Received: 9 May 2022 | Revised: 30 January 2023 | Accepted: 4 February 2023

DOI: 10.1111/desc.13377

RESEARCH ARTICLE

Developmental Science WILEY

Robustness of the cognitive gains in 7-month-old bilingual infants: A close multi-center replication of Kovács and Mehler (2009)

Sybren Spit[1] | Andreea Geambașu[1] | Daan van Renswoude[1] | Elma Blom[2,3] | Paula Fikkert[4] | Sabine Hunnius[4] | Caroline Junge[2] | Josje Verhagen[5] | Ingmar Visser[5] | Frank Wijnen[2] | Clara C Levelt[1]

[1]Leiden University, Leiden, The Netherlands

[2]Utrecht University, Utrecht, The Netherlands

[3]AcqVA Aurora, UiT The Arctic University of Norway, Tromsø, Norway

[4]Radboud University, Nijmegen, The Netherlands

[5]University of Amsterdam, Amsterdam, The Netherlands

Correspondence
Sybren Spit, Leiden University, Leiden, The Netherlands.
Email: S.B.Spit@uva.nl

Sybren Spit and Andreea Geambașu share first authorship.

Funding information
Dutch Research Council, Grant/Award Number: 401.18.044

Abstract

We present an exact replication of Experiment 2 from Kovács and Mehler's 2009 study, which showed that 7-month-old infants who are raised bilingually exhibit a cognitive advantage. In the experiment, a sound cue, following an AAB or ABB pattern, predicted the appearance of a visual stimulus on the screen. The stimulus appeared on one side of the screen for nine trials and then switched to the other side. In the original experiment, both mono- and bilingual infants anticipated where the visual stimulus would appear during pre-switch trials. However, during post-switch trials, only bilingual children anticipated that the stimulus would appear on the other side of the screen. The authors took this as evidence of a cognitive advantage. Using the exact same materials in combination with novel analysis techniques (Bayesian analyses, mixed effects modeling and cluster based permutation analyses), we assessed the robustness of these findings in four babylabs ($N = 98$). Our results did not replicate the original findings: although anticipatory looks increased slightly during post-switch trials for both groups, bilingual infants were not better switchers than monolingual infants. After the original experiment, we presented additional trials to examine whether infants associated sound patterns with cued locations, for which we did not find any evidence either. The results highlight the importance of multicenter replications and more fine-grained statistical analyses to better understand child development.

KEYWORDS

bilingualism, cognitive advantage, eye tracking, language acquisition, replication

Highlights

- We carried out an exact replication across four baby labs of the high-impact study by Kovács and Mehler (2009).

- We did not replicate the findings of the original study, calling into question the robustness of the claim that bilingual infants have enhanced cognitive abilities.
- After the original experiment, we presented additional trials to examine whether infants correctly associated sound patterns with cued locations, for which we did not find any evidence.
- The use of novel analysis techniques (Bayesian analyses, mixed effects modeling and cluster based permutation analyses) allowed us to draw better-informed conclusions.

1 | INTRODUCTION

A question that has received much attention in language acquisition, both theoretically and practically, is whether growing up with more than one language comes with drawbacks, with benefits, or both. It has been shown that while mild disadvantages may occur in some domains (e.g., smaller vocabularies in each language, slower language production in each language; Bialystok et al., 2004; Byers-Heinlein & Lew-Williams, 2013; Werker & Byers-Heinlein, 2008), bilinguals may experience a cognitive boost due to the experience they gain in managing and monitoring more than one language (Bialystok, 2009). In other words, the continued practice of flexibly using several languages may provide bilinguals with domain-general benefits in attentional control. These benefits do not require a lifetime of experience with multiple languages, but seem to emerge already within the first year of a child's life. While previous studies had shown such benefits with adults who had been bilinguals from an early age (Bialystok, 2011; Bialystok et al., 2009), Kovács and Mehler's seminal (2009) study showed that infants of 7 months old who had been exposed to two languages from birth outperformed monolingual infants in attentional control, as tested in a series of visual switch tasks.

In each of the tasks presented by Kovács and Mehler (2009), 20 monolingual and 20 bilingual infants looked at a display showing two boxes: one on the left, and one on the right. The task consisted of 18 trials, in which a visual stimulus appeared in one of these two boxes, and during which infants' eye-movements were measured using an eye-tracker. During nine pre-switch trials, this stimulus was consistently presented on one side of the screen. In nine post-switch trials, the stimulus was shown on the other side. In each of the three experiments that Kovács and Mehler conducted, the visual stimulus was preceded by a particular cue that could help children anticipate where the visual stimulus would occur. In the first experiment, these cues were 18 different auditorily-presented trisyllabic non-words. In the second experiment, the cues were again auditorily-presented trisyllabic non-words, but now all pre-switch cues had a particular structure (e.g., AAB), which was different from the structure in the post-switch trials (e.g., ABB). In the third experiment, the cue was visual: infants would see different triplets of shapes (e.g., circle-circle-triangle) for each of the 18 trials.

Results from all three experiments showed that both mono- and bilingual infants anticipated where the visual stimulus would appear during pre-switch trials and looked faster towards the correct side of the screen as the trials progressed. Yet, during post-switch trials, only bilingual children learned to anticipate that the stimulus would now appear on the other side of the screen, indicating they could better suppress the behavior learned in the first phase of the experiment. These findings provide one of the major pieces of evidence for benefits in general cognition of infants being raised bilingually (see also Comishen et al., 2019). However, several scholars have been unable to replicate the findings from Kovács and Mehler in partial or conceptual replications (e.g., D'Souza et al., 2020; Kalashnikova et al., 2021; Dal Ben et al., 2022). While all three experiments presented in Kovcs and Mehler (2009) yielded bilingual advantages, we focused on replicating Experiment 2 as closely as possible. This experiment used a linguistic pattern as a cue for visual switching, "because 7-month-old infants and even newborns discriminate these regularities" (Kovács & Mehler, 2009, p. 6557). Although the original evidence that this claim was based on (Marcus et al., 1999) has been debated (Geambașu et al., 2022), we might still argue that this experiment provides the most direct test of the idea that bilingualism leads to a cognitive advantage. The presence of structured linguistic input in the experiment could be helpful in anticipating where the visual stimulus will occur, and thus guide switching behavior. As such, it might be a miniature model of the linguistic environment in which bilinguals grow up, where they have to distinguish between input in different languages and switch between those languages. Below we first provide more background on attentional control in bilinguals and discuss our approach in replicating this experiment.

2 | ATTENTIONAL CONTROL IN BILINGUALS

Bilinguals' enhanced attentional control is thought to develop as a result of dual language management and monitoring the appropriate language for each communicative interaction, where one of the two active languages needs to be inhibited when switching to the other (Bialystok, 2001; 2015; Costa et al., 2009; see Antoniou, 2019, for an extended review). Bilinguals need to recognize the different contexts in which languages are spoken, and must inhibit the language not appropriate in a particular communicative setting. Adapting to such demands is thought to lead to an increase or more efficient use of neural reserves that translates into better attentional control (Abutalebi & Green, 2016; Green & Abutalebi, 2013). This idea is backed up by numerous experimental studies that show that bilingual adults have better working memory, and perform better than monolinguals on inhibition or switching tasks (e.g., Bialystok, 2009; Hilchey & Klein, 2011; Kroll & Bialystok, 2013; Valian, 2015, but see e.g., Paap & Greenberg, 2013 for a critical evaluation of studies showing better performance by bilinguals).

The cognitive processes that are necessary to control and plan behavior are often investigated in older children and adults with tasks like the Simon task (Simon, 1969) or the Flanker test (Eriksen & Eriksen, 1974). In the Simon task, participants see two types of visual stimuli on a screen. For one stimulus, participants need to give a leftward response, and for the other a rightward response. Importantly, stimuli are presented on the left and right side of the screen, but this is unrelated to the response participants are required to give. Typically, reaction times increase when presentation side and response side are incongruent, but bilingual participants reportedly have less difficulty resolving this incongruence (e.g., Bialystok et al., 2005; Poarch & van Hell, 2012). In the Flanker task, participants see arrows pointing in various directions on each trial, either all in the same direction, or with the middle arrow pointing in a different direction than the other arrows. When the arrows are all pointing rightwards, participants must give a rightward response, whereas when the middle arrow is pointing leftwards, in the opposite direction from the rightward arrows, they have to give a leftward response. As in the Simon task, participants typically have more difficulty responding when the presented stimulus occurs in an incongruent situation, that is, when the middle arrow is flanked by arrows pointing in the opposite direction. Again, bilinguals have less difficulty performing such a task than monolinguals (e.g., Kapa & Colombo, 2013; Yang & Yang, 2016; Yoshida et al., 2011), which is often taken as an indication of enhanced attentional control.

Because bilingual children need to monitor and manage their languages from early on, attentional control is thought to develop more rapidly in bilingual than in monolingual infants (Kovács & Mehler, 2009). However, the results reported in the literature are inconsistent, and mainly stem from studies with adults or older children. Although several studies reported better performance for bilinguals on attentional control tasks compared to monolinguals (Antoniou, 2019), others argue the bilingual advantage is an overestimated phenomenon (e.g., Paap & Greenberg, 2013). The effects are often contingent on specific task conditions, age, and experience with the language (Blom et al., 2014; De Cat et al., 2018; Hilchey & Klein, 2011; Poarch, 2018; see Ware et al., 2020 for a meta-analysis), and are likely to be restricted to specific circumstances (Paap et al., 2015; van den Noort et al., 2019; Verhagen et al., 2020). Recent meta-analyses (de Bruin et al., 2015; Lehtonen et al., 2018) and studies using larger samples than typical in studies on multilingualism (Duñabeitia et al., 2014; Gathercole et al., 2014) did not find evidence for better performance in bilingual school-aged children and young adults. This also raises the question of the replicability of better performance on tasks by bilingually-raised infants, especially because they have very little or no experience using (i.e., producing and suppressing) language in their first year.

