Effects of Bilingualism on Children’s Use of Social Cues in Word Learning

W. Quin Yow, Xiaoqian Li, Sarah Lam, Teodora Gliga, Kok Hian Tan, Lynette Shek, Fabian Yap, Yap Seng Chong, and Birit F. P. Broekman

1. Introduction

Research has demonstrated that children are attentive to social cues like eye gaze and pointing gestures and use these cues to rapidly build word-referent mappings from the early stages of language development (e.g., Baldwin, 1993; Brooks & Meltzoff, 2002; Diesendruck & Markson, 2001). Six-month-olds, for example, would follow the direction of an adult’s gaze to an object in the presence of complementary signals such as eye contact and infant-directed speech (Senju & Csibra, 2008). By the end of the first year, infants begin to understand human gaze and pointing as social and goal-directed actions (Beier & Spelke, 2012; Senju, Csibra, & Johnson, 2008; Woodward, 2003), while older toddlers and preschoolers are able to use social cues to identify referential mappings in social learning contexts, demonstrating a better understanding of the communicative intention of these social cues (Behne, Carpenter, & Tomasello, 2005; Berman, Chambers, & Graham, 2010; Jaswal, 2004).

Children also rely on other information such as linguistic and perceptual cues in addition to social cues to select the referent of a new word, as suggested by the emergentist coalition model (ECM; Golinkoff & Hirsh-Pasek, 2006; Hollich et al., 2000). The ECM proposes that children are sensitive to multiple
word-learning cues but the relative importance of these cues change over the course of development; children beginning to learn words weigh perceptual cues (e.g., the perceptual salience of objects) more than social cues and only later would they assign more weight to social cues to learn words instead. This claim is supported by studies that directly pit social cues against perceptual salience and investigate which cue children use to form a word-object mapping. For example, using an eye-tracking paradigm, 10-month-old infants were found to ignore social cues and always map a new word to a more salient object versus a less salient one (Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006). In contrast, Houston-Price, Plunkett, and Duffy (2006; Experiment 4) found that 15-month-olds used gaze over salience cues to learn new words in a looking-while-listening task. When children’s explicit responses (pointing/reaching) were measured as an index of learning, only 24-month-olds, but not 18-month-olds, succeeded in using gaze to learn the word-object mapping (Moore, Angelopoulos, & Bennett, 1999).

Despite the general developmental change in the use of a speaker’s social cues over perceptual salience in word learning, there are other factors that may affect the development of these word-learning strategies. On one hand, social and communication deficits may impair children’s ability to make use of social cues for learning (e.g., Gliga, Elsabbagh, Hudry, Charman & Johnson, 2012; Imafuku et al., 2016). In an eye-tracking study (Gliga et al., 2012), 3-year-olds at risk for autism were found to follow a speaker’s gaze to the referred object in spite of the presence of a salient distractor but failed to learn the word-object mapping, whereas typically developing 3-year-olds not only followed the gaze but also formed the mapping. On the other hand, enhanced ability to understand the underlying communicative intent of referential cues may facilitate children’s use of social cues. For instance, 3- to 4-year-old bilingual children were more likely to use a speaker’s social cues (i.e., gaze and pointing) over mutual exclusivity constraints to identify references compared to their monolingual peers who showed a preference in using linguistic cues (Healey & Skarabela, 2008; Yow & Markman, 2007). This heightened sensitivity to social cues in bilinguals is likely because bilingual children interacting with people who speak different languages are tuned to social cues from the interlocutors in order to manage their more demanding communicative environment. However, no studies have investigated whether bilinguals’ sensitivity to social cues extends to a referential word-learning task where attention to other’s gaze is required to form correct word-referent mappings in the presence of other conflicting cues (e.g., perceptual salience).

