MASTERARBEIT | MASTER'S THESIS

Titel | Title

Does Early Bilingualism Affect Habituation and Dishabituation Processes?

verfasst von | submitted by

Lea Halmer BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Psych. Dr. Stefanie Höhl

Table of Content

ABSTRACT

There is an ongoing debate about whether bilingualism provides cognitive benefits, particularly in terms of attention, memory, and learning. However, the evidence is mixed, and more research is needed, especially for young children. This study examines the effects of early bilingualism on habituation and dishabituation processes in infants. Habituation refers to the reduction of attention to a repeated stimulus, whereas dishabituation occurs when a novel stimulus regains attention. Two experiments were conducted with 7- to 8-month-old monolingual and bilingual infants using an infant-controlled habituation-dishabituation paradigm. The study aimed to replicate and extend previous findings by Singh et al. (2015), which suggested that bilingual infants may process information more efficiently, leading to faster habituation and stronger preference for novel stimuli. However, we did not find significant differences between bilingual and monolingual infants in the speed of habituation, strength of habituation, or dishabituation response. The results suggest that the expected cognitive advantages of bilingualism were not observed, possibly due to limitations such as small sample size and low statistical power. This study highlights the need for further research with larger sample sizes to better understand the impact of bilingualism on early cognitive development.

Keywords: *early bilingualism, information processing, habituation, dishabituation, bilingual advantage, infant cognition*

Introduction

Babies develop and learn at an amazing rate during the first few years of life. An incredible amount is happening in the brain that lays the foundation for future development, particularly in the areas of cognition and conceptual thinking. Even before they can speak, babies develop a rich conceptual system that allows them to categorize objects and actions, solve problems, recall events, and generalize learned behaviors across contexts. These cognitive developments are not only crucial for language acquisition, but also lay the foundation for other learning processes (Mandler, 2004). Bilingualism plays a particularly interesting role in this context, as there has been an ongoing debate for several years about whether there is a cognitive advantage to growing up bilingual. However, many studies have either found null results or those showing a significant bilingual advantage have had small sample sizes. This raises the question of whether the phenomenon exists at all, or whether it only occurs in very specific and indeterminate circumstances (Paap et al., 2015).

The original explanation for the bilingual advantage comes from Green's (1998) inhibitory control model. This model describes how bilingual speakers manage to switch between languages by suppressing the language that is not needed at the moment. However, according to the inhibitory control model, the bilingual advantage should only be observed in children who are already producing language (D'Souza et al., 2020). Nevertheless, early exposure to two languages has been shown to affect memory, cognition, and learning as early as the first year of life. This has been found in preverbal infants not only in functional areas relevant to language acquisition, such as visual language discrimination (Sebastián-Gallés et al., 2012) but also in more basic attentional processes with non-linguistic stimuli, such as in cognitive control (Kovács & Mehler, 2009, Experiment 3) or information processing (Singh et al., 2015). Because this early bilingual advantage cannot be explained by the inhibitory control model, researchers suggest that growing up in a bilingual environment, and thus being exposed to more diverse and less predictable input than monolinguals, leads to cognitive advantages in attention, encoding, and generalization (Singh et al., 2024).

In this master's thesis, we would like to investigate differences in information processing in bilingual infants compared to monolingual infants - more specifically, we are interested in *whether early bilingualism affects habituation and dishabituation processes.*

Using Infant Gaze in the Habituation Paradigm

The question is how to study these early learning processes, since babies cannot tell us directly what is going on inside their heads. Researchers have developed creative and innovative ways to study what babies learn and how their brains work. One highly effective approach to studying cognitive development in early childhood is to observe how infants direct their attention to visual stimuli. Infants' gaze is often used to measure their preference for one stimulus over another. Researchers use looking-time methods to measure a variety of topics, such as visual and auditory perception, or language, conceptual, and social development. This approach makes it possible to include very young infants and even newborns in these studies (Pahnke, 2007, Bergmann et al., 2019).

One of the most used methods involving infant gaze is the *habituation-dishabituation paradigm*. Infants naturally tend to focus on meaningful stimuli in their environment, allowing researchers to study early visual learning processes (Pahnke, 2007; Colombo & Mitchell, 2009; Oakes, 2010; Zaharieva et al., 2021). *Habituation* refers to the gradual decrease in attention to stimuli that are repeatedly presented, whereas *dishabituation* describes the renewed interest in a novel stimulus after successful habituation. (e.g., Zaharieva et al., 2021).

Theoretical Models of Habituation

According to Sokolov's (1963a) popular comparator model, non-threatening, moderately intense novel, or unexpected stimuli elicit an *orienting reflex* (Colombo & Mitchell, 2009; Kavšek, 2013). After repeated exposure to the stimulus, an internal representation is formed. When there is a discrepancy between the stimulus and its representation, another orienting reflex is activated. As the discrepancy decreases, so does the orienting reflex, and the infant is fully habituated. Thus, habituation leads to the construction of a memory trace through repeated exposure to a stimulus until the stimulus is no longer perceived as novel. A dishabituation stimulus induces a reactivation of attention by comparing the incoming information of the novel stimulus with the memory trace of the habituation stimulus (Kavšek, 2013). For this reason, habituation is considered the simplest form of learning (Pahnke, 2007). The theory proposes that differences in the rate at which individuals' orienting reflexes decline indicate differences in the rate at which infants learn. This leads to the distinction between fast and slow habituating infants (Colombo & Mitchell, 2009).

In contrast, the dual-process theory by Thompson and Spencer (1968; Groves & Thompson 1970) distinguishes between two independent opposing processes, habituation and sensitization. Whereas habituation describes the decrease in attention with repeated exposure to a stimulus, *sensitization* refers to the increase in arousal during the early habituation phase depending on stimulus intensity. Higher levels of sensitization not only extend the habituation phase, but also increase attention to the dishabituation stimulus (Kavšek, 2013). Stimulus complexity increases sensitization, while infant age also produces different sensitization responses to different stimuli (Colombo & Mitchell, 2009). According to the theory, habituation is specific to the stimulus, whereas sensitization reflects a general arousal due to a novel or intense stimulus. The dishabituation response does not reflect a discrimination between the habituation and dishabituation stimuli, because there is no internal representation of the stimulus as proposed by Sokolov. Therefore, the dual-process theory does not explain individual differences in attentional processes (Pahnke, 2007).

Hunter and Ames (1988) discuss the factors that contribute to either familiarity or novelty preferences in infants. In their model, the authors argue that infants develop a preference for a repeatedly presented stimulus over a novel stimulus and explore that stimulus until it is fully encoded. When faced with a choice between a partially encoded stimulus and a completely novel stimulus, infants prefer the familiar stimulus. Once the stimulus is fully encoded, these preferences eventually switch, and infants prefer the novel stimulus. However, habituation time, infant age, task difficulty, and their interactions influence the extent to which infants prefer the familiar stimulus over the novel stimulus (Bergmann et al., 2019; Kosie et al., 2023).

Research suggests that bilingual infants may process and encode information differently because of their early exposure to two languages – the early bilingual advantage. This exposure may influence how quickly bilinguals encode stimuli and how strongly they remember them. This study will contribute to the ongoing debate by examining whether the cognitive advantages associated with bilingualism are evident even before language production begins. Although the habituation-dishabituation paradigm is widely used to measure infant cognition, there are difficulties that should be considered.

Issues with the Habituation-Dishabituation Paradigm

There has been a lot of research on habituation over the years. However, it is still unclear whether different design factors lead to different results. Researchers have focused on novelty preferences to draw conclusions about infants' cognitive processing (Oakes, 2010), although it remains controversial what factors contribute to the emergence of novelty or familiarity effects (Bergmann et al., 2019; Paulus, 2022; Kosie et al., 2023). While most studies expect a novelty preference during the test phase, some studies report stronger familiarity preferences (Bergmann et al., 2019; Zaharieva et al., 2021). For example, Hunter and Ames (1988) suggest that infants who are not fully habituated show a preference for familiarity. The preference for novelty emerges only after the familiar stimulus has been fully encoded (Bergmann et al., 2019). This may bias the overall effect of novelty when these infants are included in the analysis. However, another perspective suggests that as long as infants show a preference for one of the stimuli (familiar or novel), they can discriminate between them, and the valence of the preference is less important (Houston-Price & Nakai, 2004; Bergmann et al., 2019).

In general, interpretation of these results is speculative, as longer viewing times may indicate various psychological states beyond preference, such as fear, anticipation, surprise, annoyance, or skepticism (Paulus, 2022). The fact that longer viewing of novel stimuli does not always result in better encoding or recognition during test suggests that the quality of attention may be more important for information processing than the mere duration of attention (Colombo et al., 2001; Begus et al., 2015). Even small changes in perception can lead to different results (Schöner & Thelen, 2006; Paulus, 2022). Different methodological factors, such as lighting or noise, can systematically influence infants' preferences for novel or familiar stimuli. Unfortunately, these factors are often not reported and are therefore difficult to detect (Bergmann et al., 2019) It is also unclear whether longer looking times for novel stimuli are caused by a preference for novelty or by avoidance of the familiar stimulus (Kucharský et al., 2022).

Nevertheless, many studies continue to make inferences about infants' cognitive processing based on novelty preferences (Oakes, 2010). This is mainly because the phenomenon of habituation itself has not been sufficiently studied. Habituation-based measures are applied beyond their original purpose, leading to conclusions that may not be fully justified. Such assumptions can lead to overinterpretation of results, with researchers attributing advanced cognitive skills to infants based on looking time alone without sufficient evidence (Paulus, 2022). Of 777 publications from 1962-2008, only one in four discusses habituation per se. However, 75% use habituation as a tool to habituate infants to a visual stimulus, e.g., to investigate cognitive abilities or skills (Colombo & Mitchell, 2009). Alternative explanations for attentional decline, such as sensory adaptation or fatigue, are often overlooked (Hartkopf et al., 2019). Additionally, habituation is influenced by various factors, including the complexity of the stimuli, age, gender, intervals between trials, the number of trials, and the method used to calculate habituation (Oakes, 2010). To address these gaps, the ManyBabies 5 initiative (Kosie et al., 2023) is conducting a large-scale study to evaluate Hunter and Ames' (1988) model of gaze preferences.

Finally, although the habituation-dishabituation paradigm has been used in numerous studies since the 1960s, there is still no single procedure that all studies can follow. The factors that contribute to novelty or familiarity preferences are still debated. Recent research claims that we should focus on the habituation process itself (Kucharský et al., 2022). For these reasons, we would like to take the opportunity in this master's thesis to investigate the habituation process itself. While previous research has focused primarily on specific cognitive skills, our study shifts the focus to the effects of early bilingualism on basic information processing and examines whether the cognitive advantages observed in bilingual infants extend beyond mere language skills. This research is particularly important because it addresses the broader implications of bilingualism for cognitive development and aims to clarify whether these advantages are real and measurable.

