	0.105	9.270	0.000
bilingual	-0.010	0.126	-0.078	0.938
age_days	0.197	0.079	2.500	0.013
bilingual * age_days	-0.096	0.123	-0.779	0.436

For latency, freq_shift, and total_look, we used a similar model with two modifications. First, we fit a linear model rather than a logistic model as these variables are continuous and unbounded. Second, we included an interaction with aoi in the fixed effects, and estimated corresponding random slopes where appropriate. This was necessary in order to estimate separate parameters for the congruous and incongruous AOIs (i.e., to model whether latency to first fixation varies as a function of whether it is to the congruous or incongruous AOI; whether infants shift more frequently to the congruous than the incongruous AOI; and whether infants fixate more on the congruous than incongruous AOI). For these three DVs, the initial model specification was:

DV ~ bilingual * age_days * aoi + (aoi|subid) + (bilingual * age_days * aoi|lab) + (bilingual * age_days * aoi |item)

First shift towards the AOI. Our final logistic model specification for first shift was:

first_shift ~ bilingual * age_days + (1|subid) + (1|lab)

Table 2 shows coefficient estimates from this model and Figure 6 visualizes this model.

Positive coefficients indicate a higher probability of making first look to the congruent object. A significant positive effect of age was found, suggesting that older infants had a greater probability of making their first looks to the congruent as opposed to the incongruent object. While there is no obvious evidence for a difference between monolingual and bilingual infants, the interaction of bilingualism and age was also not significant. Monolingual and bilingual infants, therefore, did not differ in their probabilities of first look across ages.

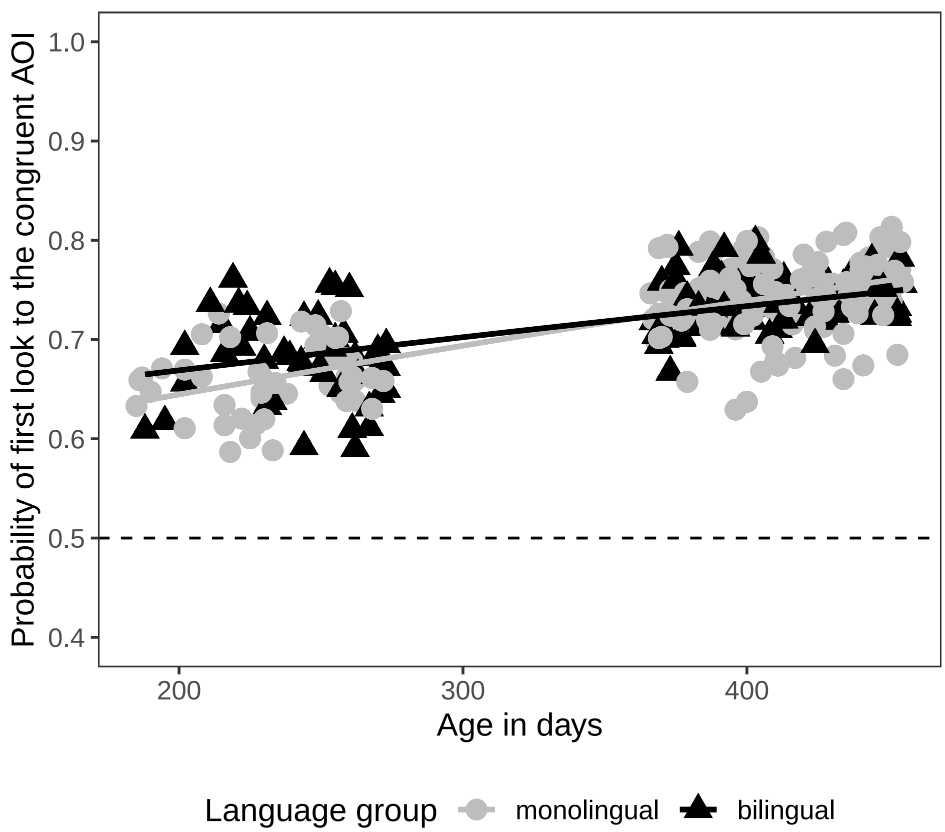

Figure 6. The logistic regression model predicting the probability of making first look to the congruent object, plotted with individual participants' probabilities.