This present study seeks to investigate the effect of bilingualism on children’s use of social cues to learn new words. We modified the novel word learning paradigm by Gliga et al. (2012), which was designed to examine both how children follow a speaker’s referential eye gaze (i.e., gaze-following) during learning and how they make use of such cues to establish new word-object mappings (i.e., word-learning). Monolingual and bilingual 4.5-year-olds were shown video clips involving pairs of novel objects. One of the objects was
perceptually salient (e.g., having changing colors or moving parts), and the other was non-salient (e.g., no physical change in state). In the learning trials, a speaker looked at the non-salient object, and gave it a novel label. In the following test trials, children were asked to find either the previously labeled non-salient object (i.e., referent test), or the perceptually salient object with a new label (i.e., mutual exclusivity test). The mutual exclusivity test trials were added as a strong test of word learning as correct target identification in this test requires children to be able to use the speaker’s gaze to form the word-object mapping first and then use this knowledge to constrain the second novel label. Past research has shown that bilingual children were more likely to rely on a speaker’s eye gaze against perceptual cues (i.e., physical similarity between objects) than monolinguals in a word generalization task (Borjde, Ahmed, & Colunga, 2012). We hypothesized that bilingual children would demonstrate better referential word learning performance than monolingual children in our novel word learning task.

2. Method
2.1. Participants

A total of 76 children (34 males; \( M_{\text{age}} = 54.9 \) months, range = 52–55 months) participated in the study as a part of the Growing Up in Singapore Towards Healthy Outcomes (GUSTO) birth cohort. Twenty-eight children were English-speaking monolinguals (14 males; \( M_{\text{age}} = 54.8 \) months, range = 52–55 months). The remaining 48 were English-Mandarin bilinguals (20 males; \( M_{\text{age}} = 55.0 \) months, range = 53–55 months).

Children’s primary caregivers were asked to list the language(s) and the amount of time (percentage) the child was exposed to each language. Children were classified as monolingual if they were exposed to English at least 90% of the time, and bilingual if they were exposed to a second language (i.e., Mandarin) regularly besides English at least 25% of the time. The mean percentage of exposure to English for monolingual children was 92.3% (\( SD = 4.0 \)), while the mean percentage of exposure to English and Mandarin for the bilingual children were 54.5% (\( SD = 15.3 \)) and 45.5% (\( SD = 15.3 \)), respectively.

All children were of Asian descent and were Singaporean residents born in Singapore. To verify that the monolingual and bilingual children were comparable in demographic characteristics, information on maternal education and each family’s monthly household income was collected. Chi-square tests showed that these two groups of children did not differ significantly from each other in terms of the number of children whose mothers completed at least high school (monolingual: 78.6%; bilingual: 62.5%), \( \chi^2(1) = 2.11, p = .15 \), Cramer’s \( V = .17 \). Average monthly household income, on a scale of 1 (less than $1000) to 5 (more than $6000), was also not significantly different between the monolingual (\( M = 3.2, SD = 0.8 \)) and bilingual (\( M = 3.9, SD = 1.0 \)) groups, \( Z = 1.00, p = .32 \).
2.2. Stimuli and design

The experiment was modeled after Gliga et al.’s (2012) eye-tracking paradigm. All stimuli were presented in English.

Eight video clips were prepared; each consisted of a learning phase and a test phase. At the beginning of each video clip, a female speaker appeared seated behind a table with two novel objects placed on it. One of the two objects had changing colors or moving parts (salient distractor) and the other one was static (non-salient target). During the learning phase, the speaker first greeted the child with a cheerful voice, “Hello!” then turned her gaze toward the non-salient object and labeled it using a novel word either two times (also known as Short learning trial, e.g., “Look at this! It’s a neem. Do you want to play with the neem?”), or three times (also known as Long learning trial, e.g., “Look at this! It’s a blicket. Wow! Look at the blicket. Do you want to play with the blicket?”) (see Figure 1). These two types of learning clips were designed to explore whether children’s word-learning performance would be affected by the amount of exposure to the novel word. There were two fixed orders of trial presentation such that half of the children received trials in the order of SSLLSSLL and the other half received trials in the order of LLSSLLSS (“S” for short learning trial, and “L” for long learning trial). The placement of the salient and non-salient objects on each side of the screen was counterbalanced across clips. At the end of each learning clip, a simple animation was shown in the center of the screen and was followed by the test clip.