Effects of Early Bilingualism

Early theories suggested that the cognitive advantages of bilingualism arise from the constant need to switch between languages and to inhibit one language while using the other, as suggested by the inhibitory control model (Green, 1998). However, more recent perspectives suggest that the environmental differences that bilingual infants encounter, such as exposure to a wider range of sounds and grammatical systems, may have broader implications for nonverbal cognitive development (Bialystok, 2017; Singh et al., 2024).

Bilingual infants are likely to receive less input from each language than monolingual infants receive from their single language. They are also exposed to more varied and less predictable input when learning two languages because of different sound or grammatical systems. Therefore, bilingual infants may need to process a wider range of information (D’Souza & D’Souza, 2021; Singh, 2021). However, the cognitive and sensorimotor systems of bilingual and monolingual children are likely to develop at similar rates (Kovács & Mehler, 2009). Therefore, it is assumed that lower input for the same output in bilingual learners indicates a higher efficiency of information processing. When bilingual infants attend to their environment more broadly (e.g., by taking in more visual information such as lip movements; D'Souza & D'Souza, 2021), the increased information may help them identify relevant aspects and generalize learned information to new contexts. Overall, the more complex environment (due to a greater number of sounds, words, grammatical patterns) faced by bilingual infants may lead to increased orientation to novelty (Singh et al., 2023).

In summary, bilingual infants receive more varied and less predictable linguistic input, which may enhance their perceptual flexibility and information processing. Studies show that bilingual infants can visually discriminate between spoken languages, without hearing them (Sebastián-Gallés et al., 2012) and show faster habituation and a stronger preference for novelty than monolinguals (Singh et al., 2015). These studies suggest that bilingual infants show different patterns of attention to visual tasks than monolingual infants.

Mixed Evidence on the Bilingual Advantage

As reflected in various meta-analyses and systematic reviews, the debate about the cognitive benefits of bilingualism remains complex. These studies present a mixed picture, suggesting that while bilingualism may offer some benefits in executive functioning, the size and consistency of these effects are far from clear. Executive functioning describes a range of cognitive abilities that can be measured, for example, by inhibiting responses, switching between tasks, and working memory tasks (Tsui & Fennell, 2019). In their meta-analysis, Donnelly et al. (2015) found moderately significant effects of bilingualism on conflict resolution tasks, such as the Simon, Flanker, and Stroop tasks. These are designed to measure inhibitory control in children and adults. However, their study did not address publication bias, which is an important factor that could bias the results and exaggerate the perceived benefits of bilingualism. While some meta-analyses initially found evidence of a bilingual advantage in children's executive function, these effects were not significant after correcting for publication bias (Gunnerud et al., 2020, Lowe et al., 2021). A review by van den Noort et al. (2019) provides a broader perspective, showing that more than half of the studies reviewed reported beneficial effects of bilingualism on cognitive control tasks. However, nearly a third of the studies showed mixed results, and about 20% found no evidence of a bilingual advantage in children and adults. This variability in results highlights the complexity of the bilingualism debate and suggests that the cognitive benefits of bilingualism are not universal.

Taken together, these findings suggest that while bilingualism may confer some cognitive benefits, these benefits are not consistently observed and may be influenced by various factors, including publication bias and study conditions. Further research is needed to understand how bilingualism affects cognitive development and whether the observed effects are truly due to language experience or other developmental factors. In particular, there is a need for studies that focus on younger children and infants, as there are currently no meta-analyses in this age group (Paap et al., 2019). For this reason, it is crucial to review studies that focus on the cognitive benefits of bilingualism in young children. These studies provide valuable insights into how early exposure to multiple languages might influence cognitive development from a young age, although the evidence for a bilingual advantage in infancy remains mixed.

Study Replication Issues

Understanding the bilingual advantage requires examining how bilingualism affects not only language skills but also basic cognitive processes. Despite the valuable insights provided by existing research, inconsistencies in findings and challenges in replication remain significant issues in the field of psychology, particularly in research with young children. Therefore, a closer look at key studies in the field is necessary to better understand these effects.

One of the first studies to investigate the bilingual advantage in infants was Kovács & Mehler (2009, Experiment 3). Eye tracking was used to assess cognitive control in 7-month-old monolingual and bilingual infants using a "switch" task. Visual patterns of circles and triangles (AAB or ABB) predicted the location of a visual reward. After switching the patterns and changing the location of the reward, only bilingual infants successfully adapted their anticipatory gaze to the new location and demonstrated better cognitive control by suppressing their responses to the previously learned location. However, attempts to replicate these findings, for example by D'Souza et al. (2020, Experiment 1), Kalashnikova et al. (2020), and Spit et al. (2023), were unsuccessful, leaving the bilingual advantage in infants unsupported. Although these studies did not confirm the specific claims about cognitive control, they did find other bilingual advantages, such as increased attentional switching between stimuli (D'Souza et al., 2020, Experiment 4) and longer viewing times of novel rewards (Kalashnikova et al., 2020).

Building on this exploration of cognitive differences between bilingual and monolingual infants, Singh et al. (2015) used a traditional visual habituation task with 6-month-old children to investigate whether monolingual and bilingual children differ in their information processing of visual stimuli (colored images of a bear and a wolf). The authors found that bilingual infants showed a greater decrease in attention during habituation and that their viewing time decreased more rapidly across trials than monolingual infants. This indicates greater efficiency in memory formation. When comparing fixation times on the novel stimulus during the test phase, bilingual infants showed a stronger preference for novel stimuli. As mentioned above, theories of habituation/dishabituation (e.g., Sokolov, 1963a or Hunter & Ames, 1988) suggest that infants prefer familiar stimuli until they are fully learned. Only then can a preference for novelty develop, and the new stimulus be learned. However, bilingual infants may seek out a greater variety of stimuli to determine the patterns and rules of their complex environment, and therefore pay more attention to familiar and novel stimuli than monolingual infants in general (Singh et al., 2024). Overall, bilingualism may speed up the process by which a new stimulus becomes familiar, leading to a preference for novelty (Singh, 2021). Singh et al. (2015) have received considerable attention for their findings on early bilingual advantages. Despite the prominence, there remains a gap in the literature: to date, no direct replications of Singh et al. (2015) have been conducted. This lack of replication limits our ability to generalize these findings.

The Current Study

In the present master's thesis, we would like to further investigate the differences in cognitive processing in bilingual infants compared to monolingual infants. Previous research, such as that of Singh et al. (2015), has shown that bilingual infants are more efficient at stimulus encoding during habituation and develop a stronger preference for novel stimuli. These findings suggest that the more diverse and less predictable linguistic environments that bilingual infants are exposed to may lead to increased perceptual flexibility and faster information processing. The current study aims to provide a conceptual replication of Singh et al. (2015) to test the robustness of their findings and increase the reliability of the results in the literature. Following Singh et al. (2015), *we hypothesize that bilingual infants habituate faster (H1a) and more strongly (H1b) than monolingual infants. Bilingual infants also show a stronger dishabituation response (H2), that is, a stronger preference for novel stimuli.*

Methods

The data used in this master's thesis were collected as part of a doctoral dissertation. We conducted two similar studies to gain new insights into early brain development and learning. Study 1 was called "OREP", which stands for object representations, and Study 2 was called "FLOREP", which stands for flickering object representations. In addition to behavioral data, we also collected EEG data, but these are not examined in this thesis. We recruited 7- and 8-month-old infants living in Vienna from an existing database of the "Wiener Kinderstudien" laboratory. The children were born after the 37th week of pregnancy and had no neurological problems or serious birth complications. The ethics committee of the University of Vienna approved our study design. Participants signed an informed consent form, and we informed them of their right to withdraw at any time. After participation, we debriefed the families about the purpose of the study.

Participants

A total of 46 (30 males and 16 females) 7- to 8-month-old infants participated in Study 1 (OREP). Of these, 33 (20 males and 13 females) participants were monolingual and 13 (10 males and 3 females) were bilingual. The languages spoken in the bilingual homes in addition to German were English (n=2), Bulgarian (n=1), Czech (n=1), Macedonian (n=1), French (n=1), Hungarian (n=1), Serbian (n=1), Italian (n=1), Portuguese (n=1) and Croatian (n=1). Two of the children are growing up trilingual. One family speaks Spanish, English and German and the other family speaks Polish, English and German with their child. An additional four infants were excluded due to tiredness, fussiness, experimenter error and not looking during the test phase.

41 (21 males and 20 females) 7 to 8-month-old infants participated in study 2 (FLOREP). 24 (11 males and 13 females) were monolingual and 17 (10 males and 7 females) were bilingual. The languages spoken in the bilingual homes in addition to German were English (n=1), Hungarian (n=1), Bosnian (n=1), Spanish (n=2), Serbian (n=1), French (n=1), Tagalog (n=1), Croatian (n=2), Romanian (n=1), Bulgarian (n=1), Portuguese (n=1) and Czech (n=1). One family does not speak German but speaks Bosnian and English with

their child. Two of the children are growing up trilingual. One family speaks Farsi, Paschtu and English and the other family speaks Spanish, English and German with their child. Of the 62 infants originally tested, 21 had to be excluded due to technical problems or because they observed less than two trials.

Study Designs: Fixed-Trial versus Infant-Controlled Designs

Before discussing the actual study procedures, we would like to discuss general procedural considerations and design choices. Habituation procedures can be either "fixed-trial" or "infant-controlled" designs. While in the first case each trial has a fixed duration and a fixed intertrial interval, in the second case the infant's gaze behavior determines the start and end of a trial. With a fixed design, the presentations are not tied to the infant's behavior or visual attention. They are therefore easy to automate, run and analyze as the data is consistent for each subject.

However, because individual infants habituate at different rates and may be at very different levels of processing, infant-controlled designs are increasingly being used. In infant-controlled designs, the trial begins when the infant looks at the stimulus and ends when the infant looks away. Researchers use online coding to mark the beginning and end of a trial. For example, during stimulus presentation, a researcher presses the space bar while the child is looking at the stimulus and stops pressing when the child looks away. Also, there is no total number of trials, but infants look at as many trials as necessary until their gaze is reduced to either an absolute (e.g., a total of 3 seconds) or relative (typically a 50% reduction in initial looking time) criterion. We use the habituation criterion to determine whether an infant has reached a significant loss of interest, indicating a state of habituation, so that the test phase of the experiment can begin. Most studies use a 50% decrease from "baseline looking time" (e.g., average of the first 3 trials) to "final looking time" (e.g., average of the last 3 trials) as their relative criterion (also called “3-3-50% criterion”). However, other percentages and looking time definitions are also used, making it difficult to generalize findings across studies (Zaharieva et al., 2021).

Using a relative criterion to measure habituation is consistent with the comparator model: The initial level of the orienting reflex reflects the response to a completely new stimulus for which there is no internal memory. Therefore, using a fixed percentage decrease to measure habituation is theoretically consistent for all individuals, regardless of how well they remember the external stimulus.