Frequency of shifting gaze towards the AOI. The final model specification for frequency of shift was:

freq_shift ~ bilingual * age_days * aoi + (1|subid) + (1|item)

Table 3

Coefficient estimates from a linear mixed effects model predicting frequency of shifting gaze towards the congruent AOI.

	Estimate	SE	t	p
Intercept	1.160	0.041	28.500	0.000
bilingual	0.066	0.039	1.700	0.090
age_days	0.087	0.025	3.460	0.001
aoi	-0.626	0.035	-17.800	0.000
bilingual * age_days	0.069	0.039	1.790	0.074
bilingual * aoi	-0.054	0.055	-0.972	0.331
age_days * aoi	-0.104	0.036	-2.920	0.003
bilingual * age_days * aoi	-0.029	0.055	-0.525	0.599

Table 3 shows coefficient estimates from this model and Figure 7 visualizes this model. The significant main effect of age indicated that older infants looked more frequently at the objects as opposed to younger infants. The interaction between age and bilingualism, although not significant, indicated a tendency for bilingual infants to increase their frequency of looks towards the objects with age, but not for monolingual infants. More centrally, there was also a significant main effect of aoi, where overall infants shifted more often to the congruent object as opposed to the incongruent object. The interaction of age and aoi was also significant, further suggesting that older infants looked more frequently to the congruent than the incongruent object compared to younger infants. The effect of bilingualism, however, was not significant, neither were its 2-way interaction with aoi nor its 3-way interaction with age and aoi; this suggests that there was not a reliable difference between bilingual and monolingual infants in the number of times they shifted gaze towards the congruent object.