For the test phase, each test clip was comprised of two test trials, a referent test trial and a mutual exclusivity (ME) test trial. On each test trial, children saw a still image of the two objects on either side of the screen—their positions reversed compared to the previous learning trial—and heard the same speaker requesting one of the two objects (see Figure 1). On the referent test trials, the speaker’s voice asked children to look at and show the previously named non-salient object, e.g., “Look at the blicket! Look at the blicket! Can you show me the blicket? Can you show me the blicket?” On the ME test trials, the speaker used a different novel label asking children to look at and show where an object is (i.e., the unnamed salient distractor), e.g., “Look at the plume! Look at the plume! Can you show me the plume? Can you show me the plume?” The image of the two objects remained on the screen until children pointed to one of the two objects, and then a new trial began. The order of the two types of test trials was counterbalanced across the eight video clips; in four video clips, the referent test trial was presented first and was followed by the ME test trial, while in the other four video clips, the ME test trial was shown first, followed by the referent test trial.
Additionally, two familiar test trials were prepared, in which children saw a pair of common objects familiar to young children (a rubber duck and a child’s shoe). Similar to the novel word test trials, children were asked to look at and show first the duck and then the shoe, or vice versa, in two separate trials. The order of the two familiar test trials was counterbalanced across participants. These familiar trials were used to familiarize children with the task and, at the same time, to test children’s knowledge of familiar word-object mappings.

2.3. Procedure

The experiment was administered as a part of the GUSTO 54-month neurodevelopment visit. Children were seated on a chair in front of a 17-inch monitor, which was mounted with a Tobii T60 eye-tracker sampling at 60 Hz. An experimenter sat next to the child and conducted the experiment. The eye-tracker was first calibrated for each child using a 5-point calibration. The calibration was run until at least 4 points were properly calibrated for each eye. The experimenter introduced the task to the child in the language most often used by the child, as reported by the child’s caregiver and confirmed with the child at the beginning of the visit (English = 58, Chinese = 18). All children were told that a “teacher” was going to teach them some new words in English that they might not have heard before.

Next, children were presented with a familiar test trial, in which they saw a pair of familiar objects (i.e., duck vs. shoe) and heard the voice of a speaker asking them to find a target object (e.g., “Look at the duck! Look at the duck! Can you show me the duck? Can you show me the duck?”). Children were then verbally prompted by the experimenter to point to one of the two objects and were reminded that he/she should point only when asked. Children were then shown two video clips that each consisted of a learning phase and a test phase and were verbally prompted to point to one of the objects at each test trial. A second familiar test trial was then shown with the same pair of familiar objects, but this time the speaker’s voice requested the other target object. Finally, preliminary analysis showed no significant differences in task performance between children who received instruction in different languages.
children saw the remaining six video clips. Thus, the entire experiment consisted of two familiar test trials and eight video clips that each included a referent test trial and a ME test trial in the test phase. Children were randomly assigned to either one of the two fixed orders of trial presentation as mentioned in Section 2.2. Children’s pointing responses to the test trials were recorded by the experimenter and their eye movements throughout the experiment were recorded by the eye-tracker. The whole session was videotaped and children’s pointing responses were independently coded from the videos by a second coder.