There are also problems with infant-controlled designs, which are more difficult to implement because the criteria change many times within a session based on the infant's gaze duration. In addition, different trial lengths per participant must be taken into account when analyzing the data. Moreover, there are no empirically based standards for design parameters or habituation criteria. Many infants do not meet the arbitrary habituation criterion and must be excluded from analyses (Colombo & Mitchell, 2009).

Oakes (2010) recommended to present both novel and familiar trials after habituation to assess the dishabituation effect. This is because presenting only novel trials after habituation may be influenced, at least in part, by regression to the mean. This is because the looking times in the last habituation trial are artificially low. By presenting both the familiar and the novel stimulus, the familiar stimulus serves as a baseline for dishabituation to the novel stimulus.

In summary, habituation procedures can be either "fixed-trial" or "infant-controlled. Fixed-trial designs are easier to automate and analyze, but do not account for individual differences in habituation rates. Infant-controlled designs, in which trials are based on the infant's gaze behavior, better account for these differences but are more difficult to implement and analyze. The habituation criterion, often a 50% reduction in looking time, is used to determine when the infant has habituated, but it lacks standardized parameters and can lead to high error rates. Researchers recommend presenting both novel and familiar stimuli after habituation to better assess dishabituation effects. The choice of an infant-controlled design for our two studies takes advantage of its adaptability to individual differences, its alignment with theoretical models, and its consistency with contemporary research standards. Although more challenging to implement and analyze, this design ultimately provides a more accurate and detailed understanding of infant habituation processes.

Stimuli

The stimuli consisted of 12 different colored cartoon characters (monsters, see Appendix: Figure C1) and 12 matching dishabituation stimuli of the same color but different shape. The order of the stimuli was randomized.

Procedure

Families recruited by telephone from the *Wiener Kinderstudien* database were asked to participate in our studies to gain insight into early learning and memory processes. Parents sat in a chair about 60 cm away from the monitor with their child on their lap. We used an infant-controlled habituation-dishabituation paradigm in both studies. See Figures 1 and 2.

Study 1 - OREP

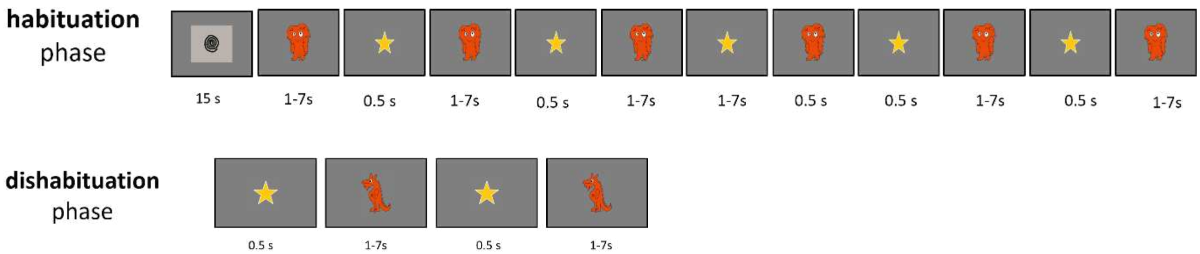
In Study 1, a stimulus was shown several times until the infant looked away (habituation phase). This part was controlled by the infant using online coding. The habituation criterion was reached when the gaze duration was reduced by 50% in three consecutive trials compared to the first three trials. There was a maximum of 12 habituation trials. Finally, in the test phase, a new stimulus of the same color but a different shape was used for 2 trials to test the infant's recovery of attention. At the beginning of each trial, a yellow star with a sound was displayed as an alert to draw the infant's attention to the monitor. Each infant saw 1 to 7 complete phases with different monsters (habituation and test), depending on how long they were satisfied and attentive.

Study 2 - FLOREP

In Study 2, a stimulation phase was added at the beginning, followed by the habituation phase and the test phase. During the stimulation phase, the habituation stimulus was shown five times for 3 seconds, flickering at either 6 Hz or 4 Hz. This resulted in the habituation phase. In Study 2, a maximum of 15 habituation trials were completed. In the test phase, a novel stimulus was shown twice, alternating with the familiar stimulus. The order was randomized (ABAB/BABA). Each infant saw at least two complete phases, one with the 6 Hz and one with the 4 Hz flickering stimulus. In Study 2, the phases could be skipped completely if the infant missed the stimulation by looking away. A new phase would then begin with a novel stimulus. The stimulation phase is not covered in this master's thesis. Only the habituation and test phases are analyzed.

Figure 1

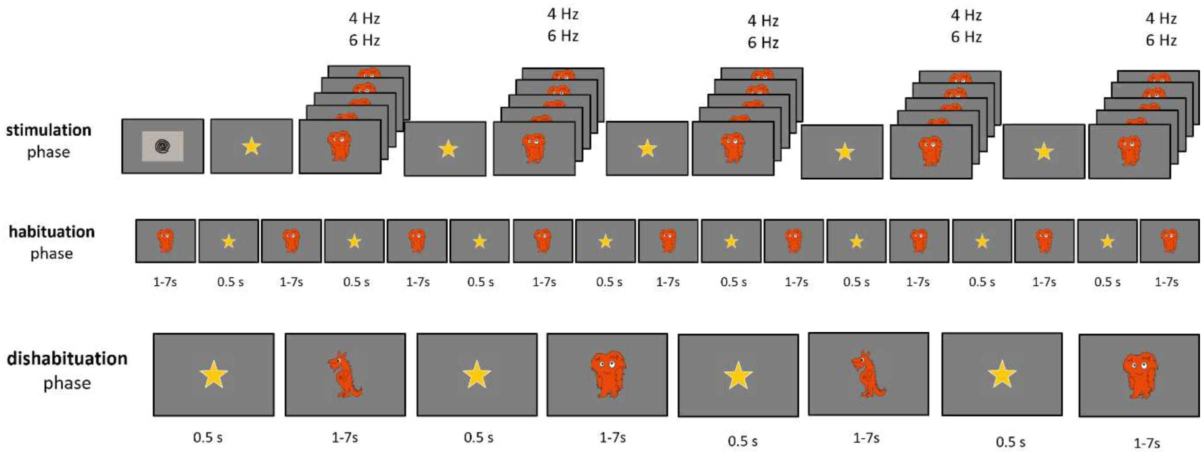
Procedure Study 1 (OREP)



Note. The habituation stimulus was shown for a maximum of 7 seconds. If the infant looked away for more than 1.5 seconds cumulatively, the next trial began. The test phase began after 12 trials or when the habituation criterion was reached (looking time in 3 consecutive trials has decreased by 50% compared to the first 3 trials). The dishabituation stimulus was presented for 2 trials.

Figure 2

Procedure Study 2 (FLOREP)



Note. During the stimulation phase, the habituation stimulus flickered 5 times for 3 seconds. Then the habituation phase began and lasted up to 15 trials or until the habituation criterion was reached (looking time in 3 consecutive trials has decreased by 50% compared to the first 3 trials). During the test phase, the habituation and dishabituation stimuli were shown alternately.

The division of the experiment into two phases (habituation and test) allows us to answer two questions independently: Whether learning occurred during the habituation phase, since the decrease in attention during habituation indicates that the infant fully grasped and memorized the familiar stimulus. And whether a novelty or familiarity effect occurred, since the increase in attention to the novel stimulus in the test phase indicates that the infant can detect differences between the two stimuli (Kucharský et al., 2022).

Data Preparation for Analyses

Trained research assistants coded the video recordings of the experimental sessions frame by frame using Mangold Interact software (Mangold, 2018). Coding consisted of identifying whether or not the infant was looking at the stimulus. We calculated Interrater reliability using Cohen's kappa. It was K=0.9685 for Study 1 and K=0.8563 for Study 2.

Habituation Criteria

Oakes (2010) recommends including only infants who habituated in the analysis. Non-habituated infants are likely to show a familiarity preference and thus reduce the overall novelty preference. Therefore, we only included the phases of participants who successfully habituated in the analysis. We calculated the habituation criterion in two ways. The *"standard criterion*" was reached when the looking time of the last three trials was less than 50% of the looking time of the first three trials. In Study 1, 28 participants reached the standard criterion (18 monolinguals and 10 bilinguals, out of a total of n = 46; in 27% of all the habituation trials). In Study 2, only 15 participants reached the standard criterion (10 monolinguals and 5 bilinguals, out of a total of n = 41; in 13% of all the habituation trials). This means that more than half of the participants in Study 2 did not meet the standard habituation criterion.

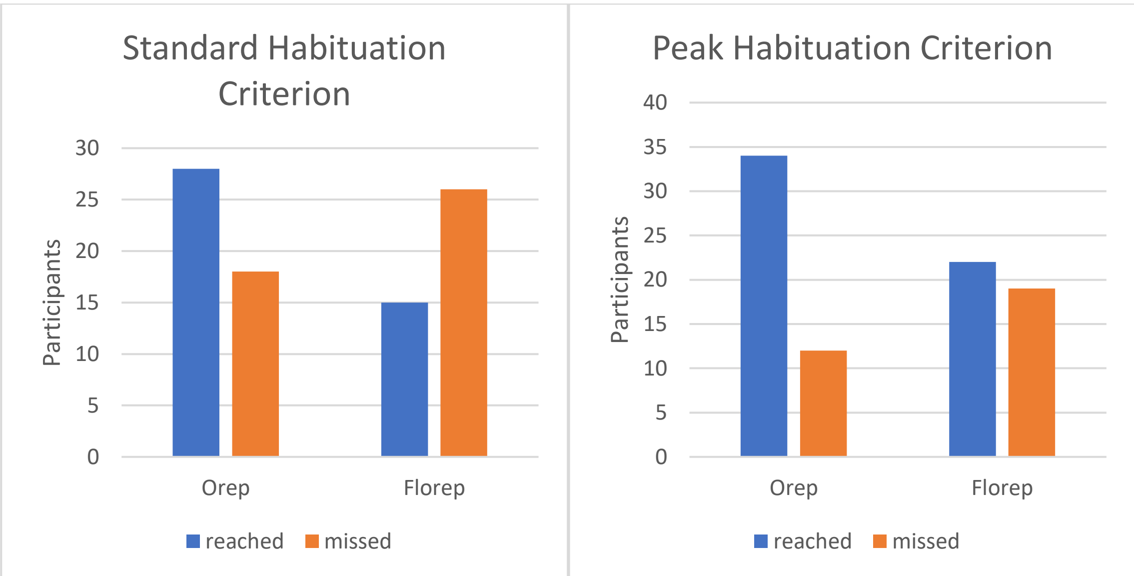
Because the group sizes are so small, making statistical analysis difficult, we decided to calculate a second version of the habituation criterion that is less conservative. This could run the risk of including infants who did not actually habituate (Oakes, 2010). However, using the standard first look criterion, some infants will not reach the criterion at all because the first look is the longest in the habituation process only 60% of the time. Longer looks often occur later in the habituation process, so instead of using the look duration of the first three trials, this criterion uses the look duration of the longest three trials (Colombo & Mitchell, 2009). The *"peak criterion"* was reached when the looking time of the last three trials was less than 50% of the longest three trials. In Study 1, 34 participants met the peak criterion (23 monolinguals and 11 bilinguals, out of a total of n = 46; in 40% of the habituation trials). In Study 2, 22 participants met the peak criterion (15 monolinguals and 7 bilinguals, out of a total of n = 41; in 22% of the trials)

Figure 3 shows a bar chart of the habituation criteria for both studies. Although a few more participants reached the peak criterion, the group sizes between monolinguals and bilinguals are still unequal. This is one of the reasons why we will use Welch's t-test instead of Student's t-test, as it better controls for type 1 errors (see Delacre et al., 2017). We will report the results of both habituation criteria.