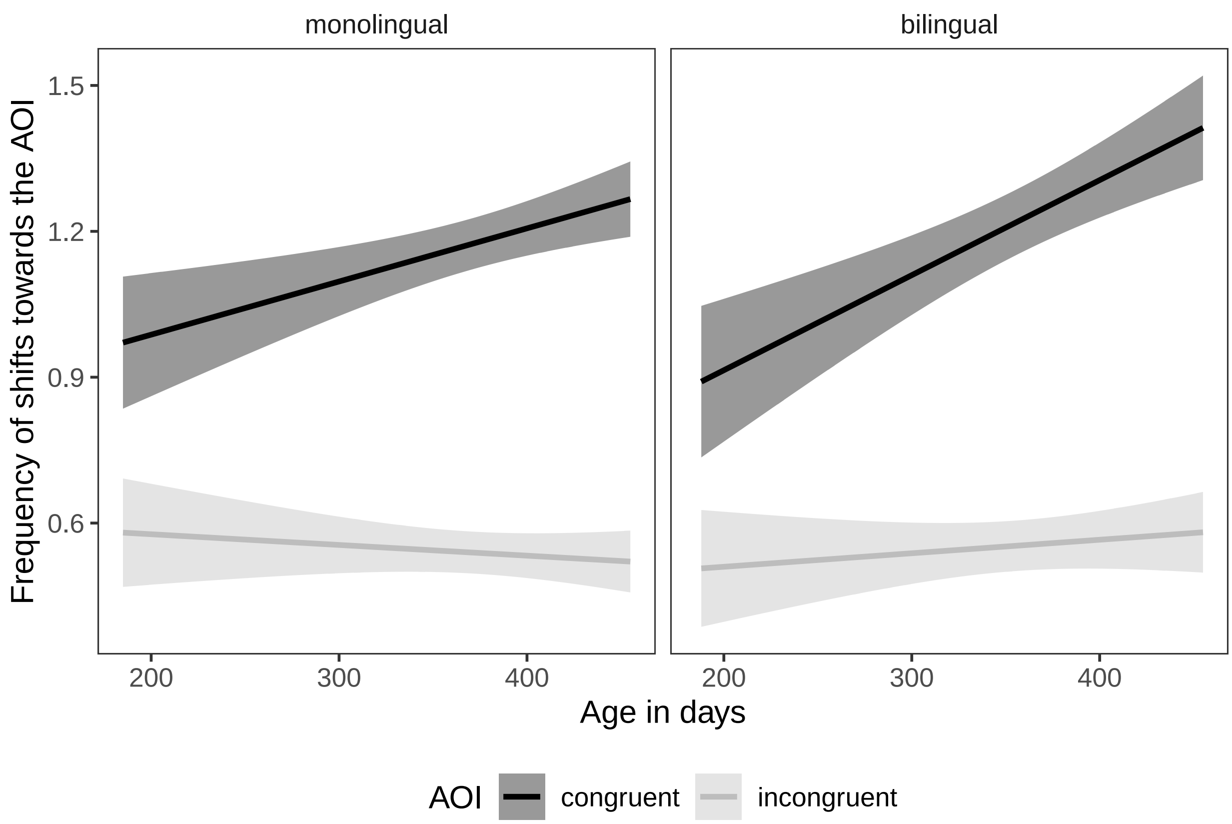

Figure 7. The linear regression model predicting the frequency of shift towards the AOI, with error bars showing 95% confidence interval.

Duration of fixations towards the AOI during the trial. The final model specification for duration of fixations was:

total_look ~ bilingual * age_days * aoi + (1|lab) + (1|item)

Table 4 shows coefficient estimates from this model and Figure 8 visualizes this model. The significant effect of age indicated that in general older infants looked longer at the objects than did the younger infants. The main effect of aoi was also significant, suggesting that infants looked longer at the congruent object than at the incongruent object. While the effect of bilingualism was significant, it was suggested that overall bilingual infants looked longer at the objects than did the monolingual infants. No significant interactions were observed.

Table 4

Coefficient estimates from a linear mixed effects model predicting duration of fixations towards the AOI during the trial.

	Estimate	*SE*	*t*	*p*
Intercept	3.950	0.147	26.900	0.000
bilingual	0.368	0.155	2.380	0.018
age_days	0.004	0.001	3.340	0.001
aoi	1.690	0.135	12.500	0.000
bilingual * age_days	0.003	0.002	1.590	0.111
bilingual * aoi	-0.225	0.213	-1.060	0.289
age_days * aoi	0.000	0.002	0.114	0.909
bilingual * age_days * aoi	-0.001	0.003	-0.501	0.616

Latency. The final model specification for latency was:

latency ~ bilingual * age_days * aoi + (1|subid) + (1|lab) + (1|item)

Table 5 shows coefficient estimates from this model and Figure 9 visualizes this model. The main effect of age was significant, suggesting that older infants were more rapid than younger infants in fixating their first look to the objects. Moreover, the interaction between age and bilingualism suggested a tendency for both bilinguals and monolinguals to make faster fixation with age. The interaction between age and aoi also pointed to a trend where, compared to younger infants, older infants were more rapid at making their first fixation at both the congruent and incongruent objects. However, the effect of aoi itself was not significant, implying that in general infants did not differ in latency of their first fixation towards the congruent or incongruent objects. Other coefficients were also non-significant.