2.4. Data analysis

Children’s eye gaze data during both learning and test trials were extracted using Tobii Studio software and were processed and analyzed using eyetrackingR packages (Dink & Ferguson, 2015). Rectangular areas of interest (AOI) were defined around the speaker’s face (learning trials) and each object (learning and test trials) with 0.6 degree of visual angle margin since the objects differ in size (see Holmqvist et al., 2015; Orquin, Ashby, & Clarke, 2016). All data were normalized for AOI sizes. The window-of-analysis for each trial began 1s after the rotation of the speaker’s head for learning trials and the onset of the speaker’s question for test trials and was 4s and 3s long for learning and test trials, respectively. This would allow children enough time to process and make looking responses. We first calculated the proportional looking time toward each AOI with respect to the total amount of time spent looking at all AOIs (i.e., speaker’s face and two objects on learning trials, and the two objects on test trials) for each trial. These proportions were then averaged across trials of the same type (learning, referent test, and ME test) for each child. To ensure that the averaged proportions were representative, individual trials were excluded from analysis if eye gaze data were missing for more than half of the duration of the window-of-analysis. Two participants did not accumulate enough looking time on more than half of the test trials, thus were excluded from the eye gaze data analysis (both from the bilingual group). Another three participants’ eye gaze data were not usable due to technical errors (one monolingual and two bilingual). The remaining 71 participants (27 monolingual and 44 bilingual) contributed data to this looking measure.

For children’s pointing responses, the experimenter and the independent coder agreed on most of the test trials (1362/1368, 99.6%). A third researcher resolved the conflicts between the experimenter and the coder on 6 trials (0.4%). Trials with correct pointing to the requested target object—the non-salient object in the referent test and the salient object in the ME test—were given a score of 1, while trials with incorrect or no pointing were given a score of 0. The percentage of correct pointing was calculated for the referent and ME test trials separately per child, as a measure of word-learning.
3. Results

We first analyzed children’s looking and pointing performance on the familiar test trials. Children in both monolingual and bilingual groups fixated at the correct target object more than 50% of the time, monolingual: $M = 67.49\%$, $SD = 18.95\%$, $t(26) = 4.80$, bilingual: $M = 71.27\%$, $SD = 21.70\%$, $t(42) = 6.43$, both $ps < .001$, $d > 0.92$. There was no significant difference between the two groups in the proportional looking time toward the correct referent, $t(68) = 0.75$, $p = 0.45$, $d = 0.18$. All but three children pointed to the target object correctly on both familiar test trials. Two monolingual children failed to point in the first familiar test trial but they pointed and chose the correct referent in the second familiar test trial. Another bilingual child chose duck when the speaker’s voice requested for either duck or shoe. Fisher’s exact test showed that the monolingual and bilingual groups did not differ significantly in terms of the percentage of children who were successful on both two familiar test trials (26/28 vs 47/48), $p = .55$. Removing data from the three children who failed in one of the familiar trials did not significantly change the results of subsequent analyses, thus they were included in the final sample for further analysis.

Preliminary analyses on children’s performance during the learning and test trials revealed no significant main effects of children’s gender, the type of learning (short-length vs. long-length), or their interaction with language group, so data were combined across gender and learning type in subsequent analyses.

3.1. Looking during learning

During the learning phase, children in both language groups spent most of the time looking at the speaker’s face, followed by the non-salient referred object, and then the salient but non-referred distractor (see Figure 2, also see Appendix B). Interestingly, while monolingual children tended to spend more time fixating on the speaker’s face than their bilingual peers, $t(69) = 1.94$, $p = .056$, $d = 0.47$, bilingual children looked longer than the monolinguals toward the referred object, $t(69) = 1.72$, $p = .090$, $d = 0.41$, although both differences were only marginally significant. The two groups did not differ significantly from each other in the proportional looking to the distractor, $t(69) = 0.84$, $p = .41$, $d = 0.20$. In addition, analyzing children’s looking distribution across the two objects (excluding the looking toward the speaker’s face) revealed that both groups of children looked at the referred object significantly more than the distractor, monolingual: $M = 0.68$, $SD = 0.14$, $t(26) = 6.72$, bilingual: $M = 0.69$, $SD = 0.15$, $t(43) = 8.32$, both $ps < .001$, $d > 1.27$, and there was no effect of language group on the proportional looking toward the referred object versus the distractor, $t(69) = 0.16$, $p = .87$, $d = 0.04$. These results suggest that the monolingual and bilingual children had comparable ability in following a speaker’s gaze to the referred object in the presence of a physically salient distractor, but monolingual children seemed to engage in looking at the speaker’s face more than the bilinguals during this learning phase.
3.2. Looking and pointing during testing