Figure 3

Participants Who Met the Habituation Criteria - Standard versus Peak

a) b)



Note. a) In Study 1, out of a total of 46 participants, 28 (61%) met the standard habituation criterion and 18 did not. In Study 2, out of a total of 41 participants, 15 (37%) met the criterion and 26 did not. b) In Study 1 (n = 46), 34 (74%) met the peak habituation criterion and 12 did not. In Study 2 (n = 41), 22 (54%) met the criterion and 19 did not.

Measures for Habituation and Dishabituation

Only behavioral data are analyzed in this thesis. We used JASP software for our statistical analyses. The dependent variable is the infant's gaze duration. The analyses focused on the difference in habituation speed (H1a) for bilingual infants, their habituation strength (H1b), and their dishabituation response (H2). Study 1 and Study 2 were analyzed separately due to different design aspects (in Study 2, the stimulation phase was added, and the familiar stimulus was also shown in the test phase).

We measured *Habituation speed* by total fixation duration, which is the sum of total gaze durations during the habituation phase (Sum Trial 1 to 12/15). It has been shown to be one of the most reliable measures of habituation in visual habituation studies. Shorter fixation durations indicate faster information processing, which means higher encoding efficiency in visual cognition tasks (Pahnke, 2007).

The decrease in looking time from the beginning to the end of the habituation phase was recorded as a measure of the *strength of habituation*. The difference between the first three and the last three trials of the habituation phase is divided by the sum of the respective trials, which allows to measure the strength of habituation relative to the absolute looking time: $\frac{(first\ 3\ trials - last\ 3\ trials)}{(first\ 3\ trials + last\ 3\ trials)}$. This results in a distribution with values between -1 and 1, with positive values indicating a decrease in viewing time and negative values indicating no habituation at all (Pahnke, 2007).

The *dishabituation response* was calculated in two different ways, because of the different design choices for each study. For Study 1, we calculated the strength of dishabituation using a recovery measure. The increase in looking time from the last 2 habituation trials to the 2 test trials was calculated: ($\frac{test\ trials - last\ 2\ trials}{test\ trials + last\ 2\ trials}$). This results in a distribution with values between -1 and 1, with positive values indicating a preference for the novel stimulus and negative values indicating a preference for the familiar stimulus (Pahnke, 2007). For Study 2, we calculated infants' preference for the novel stimulus during the test phase. Therefore we measured the proportion of looking to the novel stimulus, by calculating the mean of each test trials: $\frac{novel\ looking\ time}{(novel\ looking\ time + familiar\ looking\ time)}$ (Kosie et al., 2023).

After calculating the above measures, the mean value for each participant across all completed phases (different monsters) was calculated and used for further calculations. Independent samples t-tests were performed to determine group differences between the monolingual and bilingual infants.

Results

For all subsequent analyses, we report the effect size as Cohen's d for the Welch t-test, as rank-biserial correlation for the Mann-Whitney U-test, and as matched rank-biserial correlation for the Wilcoxon signed-rank test. Cohen's d values around 0.2 are considered small, around 0.5 are considered medium, and 0.8 or greater are considered large. Rank-biserial correlation (r) values range from -1 to 1, with values closer to -1 or 1 indicating a stronger effect. However, research has shown that the effect sizes commonly observed in psychological studies are often smaller than the general guideline values. Therefore, the suggested benchmarks are not as representative of effects in psychology (Anwari et al., 2021). But since we have no other benchmarks, we have to keep this in mind and be careful with generalizations. In addition to reporting effect sizes, we also report the 95% confidence interval for each effect size. The confidence interval represents a range of values within which we are 95% confident that the true effect size lies.

Assumption of Normality

We performed assumption tests to verify that the data met the requirements for parametric testing. In addition, we used distribution plots and Q-Q plots for graphical evaluation (see Figure D1 and Figure D2 in the appendix). Although the Shapiro-Wilk test was not significant for either study (see Table 1), indicating that the data were normally distributed, the graphical evaluation suggested that the data did not meet the requirements for normal distribution. Because the data are not clearly normally distributed, we report nonparametric tests, such as the Mann-Whitney U-Test, in addition to the Welch t test.

Table 1

Shapiro-Wilk Tests for all used Dependent Variables

	n	Shapiro-Wilk	p
Study 1 - OREP			
Habituation Speed	46	0.986	0.856
Habituation Strength	46	0.956	0.081
Dishabituation Response	46	0.970	0.275
Study 2 FLOREP			
Habituation Speed	41	0.983	0.777
Habituation Strength	41	0.951	0.075
Dishabituation Response	41	0.981	0.710

Note. Means for each participant were used for the assumption tests. Significant results suggest a deviation from normality.

Descriptive Analyses

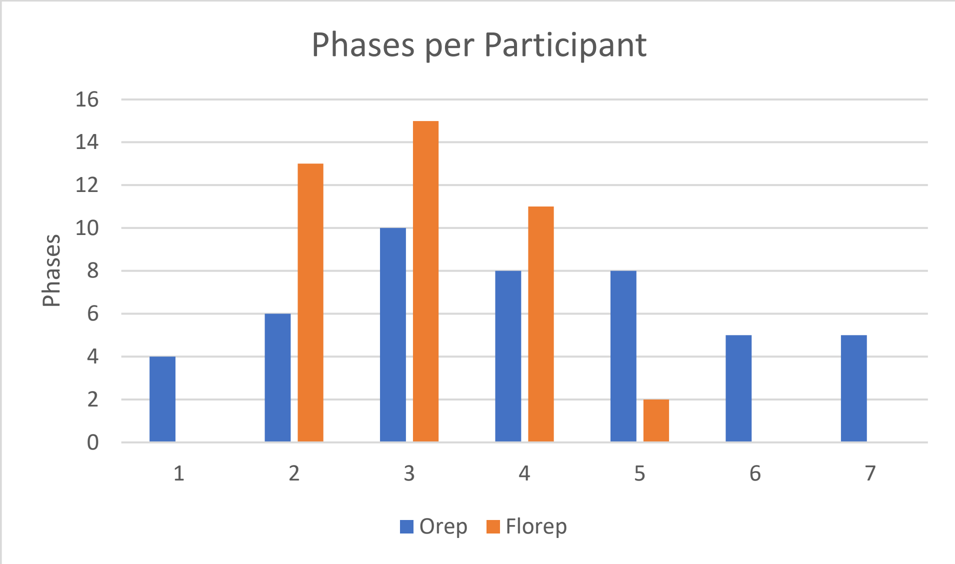
Before analyzing the main hypotheses, we performed descriptive analyses of the data to summarize and describe their main characteristics. We compared the total fixation times during the habituation and test phases of monolinguals and bilinguals. In addition, we calculated the total habituation and dishabituation effects for the whole group, as suggested by Pahnke (2007). We included all participants in the descriptive analyses, regardless of whether they met the habituation criteria. Gaze durations were measured in milliseconds but are reported in seconds for ease of interpretation and clarity.

To determine the habituation effect, we compared the average viewing time of the first three trials with that of the last three trials of the habituation phase. To determine the dishabituation effect for Study 1, we compared the looking time of the last trial of the habituation phase to the mean looking time of the test phase. For Study 2, we compared the mean looking time at the familiar stimulus to the mean looking time at the novel stimulus during the test phase. Paired samples t-tests and their non-parametric equivalent, the Wilcoxon signed-rank test.

Figure 4 shows how many phases each participant watched completely in both studies (for the corresponding table, see Table E1 in the Appendix).

Figure 4

Overview of all Fully Viewed Phases per Participant



Note. In Study 1 (OREP), between 1 and 7 phases were observed. In Study 2 (FLOREP), a minimum of 2 phases could be observed. In addition, phases could be skipped to ensure that the stimulation was viewed in its entirety. In both studies, most children viewed three phases in full (Study 1: n=10, Study 2: n=15), but the distribution looks different due to the design differences.

Study 1 – OREP

During the habituation phase in Study 1, monolingual and bilingual infants did not differ in their mean fixation times. Monolingual infants fixated on the habituation stimulus for an average of 46.52 seconds (*SD* = 10.87, n = 33). Bilingual infants fixated on the habituation stimulus for an average of 40.78 seconds (*SD* = 11.80, n = 13). In the test phase, monolingual infants fixated the novel stimulus for an average of 4.25 seconds (*SD* = 1.28), while bilingual infants fixated the novel stimulus for an average of 3.78 seconds (*SD* = 1.05). Fixation times did not differ significantly between monolingual and bilingual infants.

Table 2 shows the effect sizes and their 95% confidence intervals for all subsequent independent t-tests and Mann-Whitney U-Tests in the descriptive analyses. Figure 5 shows the mean fixation times (in milliseconds) during the habituation and test phases for Study 1 and Study 2.

Table 2

Descriptive Analyses: Independent Samples T-tests

	Test	Statistic	df	p	Effect Size	SE Effect Size	95% CI for Effect Size	
							Lower	Upper
Study 1								
Mean Hab	Welch	1.518	20.502	0.144	0.506	0.333	-0.160	1.160
	Mann-Whitney	268.000		0.199	0.249	0.189	-0.119	0.558
Mean Dishab	Welch	1.277	26.679	0.213	0.400	0.331	-0.254	1.047
	Mann-Whitney	261.000		0.262	0.217	0.189	-0.153	0.533
Study 2								
Mean Hab	Welch	-0.662	29.849	0.513	-0.214	0.318	-0.835	0.412
	Mann-Whitney	162.000		0.275	-0.206	0.183	-0.517	0.153
Mean Dishab	Welch	-0.614	31.983	0.544	-0.197	0.318	-0.818	0.428
	Mann-Whitney	172.000		0.409	-0.157	0.183	-0.479	0.202

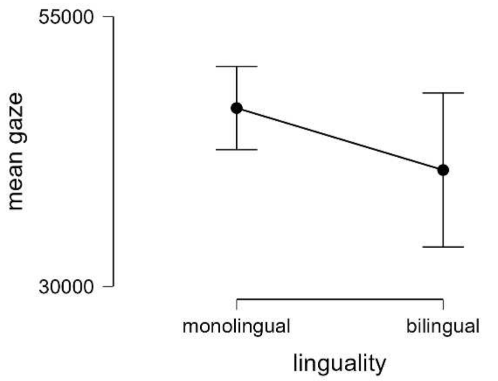
Note. Comparison of mean looking times to habituation and dishabituation stimuli between monolinguals and bilinguals. For the Welch t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