3 | CURRENT STUDY

The original study by Kovács and Mehler (2009) is often cited as an example of the benefits of bilingualism in early childhood (e.g., Antoniou, 2019; Siegler et al., 2014); during the post-switch phase, the bilingual infants were better at inhibiting what they had learned during the pre-switch phase than the monolingual control group. Several researchers carried out partial or conceptual replications of the original study (Dal Ben et al., 2022; D'Souza et al., 2020; Kalashnikova et al., 2021), but were unable to replicate the original findings. Kalashnikova et al. (2021) replicated the first and third experiment of Kovács and Mehler (2009), but constructed novel cue stimuli based on the original stimuli, and also presented children with more pre- and post-switch trials (12, as compared to 9 in the original study). Using this slightly altered set up, they found that both monolingual and bilingual children learned to anticipate where the reward would occur on the screen, but they did not find any group differences during the post-switch trials, as was observed and taken as evidence for a better performance by bilingual infants in the original study. The replication effort by D'Souza et al. (2020), which consisted of a replication of Experiment 3 with "arguably more engaging" stimuli, also did not show better performance by bilingual children.

Furthermore, Dal Ben et al. (2022) presented a reanalysis of the data from the D'Souza et al. (2020) and Kalashnikova et al. (2021) studies, supplemented by a conceptual replication of the third experiment from Kovács and Mehler (2009). This study also did not yield the same results as in the original paper. Specifically, the authors found evidence that bilinguals were better at anticipating where the reward would occur during post-switch trials than monolinguals, as was found in the original study, but they also observed that monolinguals were often better than bilinguals at such anticipation during the pre-switch trials. These findings thus do not provide clear evidence that bilingual infants are better switchers, since they were not better in the pre-switch phase in the first place. The authors rather argue that bilinguals might exploit different processing strategies than monolinguals, as they do not learn to anticipate during the pre-switch phase, but are still able to do so in the post-switch phase.

These previous replication efforts thus provide mixed results, and highlight the need for an exact replication using the same materials as the original study. Here, we try to fill this void and present an exact replication of the second experiment of Kovács and Mehler (2009), in which a verbal stimulus can be used to guide switching behavior. In order to be able to verify the robustness and generalizability of these results, a replication is necessary. Here we undertake a multi-center replication (see Frank et al., 2017; Geambașu et al., 2022; The ManyBabies Consortium, 2020 for the importance of such attempts), albeit in a different country, and with infants acquiring different languages than in the original study, which was carried out in Italy and included mostly Italian-Slovenian bilingual infants. Except for these changes, we remained as faithful as possible to the original study. Both a positive and a negative (null-result) outcome would be informative for our understanding of the impact of bilingualism on infants' general cognition. Using Bayesian statistical analyses in addition to the original analyses allowed us to more confidently interpret the results, irrespective of whether they would show significant differences between participant groups or not (Van de Schoot et al., 2014; Wagenmakers et al., 2018).

Kovács and Mehler's study suggests that a cognitive benefit of bilingualism can appear already after very limited bilingual experience. In order to investigate the robustness of the original study, we therefore conducted an exact replication of the second experiment from Kovács and Mehler (2009), in which switching was cued by an auditory stimulus following a particular pattern (either AAB or ABB in which A and B represent syllables). However, even though the auditory stimulus could cue switching in the original experiment, switching could also be predicted without this stimulus to some extent: the first nine trials always appeared on one side of the screen, and the last nine trials always appeared on the other side of the screen. Infants could have therefore simply realized that they should start looking at the other side after a certain number of trials, and did not necessarily need to associate the cue with a particular side of the screen. Kovács and Mehler (2009) observed that anticipatory looks increased more rapidly in the pre-switch phase of Experiment 2 compared to the pre-switch phase of Experiment 1, where the auditory cue was absent. While this suggests that the auditory cue was helpful in learning the contingency, it is not a direct test of the hypothesis that infants associated the structure of the cue with the appearance of the reward. Therefore, in order to test if infants learned the contingency between side of screen and a specific speech pattern (e.g., AAB predicts left side, ABB predicts right side), we introduced an additional phase after the original pre- and post-switch phases. In this new phase, the auditory stimuli (in either an AAB or ABB structure) were presented in random order, and the visual stimuli appeared on the left or right side of the screen according to the contingency that had been present in the first two phases of the experiment. If bilingual children are indeed better switchers and if they learned to associate the auditory cue with a particular side of the screen, they should show enhanced performance in these trials as well.

4 | METHODS

4.1 | Participants

For our replication, we tested 143 infants in four baby labs in the Netherlands, at the University of Amsterdam (n = 44), Leiden University (n = 24), Utrecht University (n = 54), and Radboud University in Nijmegen (n = 21). From this group, we included 98 infants (58 monolingual, 40 bilingual) in the final sample, of which a description is presented below. The reasons for excluding certain children from analysis, and the numbers of children excluded for each of these reasons, are presented below as well.

Data collection took place between February 2021 and December 2021. We aimed to test an equal number of infants for both experimental groups, evenly distributed across labs, but due to practical limitations affecting recruitment (e.g., COVID measures and rebuilding of lab rooms), we instead terminated data collection at the end date of this project, which led to an unequal distribution of infants per lab. The target population of the original study was 7-month-old infants. As in Kovács and Mehler (2009), we tested infants within ±2 weeks of this age (range = 7 months; 1 day–7 months; 30 days). Preterm infants who were born earlier than 37 weeks gestation were excluded. Infants who were reported by caregivers as having a visual impairment or chronic ear infections (more than three ear infections since birth) or who had an ear infection at the testing session were excluded.

To classify infants as either mono- or bilingual we focused on their language exposure, using an adaptation of the Language Exposure Questionnaire (LEQ, Bosch & Sebastián-Gallés, 2001; Cattani et al., 2014; the full English questionnaire can be consulted at https://osf.io/p4dwu). The LEQ questionnaire was also used in the original study (Kovács, person. comm.) and provides an estimate of children's bilingual exposure, based on the relative amount (in percentages) of exposure to each language, obtained through caregiver reports. We followed other replications of the Kovács and Mehler study (Kalashnikova et al., 2021; Dal Ben et al., 2022) to determine whether infants should be included as bilingual or monolingual participants: children whose caregivers provided at least 90% of their language exposure in one language were classified as monolinguals. Children whose caregivers provided at most 75% of exposure in the dominant language, and at least 25% in (an)other language(s) were classified as bilinguals.[1] To obtain a picture of the general background of our mono- and bilingual infants, we also gathered information about caregivers' education. This was assessed on a four-point scale with (1) 'primary school', (2) 'lower vocational training', (3) 'secondary school and/or vocational training', and (4) 'higher education (i.e., undergraduate or graduate degree)' as its scale points, and averaged for both caregivers.

Our monolingual group (n = 58, M = 232 days old, SD = 8 days, min-max = 214–245 days[2]) included 26 girls and 32 boys. Caregivers of these children had a mean education score of 3.86 (SD = 0.28 range = 3.0–4.0). Most children were from monolingual Dutch families (n = 55; 95%). The remaining children (n = 3; 5%) were from families where other languages were spoken (English, German, and Serbian).

Our bilingual group (n = 40, M = 227 days old, SD = 9 days, min-max = 208–246 days) included 20 girls and 20 boys. Caregivers of these children had a mean education score of 3.92 (SD = 0.22 range = 3.0–4.0). Most children were from families where Dutch (n = 26; 65%) or English (n = 8; 20%) was spoken most of the time, in addition to another language. The remaining children (n = 6; 15%) were from families where another language was spoken most of the time (Chinese, Icelandic, Mandarin, Spanish, Tamil, Turkish). Which languages were combined varied, with a total of 17 different languages other than Dutch.

We excluded 32 participants because of technical issues or fussiness; one child because they did not finish the first 18 trials, and a further 12 because they did not meet the language classification criteria.[3]

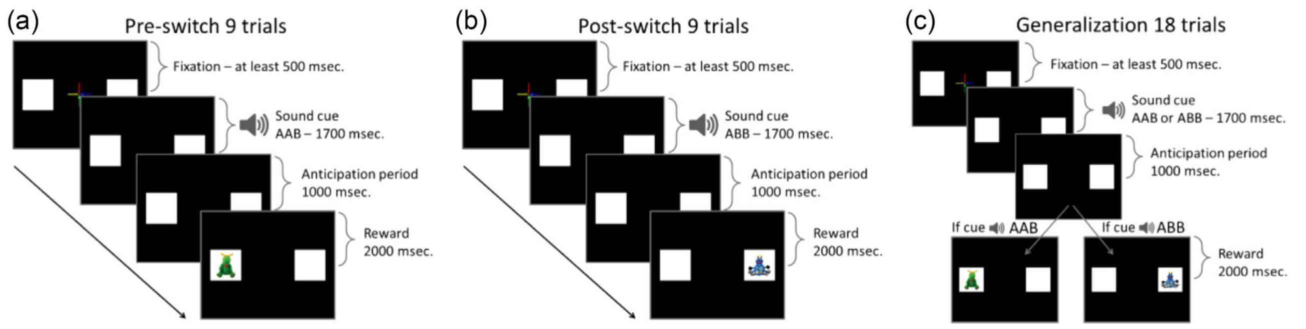

FIGURE 1 Schematic overview of the experiment.