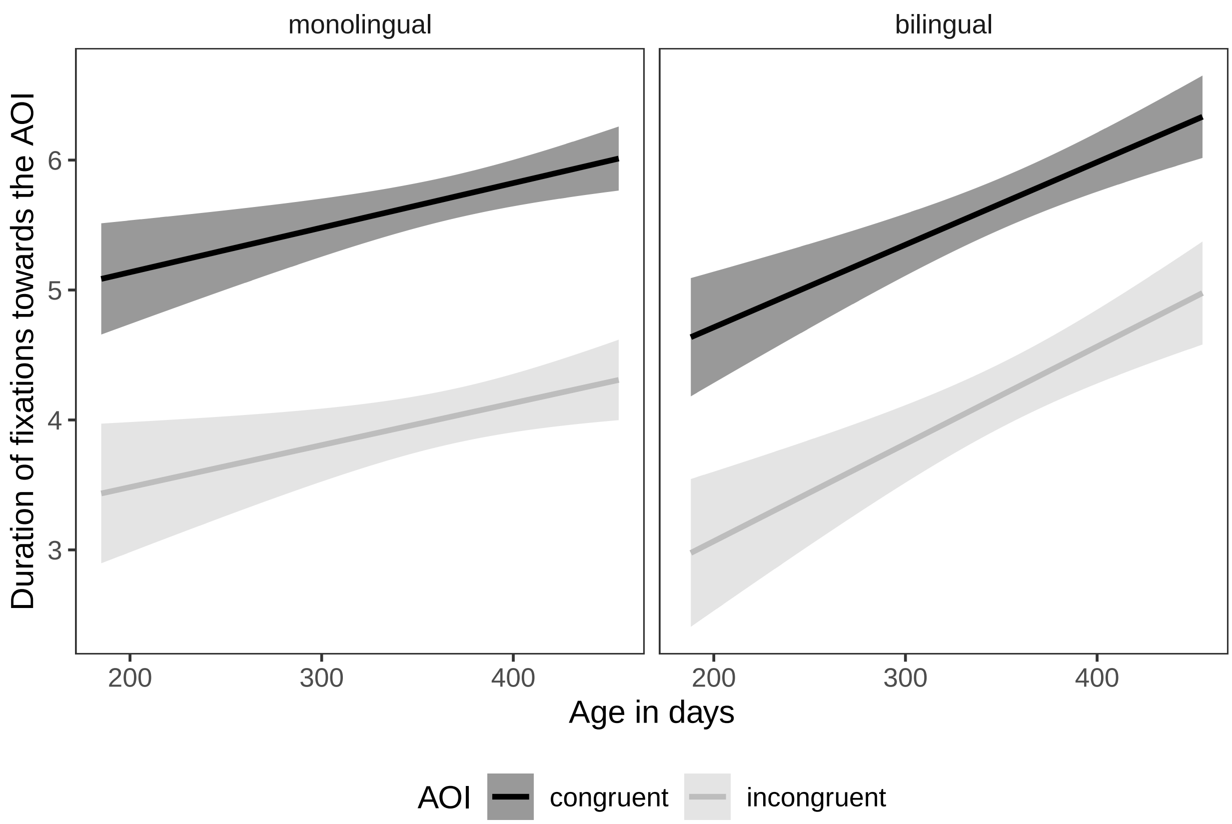

Figure 8. The linear regression model predicting duration of fixations towards the AOI, with error bars showing 95% confidence interval.

Summary of mixed-effects regression. Overall, our mixed-effects regression revealed that early gaze-following development is significantly modulated by age-related changes, where older infants showed a more reliable gaze-following ability than younger infants. In particular, older infants were shown to be more accurate and rapid than younger infants in directing their first gaze towards the congruent objects. Older infants also looked longer and more frequently at the congruent objects than at the incongruent objects. In contrast, bilingualism did not significantly predicted infants' gaze-following latency and accuracy, although there was a trend where older bilingual infants would look more often at both the congruent and incongruent objects than younger bilinguals. Bilingual significantly predicted only infants' duration of fixations, where bilingual infants

Table 5

Coefficient estimates from a linear mixed effects model predicting latency between the onset of the actor's headturn and the moment of first fixation on an object AOI.