To determine the habituation effect, the mean looking time of the first three trials (*M* = 4.63 s, *SD* = 0.97) was compared to that of the last three trials (*M* = 3.51 s, *SD* = 1.04) of the habituation sequence using a paired Student t-test. Mean looking time was significantly lower at the end of the habituation phase, $t(45) = 8.097$, $\boldsymbol{p < 0.001}$, Cohen's $d = 1.194$, 95% CI [0.810, 1.569]. We also calculated the corresponding nonparametric measure, the Wilcoxon signed-rank test. Mean looking time was also significantly lower at the end of the habituation phase, $z = 5.217$, $\boldsymbol{p < 0.001}$, $r = 0.883$, 95% CI [0.785, 0.938]. This shows that the basic assumptions of the habituation process (a reduction in looking time after repeated presentation of a stimulus) were met for all participants, although not all met the criteria completely. To determine the dishabituation effect for Study 1, the looking time of the last trial of the habituation phase (*M* = 3.34 s, *SD* = 1.23) was compared to the mean looking time of the test phase (*M* = 4.12 s, *SD* = 1.22). The alternative hypothesis was that the means would differ. A paired Student t-test showed that mean looking time was significantly higher in the test phase, $t(45) = -4.115$, $\boldsymbol{p < 0.001}$, Cohen's $d = -0.607$, 95% CI [-0.919, -0.289]. A Wilcoxon signed-rank test also showed a significantly higher mean looking time during the test phase, $z = -3.698$, $\boldsymbol{p < 0.001}$, $r = -0.626$, 95% CI [-0.788, -0.383]. The significant longer looking times to the novel object, also indicate successful habituation, as well as a preference for the novel stimulus. Figure 6 shows the boxplots of the mean looking time during the first three and the last three trials.

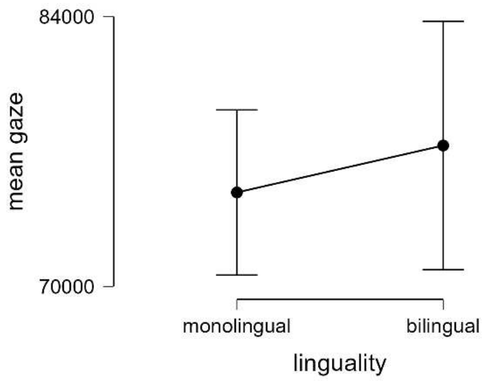
Figure 5

Mean Fixation during Habituation and Dishabituation

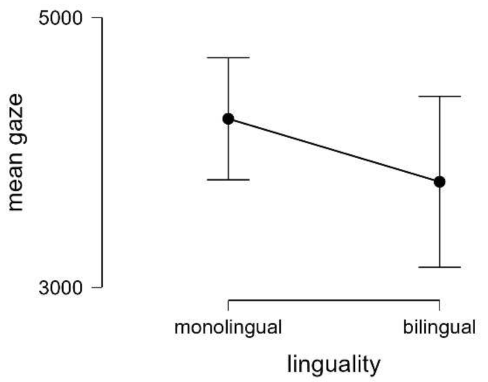
a) Study 1 OREP: Habituation



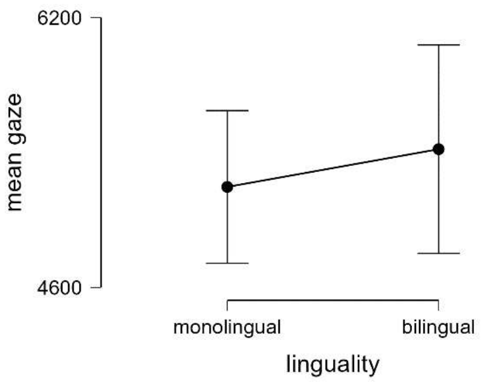
b) Study 2 FLOREP: Habituation



c) Study 1 OREP: Dishabituation



d) Study 2 FLOREP: Dishabituation

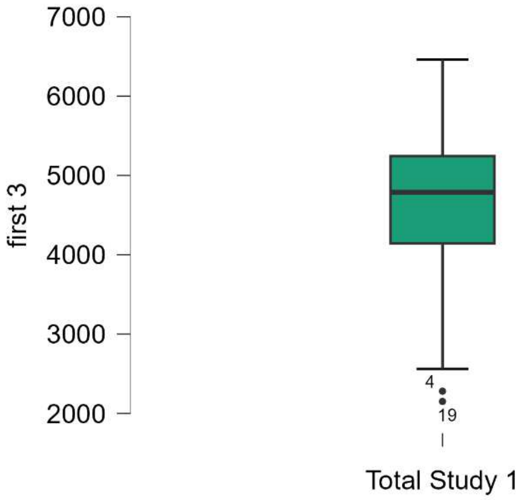


Note. No significant differences in fixation duration (in milliseconds) during habituation and dishabituation were found between monolinguals and bilinguals in either study.

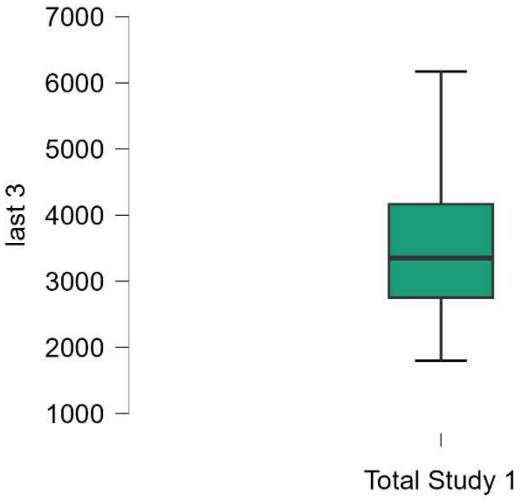
Figure 6

Study 1 OREP – Boxplots: first 3 & last 3

a)



b)

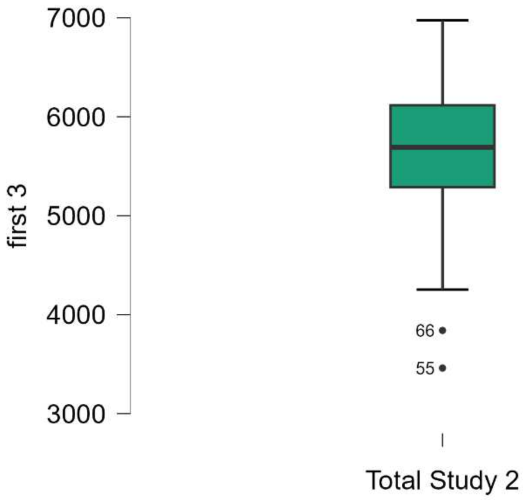


Note. a) shows the gaze in milliseconds for the first 3 trials of Study 1. b) shows the gaze in milliseconds for the last 3 trials of Study 1.

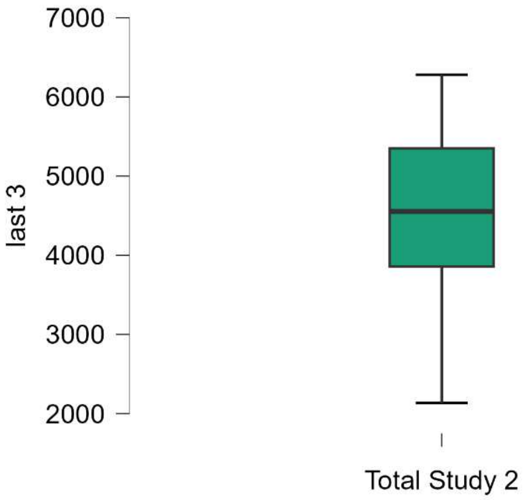
Figure 7

Study 2 FLOREP – Boxplots: first 3 & last 3

a)



b)



Note. a) shows the gaze in milliseconds for the first 3 trials of Study 2. b) shows the gaze in milliseconds for the last 3 trials of Study 2.

Study 2 – FLOREP

Study 2 also found no differences in mean fixation times during the habituation phase between monolingual and bilingual children (see Table 2 again). Monolingual infants fixated on the habituation stimulus for an average of 74.87 seconds (*SD* = 10.14, n = 24), whereas bilingual infants fixated on the habituation stimulus for an average of 77.30 seconds (*SD* = 12.52, n = 17). In the test phase, monolingual infants fixated the novel stimulus for an average of 5.20 seconds (*SD* = 1.07), while bilingual infants fixated the novel stimulus for an average of 5.42 seconds (*SD* = 1.20). Fixation times did not differ significantly between monolingual and bilingual infants.

In the habituation phase of Study 2, there was also a significant decrease in mean looking times from the first three (*M* = 5.61 s, *SD* = 0.82) to the last three trials (*M* = 4.54 s, *SD* = 1.00), *t*(40) = 6.486, ***p* < 0.001**, Cohen's *d* = 1.013, 95% CI [0.631, 1.386]. The Wilcoxon signed-rank test also showed a significant decrease in mean looking times, *z* = 5.217, ***p* < 0.001**, *r* = 0.861, 95% CI [0.785, 0.938]. In Study 2, we also found a significant dishabituation effect, as the mean looking time at the novel stimulus (*M* = 5.29 s, *SD* = 1.12) was significantly higher than the mean looking time at the familiar stimulus (*M* = 4.60 s, *SD* = 1.08) during the test phase. The paired samples t-test revealed *t*(40) = -4.203, ***p* < 0.001**, Cohen's *d* = -0.656, 95% CI [-0.991, -0.315] and the Wilcoxon signed-rank test *z* = -3.596, ***p* < 0.001**, *r* = -0.645, 95% CI [-0.807, -0.392]. Figure 7 shows the boxplots of the mean looking time during the first three and the last three trials.

Figures 8 and 9 show the course of the average gaze durations during the habituation and test phases for each study. While the gaze durations in Study 1 show a steady decrease, the gaze durations in Study 2 seem to remain constant with a slight increase at the end. This could mean that habituation did not work in Study 2. This is also consistent with the fact that so few met the habituation criteria compared to Study 1. Perhaps the infants needed more trials to habituate, or other factors, such as the stimulation phase, may have altered the process. Although the statistical calculations above show significant habituation and dishabituation effects for the entire group, this is not evident in the graph. This difference between the statistical results and the graph can be explained by the different calculation methods. The statistical analyses are based on the means of the first three and the last three trials of each phase that each child went through. The graph, on the other hand, shows the mean of each trial for all participants. Therefore, the significant results of the calculations are not directly reflected in the graph. In theory, this suggests that habituation did occur in Study 2, but the effect was primarily observed at the group level. When analyzed individually, each child's looking time did not decrease sufficiently to meet the 50% criterion, indicating that the habituation effect was not as strong or consistent across individuals.