4.2 | Experimental design

The switch task used the exact same stimuli (described below under Stimuli) and was set up exactly like Experiment 2 of the original Kovács and Mehler (2009) study. This experiment consisted of nine pre-switch and nine post-switch trials. In every trial, the infants' attention was first drawn towards the middle of the screen with an attention grabber stimulus. Once the eye-tracker registered attention towards the screen for at least 500 ms, the attention grabber disappeared and a trisyllabic sound stimulus was played, which lasted 1.7 s, while two white boxes appeared on the left and right of the screen. The sound stimuli all followed either an AAB (*le le mo*) or ABB (*le mo mo*) pattern and were predictive of the appearance of a toy puppet in one of two white boxes. After the 1.7-s time lapse, there was a 1-s anticipation period, after which the toy puppet appeared (see Figure 1A,B). In the first nine trials, the toy puppet consistently appeared on one side of the screen and the preceding sound followed the same pattern. In the following nine trials, the toy puppet was consistently shown on the other side of the screen, and all sound stimuli followed the alternative pattern. Presentation side and accompanying stimulus pattern were counterbalanced across participants.

We added 18 association trials (Figure 1C) after the original nine pre-switch and nine post-switch trials. In these trials, the contingency between sounds and the side of appearance of the toy puppet reward was exactly the same as during the previous 18 trials. However, the order of the 'AAB' and 'ABB' trials was completely random. These items were added to investigate whether switching anticipatory looks to the other side of the screen only resulted from the realization that in the post-switch trials the stimulus occurred on the other side, or whether this behavior was based on a learned association between a specific structure of the auditory stimulus and the presentation side of the visual stimulus.

4.3 | Stimuli

The auditory and visual stimuli were identical to those used by Kovács and Mehler (2009) and were provided by the first author, Agnès Kovács.[4] The attention grabber was a moving sequence of four colored arrows that pointed towards the center. A red arrow pointing down appeared first, followed by a blue arrow pointing left, a green arrow pointing up and a yellow arrow pointing right. After 1 s, these arrows formed a cross at the center of the screen, before disappearing in the same order during the next second. The individual arrows subtended approximately 3 degrees of visual angle and formed a cross 6 × 6 degrees. To draw the infant's attention, the visual attention grabber was accompanied by a 'beep' sound.

The trisyllabic cues that were played after infants directed their attention to the screen consisted of the syllables *le*, *ve*, *mo*, *zo*, *ri*, and *ni*. All syllables could occur in both the AAB and ABB sequences. Each syllable had a duration of 0.4 s, and between two syllables in the cue there was a pause of 0.25 s. As a result, all sound cues had a duration of 1.7 s.

The white boxes that displayed the puppet 1 s after the end of the cue sound were located on the left and right side of the screen and subtended approximately 7 × 7 degrees of visual angle. There were three different puppets that appeared in these boxes: a hippo, a bug, and a star. These puppets were looming for 2 s, with the start size being around 4 × 4 degrees of visual angle after which they became larger for 500 ms (6 × 6 degrees), smaller for 500 ms (4 × 4 degrees), larger for 500 ms (6 × 6 degrees), and smaller for 500 ms (4 × 4 degrees). The three puppets were always accompanied by specific sounds for the duration of 1 s: a 'ring' sound for the star and hippo, and a 'beep-beep' sound for the bug. These sounds were played twice during the reward phase.

4.4 | Procedures

Efforts were made to be as consistent as possible across labs with respect to general procedures, while at the same time allowing the labs to follow their own lab protocols as much as possible. This meant that the order of operations during the procedures was as similar as possible, while there were differences in stimulus presentation software, technical specifications of hardware, and experimentation booths. In all cases, caregivers were contacted and informed about the study prior to making an appointment to come to the lab. Before coming to the lab, they were asked to fill out a short questionnaire containing a series of questions on language background, medical history, and parental education of the caregivers (see our OSF page for the full questionnaire).

Infants and their caregivers were welcomed to the lab by the experimenter and were informed again in more detail about the goal and procedure of the study via a written document and orally (either in Dutch or in English) about the importance of remaining calm and holding their infant comfortably while not influencing or reacting to them unless their infant showed signs of distress. They were also informed in writing and orally about their rights as participants to stop the experiment and to withdraw. Before the start of the experiment, caregivers signed an informed consent form indicating that they consented to their infant's data being used for research purposes. If caregivers did not hand in the background questionnaire before their visit, they were asked to fill it out before starting the experiment.

After consent was obtained, caregivers and infants were led to the room where the experiment proceeded. Infants were seated in a baby chair in front of their caregivers or on their lap. The task was presented using Psychopy (Peirce et al., 2019). Eye tracking data was gathered using a Tobii T120 and a 24 in. screen in Leiden, a Tobii X300 and a 23 in. screen in Nijmegen, and an EyeLink 1000 (in remote mode) and a 17 in. screen in both Amsterdam and Utrecht. We calibrated the eye trackers before the start of the experiment using an infant-friendly calibration. Both EyeLink machines sampled at a rate of 500 Hz. The Tobii machine in Nijmegen sampled at a rate of 300 Hz, and the Tobii machine Leiden at a rate of 120 Hz. In all labs, infants were seated at approximately 60 cm from the screen. After the experiment, caregivers and infants were led out of the experimental area and back to the welcome area. They were then compensated for their participation with cash, reimbursed travel costs, and/or the gift of a children's book.

All procedures involving participants in this study have been approved by the Institutional Review Boards at the universities where data was collected. Each lab followed the ethical guidelines and ethics review board protocols of their own institution. Data from individual participants were stored locally at each lab. After anonymizing, data were uploaded to a protected and shared folder for central analysis. All data can be found on our OSF page (https://osf.io/p4dwu).

4.5 | Analyses

In the original experiment, results showed that the number of correct anticipations increased over the first nine trials in both mono- and bilingual infants. However, only in bilingual infants did the number of correct anticipations increase on trials 10–18 (i.e., after the targets switched sides), while the number of correct anticipations on trials 10–18 in monolingual infants did not increase. In our analyses, we investigated whether similar effects occurred in our replication study. All analyses were carried out in R (R Core Team, 2015) using the *brms* package (Bürkner, 2021), *eyetrackingR* package (Dink & Ferguson, 2015), and *rstatix* package (Kassambara, 2021) where needed. Scripts that were used for the analyses can be consulted on our OSF page as well.

4.5.1 | Original analysis

In our first analysis, we used the same methods as in the original study. We included these analyses to exactly replicate the original study. For these analyses the screen was divided into three equal parts: left, middle, and right. Every sequence of samples during which participants were looking at a particular area of interest was taken as a look. As in Kovács and Mehler (2009), we coded children's looks that occurred during the 1s time window starting 150 ms after the end of the word and ending 150 ms after the appearance of the reward. All trials with only looks to the center or non-target, or missing looks, received a value of 0. Any trial during which children looked towards the side where the visual stimulus would occur at some point within the second after the offset of the auditory stimulus received a value of 1. When infants looked both to the left and right side of the screen during the anticipatory period of the same trial, the side that received the longest uninterrupted look received a value of 1. We then calculated proportions of anticipatory looks within three blocks (trials 1–3, 4–6, 7–9) for individual participants. For example, when participants showed an anticipatory look for two of the first three trials, they would obtain a score of 0.67 for the first block. Note that in the additional analysis 1 explained below, we used a more stringent analysis where we only considered correct versus incorrect looks on the screen, and did not code missing looks as 0.

Next, following Kovács and Mehler (2009), we carried out two separate ANOVAs for the pre- and post-switch trials. For these tests we combined data from all four labs. In both ANOVAs, the proportion of anticipatory looks was taken as the dependent variable, linguistic group was taken as a between-participant factor, and block as an independent within-participant variable. If results from these analyses replicated the original findings, we should observe a main effect of block for the pre-switch trials, but no interaction effect between block and group: that is, both monolingual and bilingual children should show more anticipatory looks as the pre-switch phase progressed. For the post-switch trials, we should observe an interaction effect between group and block, but not necessarily a main effect of group. Specifically, if bilingual children are indeed better switchers, only this group should show more anticipatory looks as the post-switch trials progressed.

As noted before, we conducted these ANOVAs because they were carried out in the original study as well. The actual effect that the original paper targeted was a three-way interaction between block, group, and trial type: the anticipation effect (block) should be larger for bilinguals than for monolinguals (group) when comparing the post-switch trials to the pre-switch trials (trial type). This effect was investigated in the original study by carrying out two split analyses and comparing the *p*-values for these separate analysis, where the absence of an interaction between block and group in the pre-switch trials and the presence of an interaction between block and group in the post-switch trials was taken as evidence for a three-way interaction between block, group, and trial type, indicating that only bilingual infants were able to update their switching behavior in the second part of the experiment. Crucially, carrying out such split analyses in order to establish an interaction is erroneous (Bland & Altman, 2011; Makin & Orban de Xivry, 2019; Nieuwenhuis et al., 2011). To solve this problem, we therefore added a novel analysis to the separate ANOVAs that were carried out in the original paper, and directly tested the three way interaction between block, group, and trial type using more state-of-the-art Bayesian mixed effects models.