	Estimate	*SE*	*t*	*p*
Intercept	7.400	0.051	146.000	0.000
bilingual	-0.002	0.055	-0.034	0.973
age_days	-0.002	0.000	-3.470	0.001
aoi	0.048	0.056	0.860	0.390
bilingual * age_days	-0.001	0.001	-1.820	0.069
bilingual * aoi	0.074	0.087	0.846	0.398
age_days * aoi	-0.001	0.001	-1.790	0.073
bilingual * age_days * aoi	0.001	0.001	0.810	0.418

looked longer at both the congruent and incongruent objects compared to monolingual infants. Together, these findings suggested that, regardless of bilinguals' more frequent and longer fixation, monolingual and bilingual infants appear to follow a similar trajectory of gaze-following development despite the difference in their language experience.

Exploratory analyses

TO BE WRITTEN AFTER THE STUDY IS COMPLETED

General Discussion

XYZ

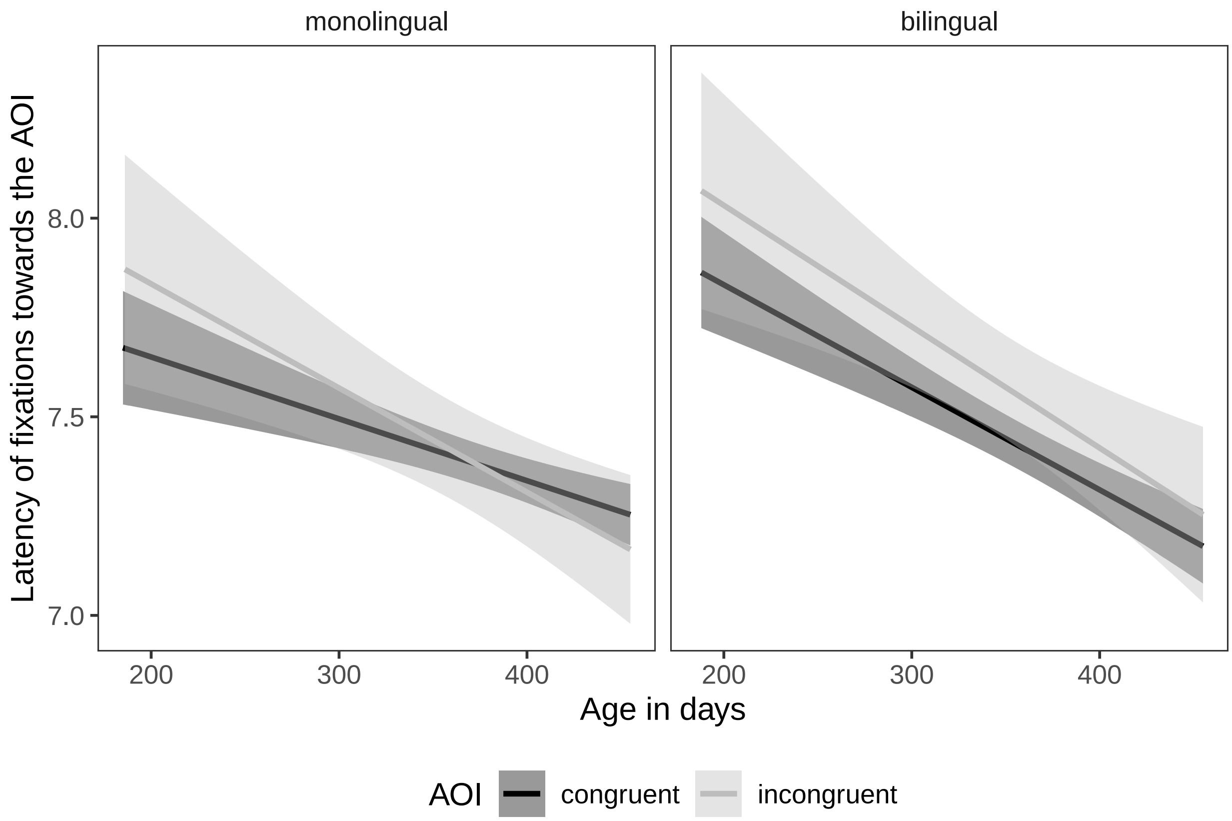

Figure 9. The linear regression model predicting latency of fixations towards the AOI, with error bars showing 95% confidence interval.