At test, children in both language groups fixated on the target object more than expected by chance for both referent and ME test trials, monolingual on referent test: $t(26) = 3.50, p = .002, d = 0.67$, monolingual on ME test: $t(26) = 4.27, p < .001, d = 0.82$, bilingual on referent test: $t(43) = 4.39, p < .001, d = 0.66$, bilingual on ME test: $t(43) = 7.50, p < .001, d = 1.13$ (see Figure 3), demonstrating success in using the speaker’s gaze to map the label to the referred object and to infer a new label for the unlabeled object using mutual exclusivity constraint. A 2 (test type: referent vs. ME) x 2 (language group: monolingual vs. bilingual) repeated measures ANOVA was conducted on children’s proportional looking toward the target object during the test. Results revealed no significant effects of test type, language group, or interaction.

Children’s pointing behavior during the test trials was analyzed in a similar way. Both groups of children pointed to the target at above-chance levels, monolingual on referent test: $t(27) = 6.84, p < .001, d = 1.25$, monolingual on ME test: $t(27) = 5.79, p < .001, d = 1.09$, bilingual on referent test: $t(47) = 15.19, p < .001, d = 2.19$, bilingual on ME test: $t(47) = 13.26, p < .001, d = 1.94$ (see Figure 4). The 2 (test type: referent vs. ME) x 2 (language group: monolingual vs. bilingual) ANOVA on the mean percentage of correct pointing showed a significant main effect of language group, $F(1,74) = 6.51, p = .013, \text{partial } \eta^2 = 0.81$. The main effect of test type and interaction of test type and language group were not significant, $Fs < 0.23, ps > .63$. Bilinguals were more accurate than monolinguals in identifying the target object as indicated by explicit pointing, and this advantage of bilingualism was evident in both referent and ME tests.
Figure 3. Proportional looking to the correct object on referent and mutual exclusivity (ME) tests by language group.

*Note.* Error bars represent the standard error of the mean. A proportion of 0.50 indicates chance performance.

![Graph showing proportional looking to target object](image)

Figure 4. Percentage of pointing to the correct object on referent and mutual exclusivity (ME) tests by language group.

*Note.* Error bars represent the standard error of the mean. A percentage of 0.50 indicates chance performance. Asterisks indicate significance level of the difference between two language groups.

*** $p < .001$
4. Discussion

This current study investigated the extent to which differences in language exposure influence how children use social cues against perceptual cues to form word-object mappings. By 4.5-year-olds, both monolingual and bilingual children were able to follow a speaker’s gaze toward a non-salient object in the presence of a salient distractor when learning a new word. While both groups of children spent a large proportion of the time during the learning phase looking at the speaker’s face (54% and 48% for monolingual and bilingual group, respectively), bilingual children tended to spend more time looking at the referred object and less at the speaker’s face compared to the monolingual children. In the subsequent test trials, both groups of children successfully mapped the learned word to the referred non-salient object and used this learned mapping to infer a new label for the unlabeled salient object. Although the two groups were comparable in their ability to identify the target object as indicated by looking times, bilingual children outperformed their monolingual peers when asked to pick the target object by overt pointing.

Our results provided novel findings that bilingual children were better able than monolingual children to use social cues when learning new words. This bilingual advantage in utilizing social cues may be a result of bilingual children’s heightened sensitivity to other people’s communicative intent in the process of language learning. In daily interactions, bilingual children are faced with a more demanding communicative environment compared with their monolingual peers; they need to adapt to different linguistic perspectives unique to individual languages and figure out what language a speaker is using. Bilinguals’ regular interactions with people who speak different languages also demand a greater need to monitor the speaker for cues in various language contexts. Indeed, past research has shown that bilingual children are better than monolingual children in understanding the referential intent of others’ through nonverbal cues (Yow & Markman, 2016) as well as in perspective-taking (Fan, Liberman, Keysar, & Kinzler, 2015). Our findings suggest that the bilingual advantage in the utilization of social cues in understanding referential intent extends to the use of these cues for successful word-learning.