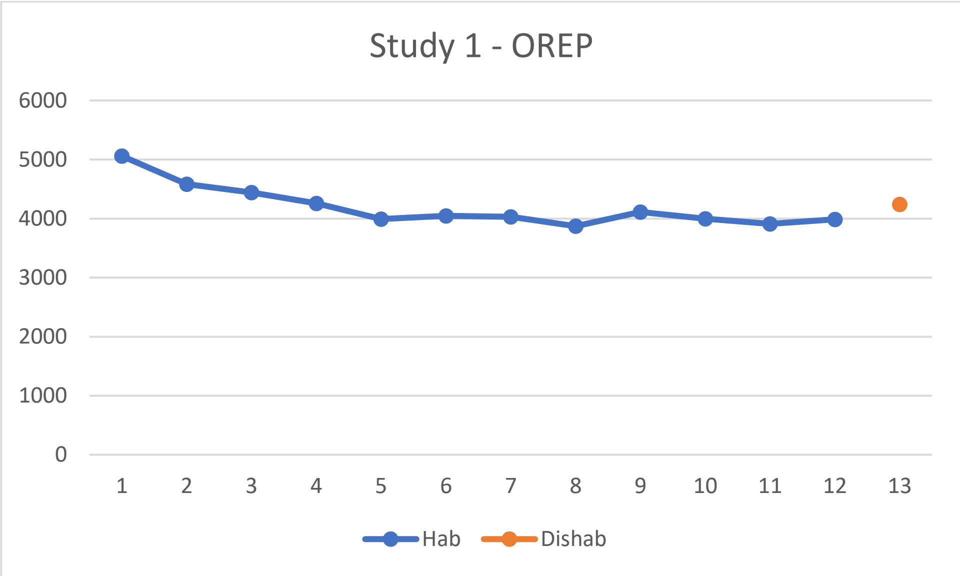
You can find a complete Table with descriptive information on the dependent variables for both studies, including the two habituation criteria, in the Appendix (see Table E2, Table E3, and Table E4).

Summary

In summary, both studies showed successful habituation and dishabituation effects for the whole group. However, we found no significant differences between monolingual and bilingual infants with respect to fixation durations during the habituation and test phases. While the significance of the habituation effect is also shown in the line graph of Figure 8, Figure 9 does not give a clear picture of successful habituation.

Figure 8

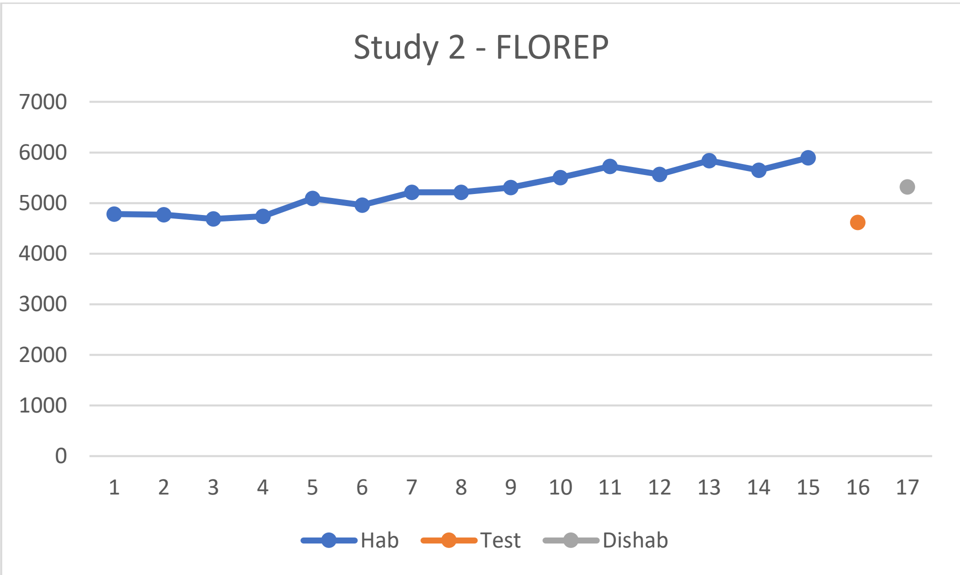
Course of Average Looking Time in Study 1



Note. Mean gaze durations (in milliseconds) during the habituation and test phases for Study 1 (N = 46, maximum 12 trials). In the test phase, the mean of the two trials with the novel stimulus ("dishab") was calculated.

Figure 9

Course of Average Looking Time in Study 2



Note. Mean gaze durations (in milliseconds) during the habituation and test phases for Study 2 (N = 41, maximum 15 trials). In the test phase, the mean of the two trials with the familiar ("test") and the two trials with the novel stimulus ("dishab") was calculated.

Main Analyses

The primary aim of this study was to investigate whether bilingualism affects habituation and dishabituation processes. We tested three hypotheses in each of the two studies (OREP and FLOREP) to assess differences in habituation speed, habituation strength, and dishabituation response.

Study 1 – OREP

First Research Question (Study 1). Our first hypothesis (H1a) proposed that *bilingual infants habituate faster than monolingual infants.* To test this, we conducted a one-tailed independent samples t-test, with *habituation speed* as the dependent variable and *bilingualism* was the independent variable. Table 3 and Table 4 show the effect sizes and their 95% confidence intervals for all subsequent independent t-tests and Mann-Whitney U-tests in the main analyses for each habituation criterion.

First, we calculated the analyses using the data for which the standard habituation criterion was reached. The Welch t-test compared the mean scores of monolinguals (M = 35.65 s, SD = 9.17, n = 18) and bilinguals (M = 27.61 s, SD = 13.66, n = 10). The results showed that there was no significant difference between the means of the two groups. Interestingly, the Mann-Whitney U-test showed a significant difference between monolingual and bilingual participants, U = 134.000, $\boldsymbol{p}$ **= 0.018**, r = 0.489, 95% CI [0.152, ∞].

A post hoc power analysis was performed using GPower 3.1.9.2 (Faul et al., 2007) to assess whether the study had adequate power to detect an effect. With an effect size of d = 0.691, an α error of 0.05, and sample sizes of 18 and 10 participants per group, the analysis indicated a power of 0.52. A sample size of 27 participants per group would have been required to achieve a power of 0.80. Normally, an α-error correction would be applied because the confirmation of only one of the sub-hypotheses is answered in terms of the global alternative hypothesis (e.g., "bilingualism influences habituation and dishabituation processes"; see intersection union testing; Dmitrienko & D’Agostino Sr, 2013). However, the smaller the α-error level, the lower the power of the test and the less meaningful the study would be. Therefore, we decided not to make a correction. Table 5 summarizes the post hoc power analyses for all subsequent analyses from Study 1 and Study 2.

We repeated the analysis using data that met the peak habituation criterion. The Welch t-test showed no significant difference between monolinguals (*M* = 36.23, *SD* = 9.64, n = 23) and bilinguals (*M* = 31.74, *SD* = 11.29, n = 11). The Mann-Whitney U-Test also showed no significant difference.

Although the Mann-Whitney U-test using the standard habituation criterion was significant, the first alternative hypothesis (H1a) must be rejected because the t-test shows no significant difference, and the power is too low for sufficient significance. This suggests that there may be subtle differences in habituation processes that were not captured by the t-test and require further investigation with a larger sample size.

Second Research Question (Study 1). Our second hypothesis (H1b), that *bilingual infants habituate stronger than monolingual infants*, was also tested using a one-tailed independent samples t-test. The dependent variable was the *habituation strength* and the independent variable was *bilingualism*.

First, using data that met the standard habituation criterion, the Welch t-test compared the mean scores of monolinguals (*M* = 0.441, *SD* = 0.089, n = 18) and bilinguals (*M* = 0.446, *SD* = 0.082, n = 10). We did not find a significant difference between the two groups. Similarly, the Mann-Whitney U-Test showed no significant difference. Negative values indicate that the effect is in the opposite direction as proposed. In this case, the monolingual infants showed stronger habituation. Repeating the analysis with the peak habituation criterion data, the Welch t-test showed no significant difference between monolinguals (*M* = 0.368, *SD* = 0.112, n = 23) and bilinguals (*M* = 0.367, *SD* = 0.099, n = 11). The Mann-Whitney U-Test also showed no significant difference.

The second alternative hypothesis (H1b) must be rejected due to no significant difference and extremely low power (see also Tables 3, Table 4, Table 5).

Third Research Question (Study 1). Our third hypothesis (H2), that *bilingual infants show a stronger dishabituation response than monolingual infants*, was again tested using a one-tailed independent samples t-test, with *dishabituation strength* as the dependent variable and *bilingualism* as the independent variable.

Using the standard habituation criterion data, the Welch t-test compared monolinguals (*M* = 0.412, *SD* = 0.155, n = 18) and bilinguals (*M* = 0.296, *SD* = 0.169, n = 10). There was no significant difference between the means. The Mann-Whitney U-Test also showed no significant difference. For the peak habituation criterion data, the Welch t-test showed no significant difference between monolinguals (*M* = 0.293, *SD* = 0.229, n = 23) and bilinguals (*M* = 0.272, *SD* = 0.185, n = 11). The Mann-Whitney U-Test also showed no significant difference.

The third alternative hypothesis (H2) must be rejected due to no significant difference and insufficient power (see also Tables 3, Table 4, Table 5).

Study 2 – FLOREP

We used the same procedure for Study 2. First, we calculated t-tests and Mann-Whitney U-tests with the standard habituation criterion and then again with the peak habituation criterion. Post hoc power analyses were also performed.

First Research Question (Study 2). The Welch t-test comparing monolinguals (*M* = 65.76 s, *SD* = 10.91, n = 10) and bilinguals (*M* = 55.85 s, *SD* = 20.24, n = 5) showed no significant difference. Similarly, the Mann-Whitney U-Test showed no significant difference. Repeated with the peak habituation criterion data, the Welch t-test showed no significant difference between monolinguals (*M* = 62.77 s, *SD* = 10.30, *n* = 15) and bilinguals (*M* = 55.52 s, *SD* = 13.36, *n* = 7). The Mann-Whitney U-test also showed no significant difference.

The first alternative hypothesis (H1a) must be rejected because there is no significant difference and the power is too low for sufficient significance (see also Tables 3, Table 4, Table 5).

Second Research Question (Study 2). Using the standard habituation criterion data, the Welch t-test compared monolinguals (*M* = 0.445, *SD* = 0.109, *n* = 10) and bilinguals (*M* = 0.505, *SD* = 0.168, *n* = 5) and showed no significant difference. The Mann-Whitney U-test also showed no significant difference. Repeated with the peak habituation criterion data, the Welch t-test showed no significant difference between monolinguals (*M* = 0.343, *SD* = 0.163, *n* = 15) and bilinguals (*M* = 0.433, *SD* = 0.210, *n* = 6). The Mann-Whitney U-test showed no significant difference.

The negative values in both subsamples indicate that the effect is in the opposite direction as proposed. This was the same as in Study 1. The second alternative hypothesis (H1b) must be rejected due to no significant difference and low power (see also Tables 3, Table 4, Table 5).