Summary of Findings

Challenges and Limitations

XYZ

Conclusion

Author Contributions

Author contribution initials reflect authorship order. KBH, LS contributed to the study concept. KBH, RB, CLW, LS contributed to the study design. KBH contributed to the final protocol. KBH contributed to study documentation. KBH, LS contributed to study management. KBH, RB, AB, SD, AG, NGG, MH, MJó, ALR, CLW, UL, LL, CN, CEP, JRH, MS, CW, LS contributed to data collection. KBH, AKB, MJó, MS, RKYT, DR, IV contributed to data management and analysis. KBH, RB, AKB, NGG, CLW, IV, LS contributed to the Stage 1 manuscript. KBH, LS contributed to the Stage 2 manuscript.

Conflicts of Interest

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

Individual participating labs acknowledge funding support from: XYZ

Disclosures

Preregistration

Our manuscript was reviewed prior to data collection; in addition, we registered our instructions and materials prior to data collection osf.io/2ey3k/.

Data, materials, and online resources

All materials, data, and analytic code are available at osf.io/2ey3k/.

Reporting

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

References

Antovich, D. M., & Graf Estes, K. (2017). Learning across languages: Bilingual experience supports dual language statistical word segmentation. *Developmental Science*, e12548. doi: 10.1111/desc.12548

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P.J. Dunham (Eds.), Joint attention: Its origins and role in development (pp. 131–58). Hillsdale, NJ: Erlbaum.

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, 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi: 10.18637/jss.v067.i01.

Bedford, R., Elsabbagh, M., Gliga, T., Pickles, A., Senju, A., Charman, T.,. . . , & the BASIS team. (2012). Precursors to social and communication difficulties in infants at-risk for autism: Gaze-following and attentional engagement. Journal of Autism and Developmental Disorders, 42, 2208–2218. doi: 10.1007/s10803-012-1450-y

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. Proceedings of the National Academy of Sciences, 109, 3253–3258. doi: 0.1073/pnas.1113380109

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. Cognition, 140(C), 122–134. doi: 10.1016/j.cognition.2015.03.013

Bloom, P. (2000). How Children Learn the Meanings of Words. Cambridge, MA:The MIT Press.

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

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

Brojde, C. L., Ahmed, S., & Colunga, E. (2012). Bilingual and monolingual children attend to different cues when learning new words. Frontiers in Psychology, 3. Doi: 10.3389/fpsyg.2012.00155

Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. Developmental Psychology, 38, 958–66. doi: 10.1037/0012-1649.38.6.958

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze-following and its relation to language. Developmental Science, 8, 535–43. doi: 10.1111/j.1467-7687.2005.00445.x

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. Journal of child language, 35, 207-220. doi: 10.1017/S030500090700829

Brooks, R., & Meltzoff, A. N. (2014). Gaze-following: A mechanism for building social connections between infants and adults. In M. Mikulincer & P. R. Shaver (Eds.), Mechanisms of social connection: From brain to group (pp. 167–183). Washington, DC: American Psychological Association.

Byers-Heinlein, K. (2015). Methods for studying infant bilingualism. In J. W. Schwieter, (Ed.). Cambridge Handbook of Bilingual Processing (pp. 133–154). Cambridge, UK: Cambridge University Press.

Byers-Heinlein, K. (2016). Bilingualism affects 9-month-old infants' expectations about how words refer to kinds. Developmental Science, 20, e12486. doi: 10.1111/desc.12486

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

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word learning heuristic. Developmental Science, 12, 815-823. doi: 10.1111/j.1467-7687.2009.00902.x

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. Cognition, 128, 407–416. doi: 10.1016/j.cognition.2013.05.010

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monographs of the Society for Research in Child Development, 63, i-174. doi: 10.2307/1166214

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish-English bilinguals from 22 to 30 months: Implications for assessment. Journal of Speech, Language and Hearing Research, 56, 1637–1649. doi: 10.1044/1092-4388(2013/11-0044)

Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P. J. Dunham (Eds.), Joint attention: Its origins and role in development (pp. 61– 83). Hillsdale, NJ: Erlbaum.