4.5.2 | Additional analysis 1: Logistic regression

To directly investigate the three-way interaction of interest in this study, we conducted a Bayesian general logistic mixed model, in which we could include the three-way interaction between group (mono- or bilingual), trial type (pre- and post-switch trials), and trial number (1–9). The advantage of using this model over an ANOVA is that we did not need to treat missing trials as non-anticipatory looks, as in an ANOVA, but could simply exclude them from the analyses. A further advantage of this technique is that it can be used to better investigate development over trials, since there is no need to aggregate trials over blocks. Finally, the added value of a Bayesian analysis is that a Bayes factor (BF) can be calculated that allows one to differentiate between evidence for the null hypothesis, evidence for the alternative hypothesis, and lack of evidence altogether. Specifically, the BF indicates the ratio of the likelihood of a given hypothesis compared to some other hypothesis (Beard et al., 2016). A BF_{10} of 10, for example, indicates that it is 10 times more likely that the data would be observed under the alternative hypothesis than under the null hypothesis. A BF_{01} of 3, on the contrary, indicates that it is 3 times more likely that the data would be observed under the null hypothesis than under the alternative hypothesis. A BF between 10 and 30 is generally seen as strong evidence, a BF of 3–10 as moderate evidence, whereas a BF between 1 and 3 merely provides anecdotal evidence for a particular hypothesis (Lee & Wagenmakers, 2013).

We used the R-package *brms* (Bürkner, 2021) to fit the following model, in which we included generic weakly-informative priors for all fixed effects (Gelman et al., 2008; Gelman, 2020):

$$\text{AnticipatoryLook} \sim \text{Trial\_Type} * \text{Trial\_Number} * \text{Linguistic\_Group}$$
$$+ (\text{Trial\_Number} + \text{Linguistic\_Group} | \text{Participant/Lab})$$
$$+ (\text{Trial\_Number} + \text{Linguistic\_Group} + \text{Trial\_Type} | \text{Sound\_Stimulus})$$

In this model, we took the anticipatory looks from the task (1 for an anticipatory look, 0 for an incorrect look) as the dependent variable, trial type (pre- vs. post-switch) and trial number (1 to 9) as within-participants fixed effects, linguistic group (monolingual vs. bilingual) as a between-participants fixed effect, participant and lab as between-participants random effects, and sound stimulus as a within-participants random effect. Orthogonal sum-to-zero contrast coding was applied to the binary fixed effects (i.e., trial type and linguistic group; Baguley, 2012, pp. 590–621; Schad et al., 2020). We aimed to keep the models as fully specified as possible (Barr et al., 2013) by including random intercepts for sound stimulus, participants, and lab, with participants nested in the different labs. Furthermore, we included random slopes for trial number and trial type per participant, as these both varied within participants. We also included random slopes for trial number, linguistic group, and trial type per sound stimulus, because these factors varied across the different sound stimuli that participants heard. The model, however, did not converge when we also included interaction effects between our random slopes, which were thus not included in the final model.

4.5.3 | Additional analysis 2: Cluster-based permutation analysis

Simplifying looking behavior during the test trials to either anticipatory (1) or non-anticipatory (0) looks, as in our previous analysis, does not do justice to the wealth of gaze behavior data that is obtained during these trials. In order to investigate looking behavior over time during the switch task, we carried out several cluster-based permutation analyses (see Dink & Ferguson, 2015; Maris & Oostenveld, 2007; Spit et al., 2022, for examples) using the *eyetrackingR* package (Dink & Ferguson, 2015). These analyses enabled us to investigate the development of the proportion of looks towards the target picture over trials during the course of the anticipation time window. Using this type of analysis, we could analyze the full set of trials from the moment children looked at the attention grabber until the offset of the visual reward on the screen (see Figure 6 below for a visual illustration of this time window). We analyzed all trials for which at least 50% of the looks were on the screen. Since we could use these analyses for a more fine-grained picture of the looking behavior children exhibit during the experiment, we took a more precise area of interest than in the previously described analyses, where the screen was divided into three equal parts. For each trial with at least 50% of available looks, we coded for all available samples whether they were directed at the white square in which the visual stimulus would appear (1) or not (0).

The cluster-based analysis involved two steps, and allowed us to determine when (within the marked time frame) monolingual versus bilingual participants looked more often towards the target square. This difference between groups could in principle occur in both directions (i.e., monolinguals or bilinguals looking more often towards the target square). In the first step of the analysis, the data were split into 50 ms time bins. For each time bin we calculated a *t*-statistic to determine whether participants from one group looked significantly more often towards the target square than participants from the other group. Adjacent time bins for which there were significant differences were clustered together and the *t*-values of the bins were summed. In a second step, the data were reshuffled 1000 times, to see how likely the observed clusters and accompanying summed statistics would be if the data were randomly generated. Using this procedure, we could correct for false alarms, while not losing much of the richness of the data. Unfortunately, a cluster-based permutation analysis allows for only one predictor, which was linguistic group in this case. Because time during the trials is already mapped out, it was not possible to look at further interactions, for example between linguistic group, trial type, and trial number simultaneously. Therefore, we ran separate analyses for each of the 18 trials. Because no single metric could be computed that expresses how group differences that occur during trials develop as the experiment progresses, these analyses do not allow us to make strong inferential claims about the development of anticipatory looking behavior in the two groups as the experiment progresses. However, these analyses do provide a more fine-grained description of children's looking behavior over the course of the experimental trials.

In addition to this set of analyses in which we mapped out looking behavior as the trials progressed, we also ran two exploratory analyses where we grouped all pre-switch and post-switch trials together. Performing these two analyses, we thus investigated whether there was a group difference in anticipatory looks for all trials during either the pre-switch or the post-switch phase. These analyses do not give an answer to the original research question, as they do not indicate whether one group adjusts their looking behavior as the experiment progresses, but they do provide more insight into general looking behavior during each of the two phases, and whether one group of infants showed different looking behavior in general during either of the two phases. The outcomes from these analyses can be consulted on our OSF page.

4.5.4 | Additional analysis 3: Association trials

As described above, we included 18 association trials in our replication. These trials were added after the original test trials, and were included to examine whether infants would associate the sound stimulus with the corresponding side of the screen. Since the sound stimulus and side of the screen were randomly distributed over these items in these trials and we were not interested in how anticipatory looking developed over the course of this part of the experiment, we grouped all items together. If there was a group difference in the ability to associate the auditory stimulus with the corresponding side of the screen, this should be indicated by either earlier looks towards the target square or more looks towards the target square for one group of infants than for the other group. As linguistic group was the only independent variable in this case, we carried out a cluster-based permutation analysis that compared the looking patterns of the monolingual and bilingual groups during the association trials. For this analysis we followed the same general procedure as for the previously described cluster-based analyses.

5 | RESULTS

5.1 | Original analysis

As described above, we first analyzed the data using the same types of ANOVAs as in the original study. Table 1 shows descriptive of the proportions of looking times per group for different blocks and trial types during the experiment as used in these analyses. Figure 2 shows these proportions of looks split out across the nine pre- and post-switch trials following the coding that was used for these ANOVAs. The first 2×3 ANOVA, which looked at pre-switch trials, and took proportion of anticipatory looks as a dependent variable, and linguistic group (monolingual, bilingual) and block (1, 2, 3) as independent variables, did not show a significant main effect of block on anticipatory looks during the pre-switch trials ($F(2, 192) = 0.625$, $p = .536$, $\eta_p{}^2 = 0.003$). This is not in line with the original finding that children got better at predicting where the reward occurred. We did not see a significant interaction between group and block ($F(2, 192) = 0.064$, $p = .938$, $\eta_p{}^2 < 0.001$), which would have been indicative of faster, or less fading, anticipation by one of the two groups. There was also no main effect of group ($F(1, 96) = 0.345$, $p = .558$, $\eta_p{}^2 = 0.002$), which would have indicated that one particular group showed more anticipatory looks during the pre-switch trials than the other group.

Our second 2×3 ANOVA, which looked at post-switch trials and again took proportion of anticipatory looks as a dependent variable, and group and block (now 4, 5, 6) as independent variables, showed a significant main effect of block on anticipatory looks during the post-switch trials ($F(2, 192) = 3.572$, $p = .030$, $\eta_p{}^2 = 0.017$). All infants together showed more anticipatory looks as the post-switch phase of the experiment progressed, which suggests that they became significantly better at predicting where the reward occurred. There was no main effect of group ($F(1, 96) = 1.269$, $p = .263$, $\eta_p{}^2 = 0.007$), which shows that, over all blocks, neither group outperformed the other group in their anticipatory looks. Furthermore, we did not observe a significant interaction between group and block ($F(2, 192) = 0.354$, $p = .702$, $\eta_p{}^2 = 0.002$). Thus, as in the pre-switch phase, we did not find any evidence that one group of children showed different anticipatory behavior compared to the other group. Note that in the original study, there was an interaction between group and block during this phase of the experiment: in that study, the bilingual children were better than monolingual children at anticipating where the reward occurred during the post-switch trials.