Previous research suggests that while children as young as 12 months old are able to follow a speaker’s gaze to a labeled target object, when faced with a competing perceptual salient object, they either map the given label to the more salient object or show no mapping at all (Hollich et al., 2000; Yurovsky & Frank, 2015). It is only later in life (about 2-years-old) that social cues begin to dominate the word learning process (Golinkoff & Hirsh-Pasek, 2006). In our study, when the speaker looked at the non-salient object instead of the salient object while providing a novel label, 4.5-year-olds were able to follow the gaze to the referred non-salient object over the perceptually salient one (which was assumed to be more attractive to children). It is possible that the ability of older children to use social cues over perceptual cues in word learning develops as
children begin to recognize people as intentional and goal-directed agents (e.g., Beier & Spelke, 2012; Gergely & Csibra, 2003).

It should be noted that while strong language group effects at test were found based on explicit pointing responses, no significant differences between the two groups were found when using implicit looking measures. Some researchers also have found diverging results from explicit and implicit measures and argued that explicit ability may reflect stronger representations in knowledge compared to implicit ability (e.g., Berman, Chambers, & Graham, 2016; Hood, Cole-Davies, & Dias, 2003; Shinskey & Munakata, 2005). It is possible that the bilingual children in our study held a stronger representation of the learned mapping than the monolinguals due to their greater sensitivity to the speaker’s referential intent. Another possibility is that bilingual children in our study had more advanced general cognitive capabilities compared to the monolingual children. Bilingual children have been found to outperform monolingual children in tasks examining executive functions particularly inhibitory control (see a recent review in Bialystok, 2016). Children’s looking patterns during the learning trials suggested that bilinguals were better able than monolinguals to refrain from being overly attentive to the speaker’s face and focus on the referred object during the word mapping process. This could in turn lead to stronger encoding of the word-object mapping that translates into stronger representation of the mapping and results in better performance of the bilingual children than monolingual children in explicit pointing measures. Thus, it remains plausible that both bilingual children’s unique cognitive processes and their sensitivity to communicative intent contribute the bilingual advantage observed in this study.

In summary, the present work provides a first test of how monolingual and bilingual children differ in their monitoring and use of social cues in the word-learning process. Our results suggest that 4.5-year-olds use social information over perceptual salience to fast map word-referent pairings. Most importantly, children’s exposure to more than one language has positive effects on their ability to learn new labels using social cues and then constrain subsequent word learning using heuristic cues. Diversity in language experiences may contribute to the development of children’s ability to understand the underlying communicative intent of referential cues and thus enhance children’s use of social cues in the word-learning process.

Appendix A

The GUSTO study team includes Allan Sheppard, Amutha Chinnadurai, Anne Eng Neo Goh, Anne Rifkin-Graboi, Anqi Qiu, Arijit Biswas, Bee Wah Lee, Birit F.P. Broekman, Boon Long Quah, Borys Shuter, Carolina Uni Lam, Chai Kiat Chng, Cheryl Ngo, Choon Looi Bong, Christiani Jeyakumar Henry, Claudia Chi, Cornelia Yin Ing Chee, Yam Thiam Daniel Goh, Doris Fok, E Shyong Tai, Elaine Tham, Elaine Quah Phaik Ling, Evelyn Xiu Ling Loo, Fabian Yap, Falk Mueller-Riemenschneider, George Seow Heong Yeo, Helen
Appendix B

Children’s looking behavior over the course of learning trials

Note. Each curve represents the mean proportionate looking of the respective AOI for one language group.

References