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

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? Psychological Science, 26, 99–107. doi:

10.1177/0956797614557866

De Houwer, A., Bornstein, M. H., & De Coster, S. (2006). Early understanding of two words for the same thing: A CDI study of lexical comprehension in infant bilinguals. International Journal of Bilingualism, 10, 331–347. doi: 10.1177/13670069060100030401

Duñabeitia, J. A., & Carreiras, M. (2015). The bilingual advantage: Acta est fabula. Cortex, 73, 371–372. doi: 10.1016/j.cortex.2015.06.009

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze-following in newborns. Infancy, 5, 39–60. doi: 10.1207/s15327078in0501_2

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, 1510–1525. doi: 10.1111/j.1467-8624.2007.01080.x

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). MacArthur-Bates Communicative Development Inventories (2nd ed.). Baltimore, MD: Brookes.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., . . . & Yurovsky, D. (2017). A Collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. Infancy, 1-15 doi: 10.1111/infa.12182

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze-cueing of attention: Visual attention, social cognition and individual differences. Psychological Bulletin, 133, 694-724. doi:10.1037/0033-2909.133.4.694

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

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. Nature, 466, 29–29. doi: 10.1038/466029a

Hessels, R. S., Andersson, R., Hooge, I. T., Nyström, M., & Kemner, C. (2015).

Consequences of eye Color, positioning, and head movement for eye-tracking data Quality in Infant Research. Infancy, 20(6), 601-633.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., . . . & Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. Monographs of the Society for Research in Child Development, 65, i-35. doi: 10.1111/1540-5834.00090.

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. Infancy, 15, 125–150. doi:10.1111/j.1532-7078.2009.00009.x

Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. Perspectives on Psychological Science, 7, 645–654. doi: 10.1177/1745691612464056

Kandhadai, P., Danielson, D.K., & Werker, J.F. (2014). Culture as a binder for bilingual acquisition. Trends in Neuroscience and Education, 3, 24-7. doi: 10.1016/j.tine.2014.02.001

Klein, R. M. (2015). Is there a benefit of bilingualism for executive functioning? Bilingualism: Language and Cognition, 18, 29–31. doi: 10.1017/S1366728914000613

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

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

Liu, L., & Kager, R. (2016). Perception of a native vowel contrast by Dutch monolingual and bilingual infants: A bilingual perceptual lead. International Journal of Bilingualism, 20(3), 335-345. doi: 10.1177/1367006914566082

ManyBabies Consortium. (2017a; Registered Report accepted pending data collection). Quantifying sources of variability in infancy research using the infant-directed speech preference. Advances in Methods and Practices in Psychological Science.

ManyBabies Consortium. (2017b; Registered Report under review.). A multi-lab study of bilingual infants: Exploring the preference for infant-directed speech. Advances in Methods and Practices in Psychological Science.

Markman, E. M. (1990). Constraints children place on word meanings. Cognitive Science, 14, 57-77. doi: 10.1207/s15516709cog1401_4

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. Cognitive Psychology, 20, 121–57. doi: 10.1016/0010-0285(88)90017-5

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. doi: 10.1016/j.jml.2017.01.001

Meltzoff, A. N., & Brooks, R. (2007). Eyes wide shut: The importance of eyes in infant gaze-following and understanding other minds. In R. Flom, K. Lee & D. Muir (Eds.), Gaze-following: Its development and significance, 217–41. Mahwah, NJ: Erlbaum.