5.2 | Additional analysis 1: Logistic regression

Next, we carried out a Bayesian logistic regression with mixed effects, which investigated the effects of trial type, group, and trial number in a single analysis. Figure 3 shows these proportions of looks split out across the nine pre- and post-switch trials following the coding that was used for the Bayesian analysis. This analysis showed strong evidence for a main effect of trial number on anticipatory looks ($\beta = 0.13$, 95% CCI [0.06, 0.20], $BF_{10} > 100$), but no interaction between trial type (pre- vs. post-switch) and trial number ($\beta = 0.01$, 95% CCI [−0.09, 0.11], $BF_{01} = 19.14$). The BF for the main effect of trial number indicates that these results can be interpreted as evidence for the hypothesis that infants across groups learned to anticipate where the reward would occur as the trials proceeded. However, the BF for the interaction between trial type and trial number indicates that the effect of trial number is equal for pre-switch and post-switch trials. We found no evidence that either mono- or bilinguals were better at anticipating where the reward would occur in general ($\beta = -0.06$, 95% CCI [−0.66, 0.56], $BF_{01} = 3.25$). Furthermore, the lack of an interaction between group and trial number ($\beta = 0.04$, 95% CCI [−0.08, 0.16],

TABLE 1 Proportion of anticipatory looks during the six blocks of the experiment for the monolingual and bilingual infants.

		Monolingual (*n* = 58)			Bilingual (*n* = 40)		
		M	*SD*	min-max	*M*	*SD*	min-max
Pre-switch phase	*Block 1*	0.49	0.33	0–1	0.54	0.34	0–1
	Block 2	0.53	0.30	0–1	0.55	0.36	0–1
	Block 3	0.48	0.35	0–1	0.51	0.38	0–1
Post-switch phase	*Block 4*	0.35	0.30	0–1	0.31	0.31	0–1
	Block 5	0.40	0.32	0–1	0.37	0.36	0–1
	Block 6	0.48	0.34	0–1	0.39	0.32	0–1

Note: Each block consisted of three trials. If participants anticipated where the reward would occur for two out of those three trials, they would obtain a score of 0.67.

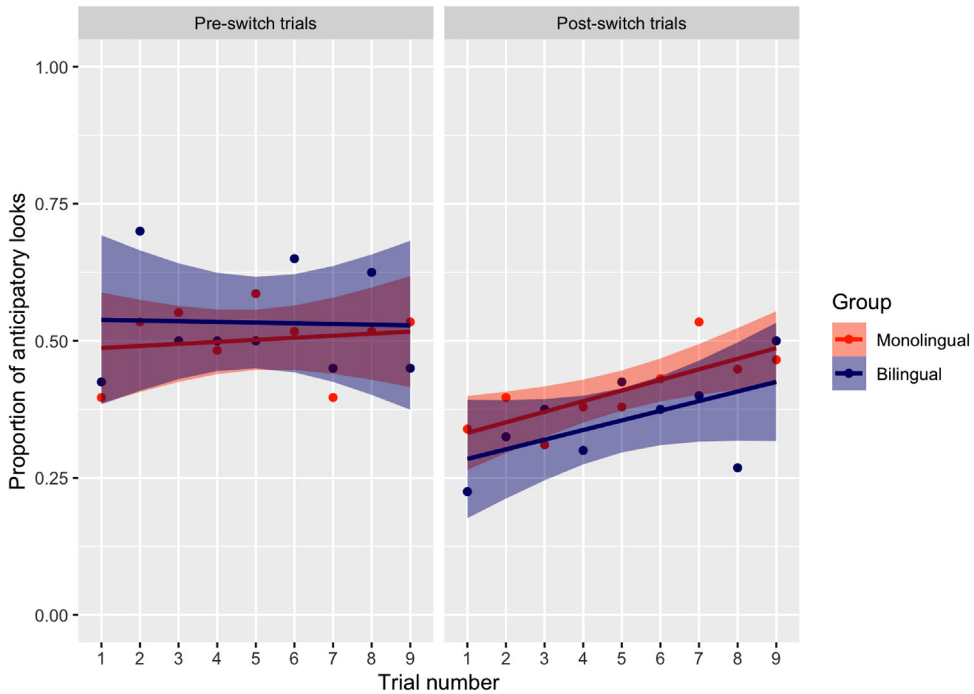

FIGURE 2 Plots of the proportion of anticipatory looks during the pre- and post-switch trials across trials as coded for the original ANOVAs. ANOVA: analysis of variance.

BF_{01} = 13.11) indicates no group difference in learning to anticipate where the reward would occur as the experiment progressed.

Furthermore, we found evidence for a main effect of phase: all infants together showed less anticipatory looks for post-switch trials, than for the pre-switch trials (β = –0.80, 95% CCI [–1.0, –0.12], BF_{10} = 4.17), which suggests that infants were actually not very good at switching at all, as they did not adjust their looking behavior during the post-switch trials to the level that was reached during the pre-switch trials. The absence of interactions between trial type and group (β = 0.06, 95% CCI [–1.09, 1.22], BF_{01} = 1.62), and between trial type, group, and trial number (β = –0.16, 95% CCI [–0.34, 0.04], BF_{01} = 2.97), also entails that we have no evidence that one of the two groups of children was better able to switch during the experiment.

5.3 | Additional analysis 2: Cluster based permutation analyses

To get a more fine-grained picture of when exactly the two groups showed anticipatory eye movements, we conducted a series of cluster-based permutation analyses. These analyses can demonstrate how looking behavior changes over time across trials. Specifically, these analyses allowed us to determine the intervals of a trial during which children from one group looked more often towards the target side of the screen than children from the other group. Results are shown in Figure 4 and Figure 5. All intervals during which one group of infants looked more towards the target side of the screen than the other group can be consulted in the supplementary materials on our OSF page. Although for nearly all trials there were small intervals during which

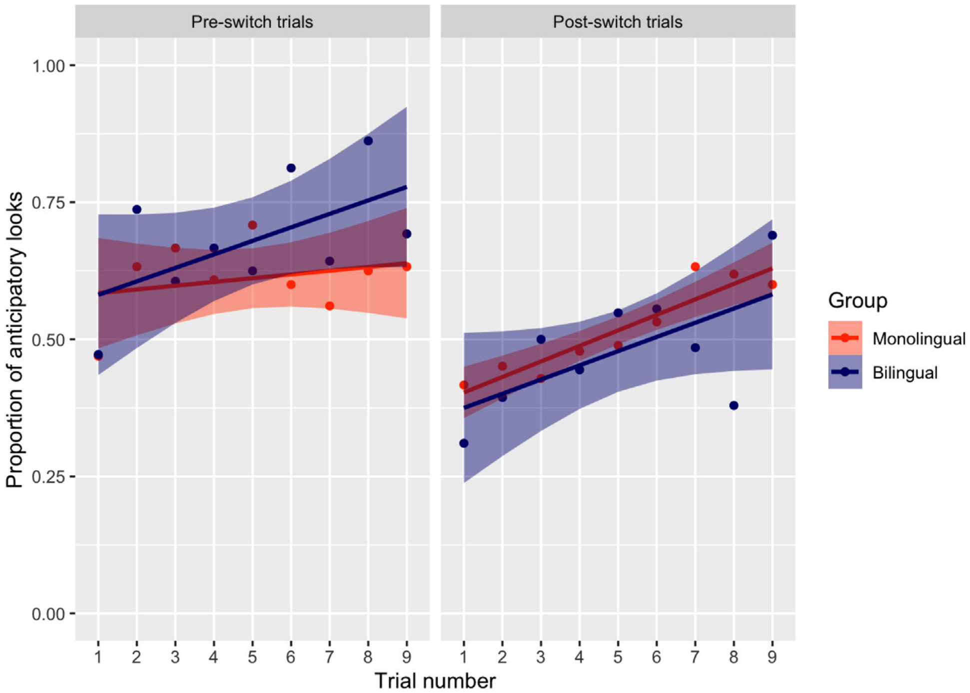

FIGURE 3 Plots of the proportion of anticipatory looks during the pre- and post-switch trials across trials as coded for the additional Bayesian analysis.