Third Research Question (Study 2). Using the standard habituation criterion data, the Welch t-test compared monolinguals (*M* = 0.567, *SD* = 0.138, n = 10) and bilinguals (*M* = 0.556, *SD* = 0.066, *n* = 5) and showed no significant difference. The Mann-Whitney U-Test also showed no significant difference. Repeated with the peak habituation criterion data, the Welch t-test showed no significant difference between monolinguals (*M* = 0.513, *SD* = 0.162, n = 15) and bilinguals (*M* = 0.563, *SD* = 0.059, n = 7). The Mann-Whitney U-Test also showed no significant difference.

The third alternative hypothesis (H2) has to be rejected due to no significant difference and insufficient power (see also Tables 3, Table 4, Table 5).

Summary

In summary, none of the hypotheses could be accepted in either study. Except for one significant result in a Mann-Whitney U-test for H1a in Study 1, all tests showed no significant differences between the groups of monolingual and bilingual infants with respect to habituation and dishabituation. The lack of significant findings may be due to low statistical power and small sample sizes, which limited the ability to detect potential effects. Future research should aim to address these limitations by increasing sample sizes and considering alternative methods of assessing habituation and dishabituation.

Table 3

Main Analyses: Independent Samples T-tests using the Standard Habituation Criterion

	Test	Statistic	df	p	Effect Size	SE Effect Size	95% CI for Effect Size	
							Lower	Upper
Study 1 - OREP								
Habituation Speed (H1a)	Welch	1.665	13.613	0.059	0.691	0.411	-0.004	∞
	Mann-Whitney	134.000		**0.018***	0.489	0.228	0.152	∞
Habituation Strength (H1b)	Welch	-0.154	20.033	0.440	-0.060	0.395	-∞	0.590
	Mann-Whitney	86.000		0.435	-0.044	0.228	-∞	0.325
Dishabituation Response (H2)	Welch	1.799	17.379	0.955	0.718	0.412	-∞	1.387
	Mann-Whitney	121.000		0.934	0.344	0.228	-∞	0.629
Study 2 - FLOREP								
Habituation Speed (H1a)	Welch	1.023	5.198	0.176	0.610	0.564	-0.367	∞
	Mann-Whitney	29.000		0.339	0.160	0.315	-0.359	∞
Habituation Strength (H1b)	Welch	-0.729	5.742	0.247	-0.426	0.556	-∞	0.515
	Mann-Whitney	19.000		0.257	-0.240	0.315	-∞	0.284
Dishabituation Response (H2)	Welch	0.216	12.995	0.584	0.106	0.548	-∞	1.005
	Mann-Whitney	25.000		0.523	0.000	0.315	-∞	0.491

Note. H1a, specifies that group *monolingual* is greater than group *bilingual*. H1b and H2, specifies that group *monolingual* is less than group *bilingual*. For the Welch t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Table 4

Main Analyses: Independent Samples T-tests using the Peak Habituation Criterion

	Test	Statistic	df	p	Effect Size	SE Effect Size	95% CI for Effect Size Lower	95% CI for Effect Size Upper
Study 1 - OREP								
Habituation Speed (H1a)	Welch	1.135	17.229	0.136	0.428	0.372	-0.193	∞
	Mann-Whitney	164.000		0.088	0.296	0.211	-0.048	∞
Habituation Strength (H1b)	Welch	0.036	22.148	0.514	0.013	0.367	-∞	0.616
	Mann-Whitney	128.000		0.529	0.012	0.211	-∞	0.350
Dishabituation Response (H2)	Welch	0.285	24.176	0.611	0.101	0.367	-∞	0.703
	Mann-Whitney	136.000		0.644	0.075	0.211	-∞	0.404
Study 2 - FLOREP								
Habituation Speed (H1a)	Welch	1.272	9.480	0.117	0.609	0.471	-0.193	∞
	Mann-Whitney	67.000		0.166	0.276	0.264	-0.160	∞
Habituation Strength (H1b)	Welch	-0.936	7.533	0.189	-0.476	0.491	-∞	0.358
	Mann-Whitney	34.000		0.212	-0.244	0.278	-∞	0.217
Dishabituation Response (H2)	Welch	-1.039	19.462	0.156	-0.404	0.464	-∞	0.361
	Mann-Whitney	44.000		0.291	-0.162	0.264	-∞	0.274

Note. H1a, specifies that group *monolingual* is greater than group *bilingual*. H1b and H2, specifies that group *monolingual* is less than group *bilingual*. For the Welch t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Table 5

Post Hoc Power Analyses

Hypothesis	Criterion	t-test	Mann-Whitney U-test	Effect Size Cohen's *d*	alpha	n1	n2	actual Power	required N per group
Study 1									
H1a	standard	not significant	**significant: p = 0.018**	0.691	0.05	18	10	0.524	27
H1a	peak	not significant	not significant	0.428	0.05	23	11	0.308	69
H1b	standard	not significant	not significant	-0.060	0.05	18	10	0.067	3436
H1b	peak	not significant	not significant	0.013	0.05	23	11	0.054	73168
H2	standard	not significant	not significant	0.718	0.05	18	10	0.551	25
H2	peak	not significant	not significant	0.101	0.05	23	11	0.085	1213
Study 2									
H1a	standard	not significant	not significant	0.610	0.05	10	5	0.278	34
H1a	peak	not significant	not significant	0.609	0.05	15	7	0.360	35
H1b	standard	not significant	not significant	-0.426	0.05	10	5	0.182	69
H1b	peak	not significant	not significant	-0.476	0.05	15	7	0.261	56
H2	standard	not significant	not significant	0.106	0.05	10	5	0.072	1102
H2	peak	not significant	not significant	-0.404	0.05	15	7	0.214	77

Note. Effect sizes (Cohen's d) are from Welch t-tests. An α error of 0.05 was used for all power analyses. Required N per group indicates the sample size required to achieve a power of 0.80.

Exploratory Analyses

Since the original idea was to analyze the two studies as a joint sample, but we decided against it, we wanted to take another exploratory look at the extent to which the two studies differ. While no significant differences in total fixation during habituation phase were found between monolinguals and bilinguals in either study, we found that there were significant differences in mean fixation times during habituation phase between the studies (Welch t-test: $t(84.192) = -12.875$, $\boldsymbol{p < 0.001}$; Mann-Whitney U-test: $U = 52.000$, $\boldsymbol{p < 0.001}$). In Study 1, infants fixated the stimulus for an average of 44.89 seconds ($SD = 11.31$), whereas in Study 2, infants fixated the stimulus for an average of 75.88 seconds ($SD = 11.11$). This confirms the decision to analyze the studies separately. In addition, we found significant differences in the mean fixation on the dishabituation stimulus in the test phase studies (Welch t-test: $t(84.939) = -4.666$, $\boldsymbol{p < 0.001}$; Mann-Whitney U-test: $U = 460.500$, $\boldsymbol{p < 0.001}$). Infants in Study 1 fixated the stimulus for an average of 4.12 seconds ($SD = 1.22$), whereas infants in Study 2 fixated the stimulus for an average of 5.29 seconds ($SD = 1.12$).

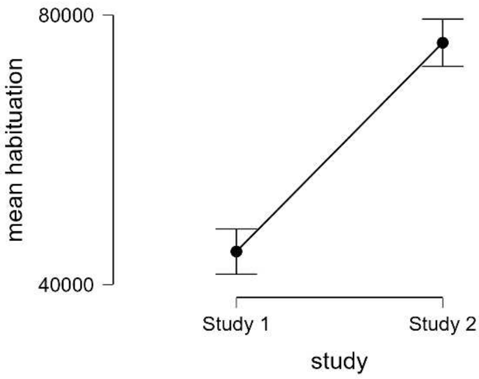
We also found that the first 3 means in Study 1 were significantly different from the first 3 means in Study 2 (Welch t-test: $t(84.685) = -5.132$, $\boldsymbol{p < 0.001}$; Mann-Whitney U-test: $U = 394.500$, $\boldsymbol{p < 0.001}$). In Study 1, infants fixated the first three trials for a mean of 4.63 seconds ($SD = 0.97$), whereas in Study 2, infants fixated the first three trials for a mean of 5.61 seconds ($SD = 0.82$).

In summary, infants in Study 2 showed significantly longer looking times for both the familiar and the novel stimulus throughout the experiment than infants in Study 1 (see Figure 10).

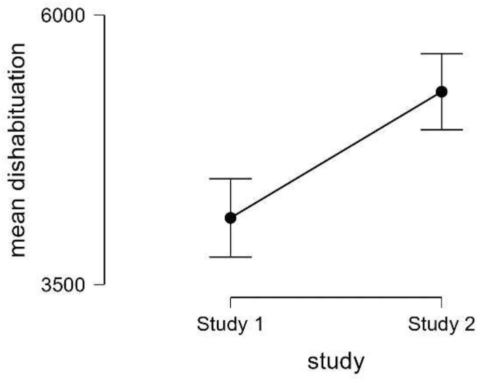
Figure 10

Mean Fixation Times Comparing Study 1 and Study 2

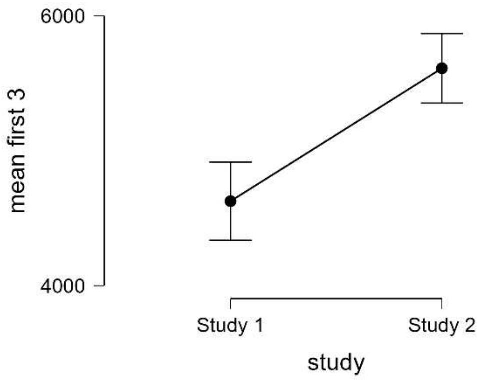
a) Habituation



b) Dishabituation



c) First 3 Trials



Note. In Study 2, the looking times were significantly higher than in Study1.

Habituation Criteria

Since far fewer infants in Study 2 reached the habituation criterion, we wanted to compare this between studies. Indeed, infants in Study 2 were significantly less likely to reach the standard habituation criterion (Welch t-test: $t(71.120) = 3.527$, $\boldsymbol{p < 0.001}$; Mann-Whitney U-test: $U = 1261.500$, $\boldsymbol{p = 0.003}$) and the peak criterion (Welch t-test: $t(67.222) = 3.966$, $\boldsymbol{p < 0.001}$; Mann-Whitney U-test: $U = 1299.000$, $\boldsymbol{p = 0.001}$). The mean for the standard criterion was 1.06 ($SD = 1.08$) in Study 1 and 0.42 ($SD = 0.59$) in Study 2. For the peak criterion, the means were 1.59 ($SD = 1.41$) in Study 1 and 0.66 ($SD = 0.69$) in Study 2.

A complete table of all exploratory tests for both studies, including the effect sizes and their 95% confidence intervals, can be found in the Appendix (see Table E5).

Correlations

Table 6 shows the correlation of Study 1 and Study 2 with the variables sex, age, linguality (monolingual/bilingual), phases the infant fully watched, study affiliation (Study 1/Study 2), mean habituation, mean dishabituation and mean first three trials.