Moore, C. (2008). The development of gaze-following. Child Development Perspectives, 82, 661–673. doi: 10.1111/j.1750-8606.2008.00052.x

Moore, C., & Corkum, V. (1998). Infant gaze-following based on eye direction. British Journal of Developmental Psychology, 16, 495–503. doi: 10.1111/j.2044-835x.1998.tb00767.x

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Neal, R. & Schwartz, H. K. (2000). Gaze-following, temperament, and language development in 6-month-olds: A replication and extension. Infant Behavior & Development, 23, 231–6. doi: 10.1016/s0163-6383(01)00038-8

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A.V., & Parlade, M.V. (2007). Individual differences and the development of joint attention in infancy. Child Development, 78, 938–954. doi: 10.1111/j.1467-8624.2007.01042.x

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed processing model of joint attention, social-cognition and autism. Autism Research, 2, 2–21. doi:10.1002/aur.61

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. Cortex, 69, 265–278. doi: 10.1016/j.cortex.2015.04.014

Paulus, M., & Fikkert, P. (2014). Conflicting social cues: 14- and 24-month-old infants' reliance on gaze and pointing cues in word learning. Journal of Cognition and Development, 15, 43–59. doi: 10.1080/15248372.2012.698435

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. Language Learning, 43, 93–120. doi: 10.1111/j.1467-1770.1993.tb00174.x

Petitto, L. A., Katerelos, M., Levy, B. G., Gauna, K., Tetreault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. Journal of Child Language, 28, 453–496. doi: 10.1017/S0305000901004718

Pons, F., Bosch, L., Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. Psychological Science, 26, 490–498. doi: 10.1177/0956797614568320.

Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. NeuroReport, 15, 2553–2555. doi:10.1097/00001756-200411150-00025 van Renswoude, D. R., Raijmakers, M. E., Koornneef, A., Johnson, S. P., Hunnius, S., & Visser, I. (2017). Gazepath: An eye-tracking analysis tool that accounts for individual differences and data quality. Behavior Research Methods, 1-19.

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, 994–999.

Senju, A., & Csibra, G. (2008). Gaze-following in human infants depends on communicative signals. Current Biology, 18, 668–671. doi: 10.1016/j.cub.2008.03.059

Senju, A., Tucker, L., Pasco, G., Hudry, K., Elsabbagh, M., Charman, T., & Johnson, M. H. (2013). The importance of the eyes: communication skills in infants of blind parents. Proceedings of the Royal Society B: Biological Sciences, 280, 1-7. doi: 10.1098/rspb.2013.0436

Singh, L., Fu, C. S. L., Tay, Z.W., & Golinkoff, R. M. (2017) Novel word learning in bilingual and monolingual infants: Evidence for a bilingual advantage. Child Development. doi: 10.1111/cdev.12747

Singh, L., Fu, C. S., Rahman, S. A., Hameed, W., Sanmugam, S., Agarwal, P.,..& Rifkin-Graboi, A. (2015). Back to Basics: A bilingual advantage in infant visual habituation. Child Development, 86, 294–302. doi: 10.1111/cdev.12271

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? Nature Publishing Group, 4, 1–4. doi: 10.1038/srep05304

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Shah, R. J., Malle, B. F. & Morgan, J. L. (2015). Attention to the mouth and gaze-following in infancy predict language development. Journal of Child Language, 42, 1173–1190.doi: 10.1017/s0305000915000501

Tomasello, M. (2003). Constructing a language: A usage-based theory of language acquisition. Cambridge, MA: Harvard University Press.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36, 1-48. doi: 10.18637/jss.v036.i03

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. Behavior Research Methods, 45(1), 229-250.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. Science, 316, 1159. doi: 10.1177/0956797612436817

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

Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. Developmental Science, 6, 297–311. doi: 10.1111/1467-7687.00286

Yow, W. Q., & Markman, E. M. (2011). Young bilingual children's heightened sensitivity to referential cues. Journal of Cognition and Development, 12, 12–31. doi:10.1080/15248372.2011.539524

Yow, W. Q., Li, X., Lam, S., Gliga, T., Chong, Y. S., Kwek, K., & Broekman, B. F. P. (2017). A bilingual advantage in 54-month-olds' use of referential cues in fast mapping. Developmental Science, 20. doi:10.1111/desc.12482