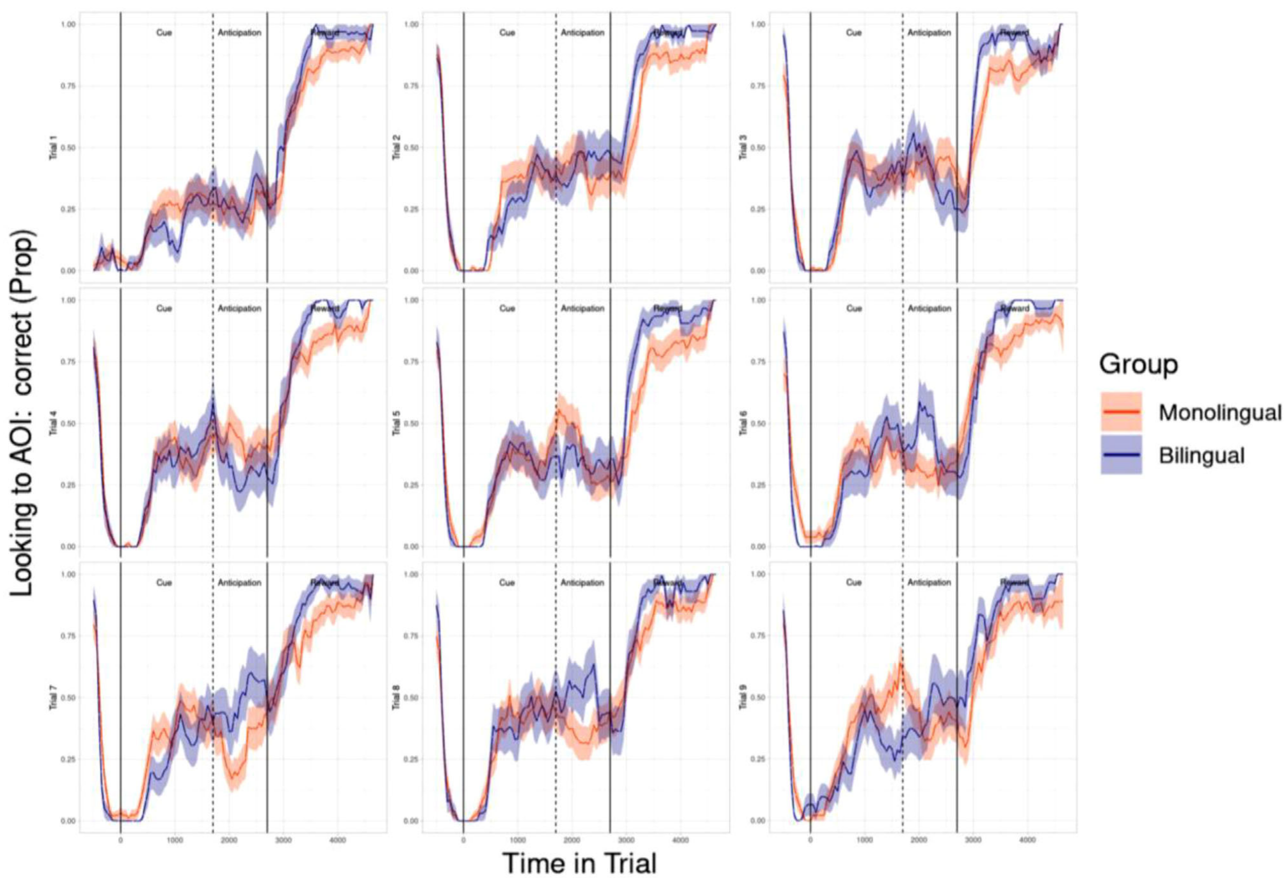

FIGURE 4 Proportion of looks toward the target picture over time for the pre-switch trials.

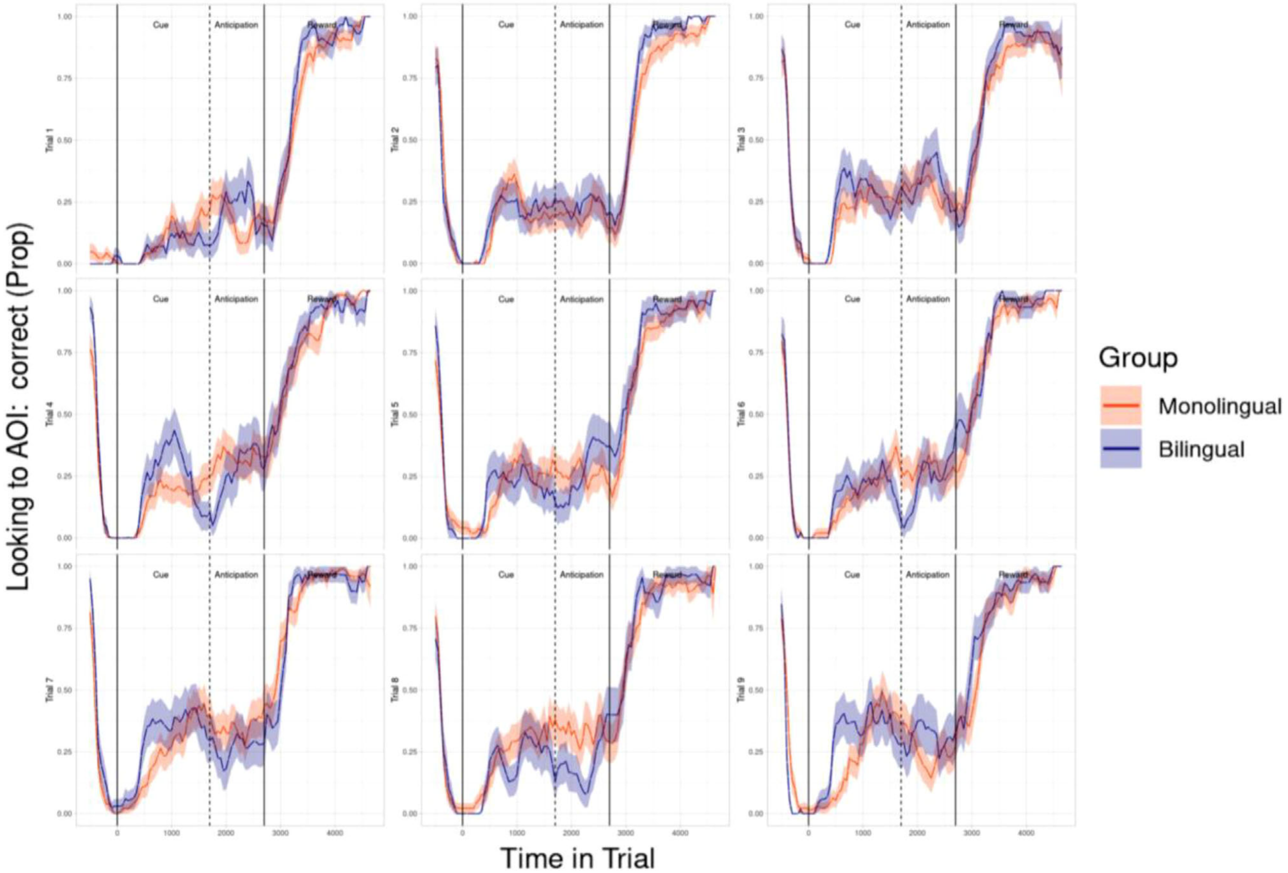

FIGURE 5 Proportion of looks toward the target picture over time for the post-switch trials.

one group of infants looked slightly more towards the target side of the screen than the other group, none of the differences in these intervals were significant in the second step of the cluster based permutation analyses in which we calculated *p*-values for intervals in which we observed a group difference. These results thus provide no evidence that bilingual and monolingual infants differed in anticipating where the reward would occur at any point in time.

However, these analyses do offer some sort of sanity check with regard to whether the stimuli were interesting enough for children to pay attention to in the first place. During each trial, virtually no looks were directed at the target at the start of the trisyllabic sequence, when the attention grabber was playing in the center of the screen. Moreover, looks towards the target quickly reached ceiling scores in each of the trials when the reward was shown on the screen. Infants thus showed interest in the reward, whenever it was visible. Additional analyses in our supplementary materials, in which we grouped trials for the respective phases together, show that during the pre-switch trials bilingual children looked significantly more towards the target during an interval that lasted from 3350 ms until 4050 ms after the start of the trial ($p = .029$), which fell entirely within the reward period. No such differences between the groups were observed for the post-switch trials.

TABLE 2 Results of the cluster based permutation analysis for the association trials.

Time cluster	*Sum statistic*	*Time range in ms*	Probability
1	−7.43	2950–3100	.339
2	−2.04	3150–3200	.737
3	−4.44	3450–3550	.504

Note: The time range is measured from the start of the trial. No clusters turned out significant after the second step of the analysis. Negative sum statistics indicate that the bilingual group shows more looks towards the target during the particular cluster.

5.4 | Additional analysis 3: Association trials

Our final analysis concerned a cluster-based permutation analysis on the 18 trials that were added to the original 18 trials. Here we wanted to see whether infants would use the auditory stimulus as a cue for anticipating where the visual reward would appear. Results from this final cluster-based permutation analysis are shown in Figure 6 and Table 2. In accordance with the lack of differences in switching between the groups during the first 18 trials, we did not find any such differences during the association trials either. There were several

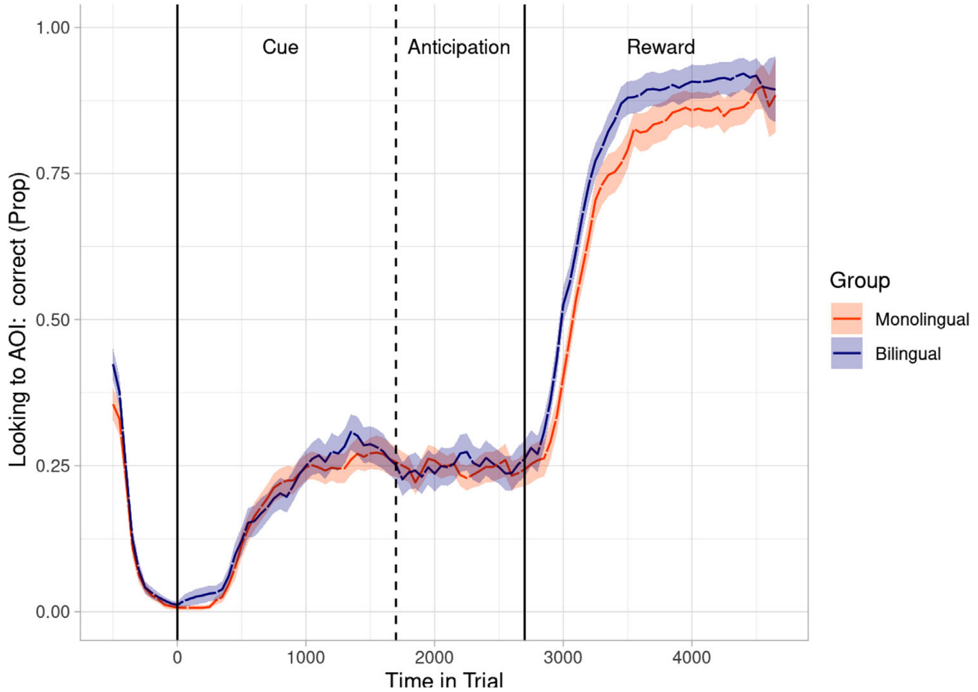

FIGURE 6 Proportion of looks toward the target picture over time for the additional association trials.

small intervals where bilingual children looked slightly more towards the target than monolingual children, but these all occurred when the reward was already on the screen, and these differences were not significant.

6 | DISCUSSION

In our replication of Kovács and Mehler (2009), we did not find evidence that a cognitive advantage is present in 7-month-old bilingual children. We conducted an exact replication of the original switch task (i.e., Experiment 2 in Kovács and Mehler, 2009), in which anticipatory looks were measured. In the original study, all infants became better at anticipating where a reward would occur during the course of pre-switch trials, but bilingual infants were better at anticipating this during post-switch trials than monolingual infants. When carrying out the original ANOVAs for each separate phase, we observed a pattern that did not match with the original results: we found no evidence for increased anticipations during the pre-switch phase, but we did see increased anticipations during the post-switch phase. Importantly, we did not observe any significant interactions between group and block during either phase, which had been taken as indicative of better switching in bilingual children in the original paper. However, because it is considered undesirable to run split analyses in order to investigate whether only bilinguals adapt their behavior during one part of the experiment (Bland & Altman, 2011; Makin & Orban de Xivry, 2019; Nieuwenhuis et al., 2011), we also conducted Bayesian analyses in which we directly tested the three-way interaction of interest (between trial type, group, and trial number) and in which we could take each trial into account in the analyses rather than aggregated trials over blocks. Moreover, we ran cluster-based permutation analyses, which made better use of the richness of the eye-tracking data, because we could track eye movements over the course of each of the trials rather than reducing it to a categorical variable (anticipatory look/no anticipatory look). Since the combination of these analyses is more appropriate for addressing the original research question of whether bilinguals show a cognitive advantage, we will mainly focus on those outcomes in the remainder of the discussion.

The additional analyses of the first two phases did not yield a substantial difference between groups at any point in the experiment. Our Bayesian mixed-effects model, which could be conducted without taking into account trials for which looks were missing in this period and encompassed a direct test of the three-way interaction between trial number, linguistic group, and trial type, showed that both mono- or bilinguals became better at anticipating where the reward would occur as trials progressed, and that they did this in both parts of the experiment. As such, the current findings add to earlier replication attempts of the Kovács and Mehler study that did not replicate the original results (Dal Ben et al., 2022; D'Souza et al., 2020; Kalashnikova et al., 2021), and are in line with the idea that claims about the existence of a bilingual advantage are not robustly substantiated (e.g., Paap & Greenberg, 2013). Moreover, our results suggest that it is unlikely that the slightly altered cue stimuli that were used in D'Souza et al. (2020) and Kalashnikova et al. (2021) are the reason that these studies failed to replicate the original results by Kovács and Mehler (2009). Following the reasoning of the original study, these results imply that we have no evidence that bilinguals are better switchers than monolinguals, which would have indicated a cognitive advantage.

We also ran cluster-based analyses, in which we took a more precise area of interest for anticipatory looks; we compared looks towards the target white box in which the stimulus would appear versus other looks towards the screen. The key advantage of cluster-based analyses is that they can take many more data points into account than the other analyses we conducted, enabling us to track infants' looks over the course of each of the trials. Although we could not inferentially test for group differences across the course of the experiment, our analyses and plots thus serve a descriptive goal. The results of these analyses suggested that both groups behaved similarly: neither group became better at anticipating where the stimulus would occur over the course of the experiment than the other group, suggesting that neither monolinguals nor bilinguals demonstrated a cognitive advantage over the other group. Only the finding that bilinguals tended to look more towards the target during the reward period of pre-switch trials hinted at a group difference, but this occurred during a phase of the experiment that was not relevant for any of the hypotheses. Moreover, across groups there were fewer anticipatory looks during post-switch trials than during pre-switch trials. This could indicate that infants in general were not very good at switching, but might also show that attention decayed during the second part of the experiment. In any case, we did not observe any meaningful group differences.

An analysis of the 18 trials that we added to the original 18 trials yielded results that are in line with these findings. During these additional association trials the reward would occur randomly on the left or right side of the screen, but the mapping between cue and reward was kept intact. These trials were added to investigate whether the bilinguals might be better able to build an association between the linguistic structure of the cue and the location of the reward during the post-switch trials. Yet, since we did not see group differences in the first 18 trials in switching behavior, it was not surprising that we also did not see any group differences in anticipatory looks during the 18 association trials.

The cluster-based analyses also indicated that looking behavior in general was reactive: children quickly reached ceiling levels of target looks, but only once the visual reward stimulus appeared on the screen. The fact that the infants became interested in the visual stimulus means that the stimulus itself was sufficiently rewarding for the children; hence, we can conclude that they were paying attention to what was happening on the screen. However, this reactive looking behavior warrants the question of whether we can expect children of this age to show strong anticipatory looking behavior when they learn novel spatially-cued mappings in a fixed familiarization period. Perhaps this type of fast-mapping of multimodal information is asking too much from these young infants (Kriengwatana et al., 2015), making it difficult to observe strong group differences in their anticipatory looking data. This could mean that this method may not be suited to investigate whether there is a cognitive advantage in this age group. This reasoning would be in line with the common observation that the presence and size of the effects found in tasks that compare mono- and bilinguals depend on various factors such as task parameters and age (Ware et al., 2020).

In this study we investigated the data in three different ways, which all showed a lack of group differences, but also yielded unique insights. The contrast between the results of the original analyses and our additional analyses highlights the need for more advanced statistical approaches in general, especially when it comes to the rich eye tracking data that were collected in this particular experiment. Indeed, one recent replication (Dal Ben et al., 2022) analyzed group differences over the course of the experiment and found that while monolinguals anticipated in the first part of the experiment, bilinguals only anticipated in the latter part of the experiment, showing a difference in processing. However, our additional analyses are also not in line with this alternative finding. Yet, since neither the studies that Dal Ben et al. re-analyzed nor their added conceptual replication were exact replications, it is difficult to pinpoint the precise cause of these different results across studies.

Another reason for our inability to replicate the original findings could lie in the language background of the children. Classifying individuals as either mono- or bilingual has proven to be debatable (Byers-Heinlein, 2015; Luk & Bialystok, 2013). Bilingualism is a multifaceted phenomenon in which many factors play a role, including language use and language exposure. These dimensions are typically gradual with no clear cut-off points. Using a criterion like language use in order to assess bilingualism in children of this age is not possible, which in turn makes it difficult to assess whether they are 'sufficiently bilingual' to show a possible cognitive advantage. Also, the amount of input is not the only factor that determines to what extent children monitor their languages at home; qualitative aspects of children's input play a role as well, but we did not capture these in our questionnaire. Furthermore, even if we would have captured these infants' linguistic status correctly, their language background is different from those in Kovács and Mehler (2009). Our sample was quite heterogeneous in terms of linguistic combinations, but homogeneous in terms of the educational background of the parents. It is not completely clear how our sample can be compared to the sample of Italian-Slovenian infants in the original study, but a one-to-one comparison might not be completely justified. Studies have shown that cultural or linguistic differences might affect cognitive advantages (Sabbagh et al., 2006; Ware et al., 2020) and this could have influenced our inability to replicate the original findings.

Our replication of the seminal Kovács and Mehler (2009) study stresses the importance of exact replications in the field of developmental science in multiple labs (Frank et al., 2017; Geambașu et al., 2022; The ManyBabies Consortium, 2020) in combination with more advanced statistical techniques (Bayes; Cluster-based; logistic regression). Our inability to replicate the original findings not only for this study, but also for another infant study on rule learning (i.e., Marcus et al. (1999)—see Geambașu et al., 2022) highlights the necessity of repeated replication attempts to obtain a better understanding of the factors that may or may not contribute to the building blocks of infant cognition.

ACKNOWLEDGMENTS

This research was funded by the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek; Grant Number 401.18.044) under a funding stream for replication studies. Our project was funded under the title "The building blocks of cognition: Core debates in infancy research". We would like to thank the infants and their caregivers, research assistants, and lab managers who have made this study possible. We also thank Ágnes Kovács for her willingness to share the original stimuli with us, and for our open discussion about the results of this study.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data, stimuli and analysis scripts are available on our OSF page: https://osf.io/p4dwu.

ETHICS STATEMENT

The proposed research has been approved by the Ethics Committees of each participating university with the following approval codes: Amsterdam: 2020-DP-11839; Leiden: 2019/02; Nijmegen: ECSW2017-3001-470; Utrecht: geamb001-01-01-2020.

ORCID

Andreea Geambașu https://orcid.org/0000-0002-5883-9875
Ingmar Visser https://orcid.org/0000-0003-3855-2778

ENDNOTES

[1] For instance, an infant who received 90% exposure to Dutch and 10% exposure to English was regarded as monolingual. An infant with 30% exposure to Dutch, 40% exposure to English, and 30% exposure to Spanish was included as bilingual. We would actually consider these infants as multilingual, but we here followed the classification used in the original study. An infant with 80% exposure to Dutch and 20% to English could not be categorized and was therefore not included in the analysis of our replication study.

[2] We report age in days rather than in months; days, because a specific month; day notation might refer to different numbers of days, depending on the month a child was born in.

[3] From the 143 children that were tested, 76 were monolingual (42 boys and 34 girls, $M = 232$ days old, $SD = 9$ days, min-max = 208–245 days, M caregivers' education = 3.84, $SD = 0.30$, range: 3.0–4.0), 46 were bilingual (23 boys and 23 girls, $M = 227$ days old, $SD = 9$ days, min-max = 208–246 days, M caregivers' education = 3.93, $SD = 0.20$, range: 3.0–4.0), and 12 were unbalanced bilingual (8 boys and 4 girls, $M = 233$ days old, $SD = 9$ days, min-max = 216–244 days, M caregivers' education = 3.68, $SD = 0.34$, range: 3.0–4.0). For the final 9 participants, questionnaire data was missing.

[4] Only calibration stimuli differed, also between our four labs, depending on hardware and software differences.

REFERENCES

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*(4), 689–698. https://doi.org/10.1017/S1366728916000225

Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics*, *5*, 395–415. https://doi.org/10.1146/annurev-linguistics-011718-011820

Baguley, T. (2012). *Serious stats*. Palgrave Macmillan.

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

Beard, E., Dienes, Z., Muirhead, C., & West, R. (2016). Using Bayes factors for testing hypotheses about intervention effectiveness in addictions research. *Addiction*, *111*(12), 2230–2247. https://doi.org/10.1111/add.13501

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(1), 3–11. https://doi.org/10.1017/S1366728908003477

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *65*(4), 229. https://doi.org/10.1037/a0025406

Bialystok, E. (2015). Bilingualism and the development of executive function: The role of attention. *Child Development Perspectives*, *9*(2), 117–121. https://doi.org/10.1111/cdep.12116

Bialystok, E., Craik, F. I., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage*, *24*(1), 40–49. https://doi.org/10.1016/j.neuroimage.2004.09.044

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10*(3), 89–129. https://doi.org/10.1177/1529100610387084

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the simon task. *Psychology and Aging*, *19*(2), 290–303. https://doi.org/10.1037/0882-7974.19.2.290

Bland, J. M., & Altman, D. G. (2011). Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials*, *12*, 264. https://doi.org/10.1186/1745-6215-12-264

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish–Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. https://doi.org/10.1016/j.jecp.2014.06.007

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

Bürkner, P. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Byers-Heinlein, K. (2015). Bilingual advantages, bilingual delays: Sometimes an illusion. *Applied Psycholinguistics*, *35*, 902–905. https://doi.org/10.1017/S0142716414000204

Byers-Heinlein, K., & Lew-Williams, C. (2013). Bilingualism in the early years: What the science says. *LEARNing landscapes*, *7*(1), 95.

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

Comishen, K. J., Bialystok, E., & Adler, S. A. (2019). The impact of bilingual environments on selective attention in infancy. *Developmental Science*, *22*(4), e12797. https://doi.org/10.1111/desc.12797

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*(2), 135–149. https://doi.org/10.1016/j.cognition.2009.08.001

Dal Ben, R., Killam, H., Pour Iliaei, S., & Byers-Heinlein, K. (2022). Bilingualism affects infant cognition: Insights from new and open data. *Open Mind*, *6*, 88–117. https://doi.org/10.1162/opmi_a_00057

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*(1), 99–107. https://doi.org/10.1177/0956797614557866

De Cat, C., Gusnanto, A., & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in*

Second Language Acquisition, *40*(1), 119–151. https://doi.org/10.1017/S0272263116000486

Dink, J. W., & Ferguson, B. 2015. eyetrackingR; An R library for eye-tracking data analysis. http://www.eyetrackingr.com

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental Psychology*, *61*(3), 234–251. https://doi.org/10.1027/1618-3169/a000243

D'Souza, D., Brady, D., Haensel, J. X., & D'Souza, H. (2020). Is mere exposure enough? The effects of bilingual environments on infant cognitive development. *Royal Society Open Science*, *7*(2): 180191. https://doi.org/10.1098/rsos.180191

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

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

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas-Guasch, N., Roberts, E. J., Hughes, E. K., & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*(11), 1–13. https://doi.org/10.3389/fpsyg.2014.00011

Geambașu, A., Spit, S., van Renswoude, D., Blom, E., Fikkert, P., Hunnius, S., Junge, C., Verhagen, J., Visser, I., Wijnen, F., & Levelt, C. C. (2023). Robustness of the rule-learning effect in 7-month-old infants: A close, multicenter replication of Marcus et al. (1999). *Developmental Science*, *26*(1), e13244. https://doi.org/10.1111/desc.13244

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-aoas191

Gelman, A. (2020). Prior choice recommendations. Blog post. https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations#5-levels-of-priors

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, *18*(4), 625–658. https://doi.org/10.3758/s13423-011-0116-7

Kalashnikova, M., Pejovic, J., & Carreiras, M. (2021). The effects of bilingualism on attentional processes in the first year of life. *Developmental Science*, *24*(2), e13011. https://doi.org/10.1111/desc.13011

Kapa, L. L., & Colombo, J. (2013). Attentional control in early and later bilingual children. *Cognitive Development*, *28*(3), 233–246. https://doi.org/10.1016/j.cogdev.2013.01.011

Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.0. https://CRAN.R-project.org/package=rstatix

Kovács, A. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Psychological and Cognitive Sciences*, *106*(16), 6556–6560. https://doi.org/10.1073/pnas.0811323106

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497–514. https://doi.org/10.1080/20445911.2013.799170

Kriengwatana, B., Junge, C., Chladkova, K., Mulak, K., Bonn, C., Aslin, R., & Escudero, P. (2015, June 10–12). Keep looking: No robust anticipatory looking paradigms yet in infant speech sound learning. Workshop on infant language development 2015 [Poster presentation]. Department of Linguistic, Stockholm, Sweden.

Lee, M., & Wagenmakers, E. (2013). *Bayesian modeling for cognitive science: A practical course*. Cambridge University Press.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, *144*(4), 394–425. https://doi.org/10.1037/bul0000142

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Makin, T. R., & Orban de Xivry, J. J. (2019). Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. *eLife*, *8*, e48175. https://doi.org/10.7554/eLife.48175

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

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *38*(4), 843–847.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. https://doi.org/10.1038/nn.2886

van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., Vermeire, K., Lee, S. H., & Lim, S. (2019). Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? A systematic review. *Behavioral Sciences*, *9*(3), 27. https://doi.org/10.3390/bs9030027

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(22), 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R, Myuz, H., & 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. https://doi.org/10.1016/j.cortex.2015.04.014

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. Behavior Research Methods. https://doi.org/10.3758/s13428-018-01193-y

Poarch, G. J. (2018). Multilingual language control and executive function: A replication study. *Frontiers in Communication*, *3*(46). https://doi.org/10.3389/fcomm.2018.00046

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*(4), 535–551. https://doi.org/10.1016/j.jecp.2012.06.013

R Core Team. (2015). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind. A comparison of Chinese and U.S. preschoolers. *Psychological Science*, *17*(1), 74–81. https://doi.org/10.1111/j.1467-9280.2005.01667.x

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Siegler, R. S., Eisenberg, N., DeLoache, J. S., & Saffran, J. (2014). *How children develop*. Worth Publishers.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, *81*(1), 174–176. https://doi.org/10.1037/h0027448

Spit, S., Andringa, S., Rispens, J., & Aboh, E. O. (2022). The effect of explicit instruction on implicit and explicit linguistic knowledge in kindergartners. *Language Learning and Development*, *18*(2), 201–228. https://doi.org/10.1080/15475441.2021.1941968

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, *18*(1), 3–24. https://doi.org/10.1017/S1366728914000522

Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*(3), 842–860. https://doi.org/10.1111/cdev.12169

Verhagen, J., de Bree, E., & Unsworth, S. (2020). Effects of bilingual language use and language proficiency on 24-month-olds' cognitive control. *Journal of Cognition and Development*, *21*(1), 46–71. https://doi.org/10.1080/15248372.2019.1673752

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Ware, A. T., Kirkovski, M., & Lum, J. (2020). Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Frontiers in Psychology*, *11*, 1458. https://doi.org/10.3389/fpsyg.2020.01458

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

Yang, S., & Yang, H. (2016). Bilingual effects on deployment of the attention system in linguistically and culturally homogeneous children and adults. *Journal of Experimental Child Psychology*, *146*, 121–136. https://doi.org/10.1016/j.jecp.2016.01.011

Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Frontiers in Psychology*, *2*, 210. https://doi.org/10.3389/fpsyg.2011.00210

How to cite this article: Spit, S., Geambașu, A., Renswoude, D., Blom, E., Fikkert, P., Hunnius, S., Junge, C., Verhagen, J., Visser, I., Wijnen, F., & Levelt, C. C. (2023). Robustness of the cognitive gains in 7-month-old bilingual infants: A close multi-center replication of Kovács and Mehler (2009). *Developmental Science*, e13377. https://doi.org/10.1111/desc.13377