Age was found to correlate with several other variables: A moderate positive association was found between age and gender (r = 0.223, $\boldsymbol{p = 0.038}$), indicating that “older” infants were more likely to be female. Another moderate positive correlation (r = 0.248, $\boldsymbol{p = 0.021}$) was found between age and the study in which the infant was enrolled, suggesting that "older" infants were more likely to be in Study 2. A final small correlation between age and performance (r = 0.214, $\boldsymbol{p = 0.046}$) was found in the first three trials, suggesting that “older” infants showed more interest in the initial trials.

Study affiliation was also found to correlate with several other variables: A strong positive correlation (r = 0.813, $\boldsymbol{p < 0.001}$) was found between study affiliation and mean habituation scores, indicating that infants in Study 2 had longer mean looking times during habituation. A similar relationship was found between study affiliation and the first 3 trials (r = 0.483, $\boldsymbol{p < 0.001}$), meaning that infants in Study 2 also looked longer during the first 3 trials, which makes sense because the first 3 trials were part of the habituation phase.

Mean gaze durations during habituation and dishabituation were also positively correlated (r = 0.683, $\boldsymbol{p < 0.001}$), indicating that infants who looked longer during habituation also looked longer during dishabituation, which is also consistent with the habituation-dishabituation paradigm, where longer gaze durations during dishabituation were expected.

Mean gaze during the first 3 trials was also correlated positively with mean gaze during habituation (r = 0.777, $\boldsymbol{p < 0.001}$) and mean gaze during dishabituation (r = 0.678, $\boldsymbol{p < 0.001}$), again suggesting that longer gaze durations in the beginning lead to longer gaze durations in general.

We also found negative correlations between some of the variables: A negative relationship between the phases the participants fully watched and study affiliation (r = -0.311, $\boldsymbol{p = 0.003}$) was found, meaning that infants in Study 2 watched less phases, this pattern was also shown in Figure 4. Interestingly, the more phases an infant watched, the less time they spent looking during the habituation phase (*r* = -0.218, ***p* = 0.042**).

Table E6 shows only the correlations of Study 1 with the variables *sex*, *age, linguality, phases* the infant observed completely, *study affiliation, habituation speed, habituation strength, dishabituation response, mean first three trials, mean last three trials*, and *mean dishabituation*. Table E7 shows only the correlations of Study 2 with the same variables as Study 1 and additionally with *mean stimulation* and *mean Test*. These two tables can be found in the Appendix.

Summary

The aim of the research was to investigate whether bilingualism influences habituation and dishabituation processes in infants. We used parametric (Welch's t-test) and non-parametric (Mann-Whitney U-test) tests to compare differences between groups, as the normal distribution in the data was not fully given. However, the results of the main analyses did not show significant differences between monolingual and bilingual infants, likely due to small sample sizes and low statistical power. The descriptive analyses focused on the comparison of fixation times between monolingual and bilingual children during the habituation and test phases and did not find a significant difference. However, significant habituation and dishabituation effects were found for the whole group.

The exploratory analysis revealed significant differences between the two studies (OREP and FLOREP), justifying the decision to analyze them separately. Infants in Study 2 had longer mean fixation durations during the habituation phase and were less likely to reach the habituation criterion compared to infants in Study 1. In addition, correlations showed that older infants were more likely to be female and to be enrolled in Study 2. There were strong positive correlations between study affiliation and mean gaze durations during habituation and the first three trials, indicating longer gaze durations in Study 2. Negative correlations suggested that infants in Study 2 looked at fewer phases and that longer overall looking was associated with shorter habituation times.

Table 6

Pearson's Correlations of Study 1 OREP and Study 2 FLOREP

Variable		sex	age	linguality	phases	study	hab	dishab	first 3
1. sex	Pearson's r	—							
	p-value	—							
2. age	Pearson's r	**0.223**	—						
	p-value	**0.038** *	—						
3. linguality	Pearson's r	-0.119	-0.013	—					
	p-value	0.274	0.902	—					
4. phases	Pearson's r	0.071	-0.108	0.110	—				
	p-value	0.512	0.322	0.310	—				
5. study	Pearson's r	0.142	**0.248**	0.139	**-0.311**	—			
	p-value	0.190	**0.021** *	0.200	**0.003** **	—			
6. hab	Pearson's r	0.134	0.153	0.075	**-0.218**	**0.813**	—		
	p-value	0.216	0.156	0.487	**0.042** *	**< .001** ***	—		
7. dishab	Pearson's r	0.147	0.138	0.023	0.014	**0.450**	**0.683**	—	
	p-value	0.175	0.203	0.835	0.895	**< .001** ***	**< .001** ***	—	
8. first 3 trials	Pearson's r	0.055	**0.214**	-0.086	-0.021	**0.483**	**0.777**	**0.678**	—
	p-value	0.611	**0.046** *	0.427	0.850	**< .001** ***	**< .001** ***	**< .001** ***	—

Note. * p < .05, ** p < .01, *** p < .001

Discussion

The aim of this master's thesis was to investigate whether bilingualism affects habituation and dishabituation processes in infants. A conceptual replication of Singh et al. (2015) was the goal. Specifically, we hypothesized that bilingual infants would *habituate faster* (H1a) and *more strongly* (H1b) than monolingual infants. In addition, we predicted that bilingual infants would also show a *stronger dishabituation response* (H2). We conducted two studies with 7- and 8-month-old bilingual and monolingual infants from Vienna using a visual habituation design. For the full sample, the habituation and dishabituation effects were significant for both studies, indicating that the infants successfully memorized the habituation stimulus and were able to discriminate between the novel and the familiar stimulus. However, we found no significant differences in fixation times between bilingual and monolingual infants during habituation and dishabituation. In both studies, the results showed that there were no significant differences in habituation speed, strength, or novelty preference between bilingual and monolingual infants. Interestingly, only in Study 1 did the Mann-Whitney U-test of the main analyses indicate that bilingual infants habituated significantly faster than monolinguals. Thus, none of our hypotheses was fully supported.

To the best of our knowledge, our study is the first to attempt to replicate the findings of Singh et al. (2015). In addition, our study contributes to the interest in testing the reliability and generalizability of the proposed early bilingual advantage. However, our results did not provide support for this early bilingual advantage. This lack of replication is consistent with other research, which has also produced mixed results about whether bilingualism offers cognitive benefits. More rigorous and large-scale replications are needed to better understand the reliability of any potential bilingual advantage. The mixed evidence in the literature - with some studies reporting benefits and others not - suggests that the bilingual advantage may be influenced by several factors, including the specific methods used, the age of the participants, and the languages involved. While bilingual environments may expose infants to more diverse stimuli, this does not necessarily result in consistently measurable cognitive advantages in habituation and novelty preference.

However, it is important to note the limitations of this study to avoid falsely rejecting the stated hypotheses. Our research adds to the ongoing debate by providing a valuable attempt to replicate previous findings. However, the lack of support for early bilingual advantages in our study suggests that the evidence for such advantages remains inconclusive. The variability in research findings suggests that more studies are needed to understand the conditions under which bilingualism might provide cognitive benefits.

Limitations

Our study has several limitations that may affect the interpretation of the results. The main limitation of our study is the sample size and power, which is directly related to the sample size. The power of the statistical tests was often low, indicating that larger sample sizes may be needed to detect potential effects more reliably. Of the 50 infants originally recruited for Study 1 and 62 infants for Study 2, approximately 40 infants remained for each study. However, the standard habituation criterion was met in only 28 (18 monolinguals and 10 bilinguals, Study 1) and 15 (10 monolinguals and 5 bilinguals, Study 2) participants, respectively. The use of a second habituation criterion to obtain larger sample sizes was not successful. In the end, we had 34 (23 monolinguals and 11 bilinguals, Study 1) and 22 (15 monolinguals and 7 bilinguals, Study 2) participants who met the peak habituation criterion. Because we had to split the participants into two groups (monolingual versus bilingual), the sample size became even smaller, making it difficult to obtain reliable and valid results in the analyses. However, low power due to small sample sizes is common in developmental research, particularly in looking time studies. This is because data collection with young children is quite difficult, expensive, and time consuming (Oakes, 2017; Bergmann et al., 2019).

Oakes (2017) suggests several ways to address this issue. One obvious solution to improve power is to increase the sample size to 20 to 32 participants per cell. However, as noted above, this is not always possible. A compromise may be to use careful sequential hypothesis testing, where a target sample size is set but a test is scheduled at an intermediate point in data collection. Another solution may be to use Bayes factors instead of p-values, which indicate the relative probability of the observed results for the null hypothesis.

Habituation Criteria

We have already mentioned that it was a problem that in many cases the infants did not meet the habituation criterion (see Figure 3). Specifically, in Study 2, only 15 of 41 infants achieved standard criterion, or about 37% of the sample. Although we were able to make some progress with the peak criterion, only 54% of the sample (22 of 41 infants) successfully habituated to this criterion. As a reminder, in Study 1, 28 infants (about 61%) met the standard criterion, and 34 infants (about 74%) met the peak criterion out of 46 participants.

Why did significantly fewer infants in Study 2 meet the habituation criterion? The reliability of habituation criteria is often limited, as discussed in previous studies (Lavoie & Desrochers, 2002; Kucharský et al., 2022), for example, the criteria often lead to high error rates, thus affecting the accuracy and consistency of research results. Ashmead and Davis (1996) ran simulations of habituation criteria and obtained mean test-retest correlations close to zero. This indicates that the habituation criteria tested have very low test-retest reliability. Furthermore, Dannemiller (1984) found that Type-I-errors (false positives, i.e., randomly reporting habituation when none has occurred) increase with increasing noise in the data. However, low noise in the data can lead to higher Type-II-errors (false negatives, i.e., failing to detect habituation when it has occurred). Even when a person is fully habituated, they may still respond to the stimulus to some degree, and this response can vary widely from person to person. The level of attention that remains after someone has habituated is called “residual attention”. If this residual attention is higher than the threshold set by the habituation criteria, then the criteria will have difficulty detecting habituation unless there is a lot of noise in the data. This means that the criterion may not be sensitive enough to detect subtle changes in response that indicate that the person has habituated to the stimulus (Kucharský et al., 2022). This means that infants will randomly meet or fail the habituation criterion, which is problematic because only infants who meet the criterion can be included in the final analyses (Oakes, 2010). When simulating the characteristics of a classical habituation process using the 3-3-50% criterion, Kucharský et al. (2022) found that the criterion was relatively strict and excluded many participants who actually habituated (54 out of 89). This may also be the case in our study, as the analyses in the whole group showed successful habituation effects using the means of all infants, regardless of whether they reached the criterion.

Since the habituation criteria are so arbitrary and random, it would make sense to rely less on them in the future and instead consider the entire habituation process more comprehensively. Perhaps future studies should consider thinking of habituation as a process in which a habituating person shows a decrease in attention over time. Instead of filtering out participants who fail to reach one of the habituation criteria. It may not be possible to achieve the desired decrease in attention within a given time to reach a criterion (Kucharský et al., 2022). We will discuss this topic in more detail later.

Another possibility for why fewer infants in Study 2 met the habituation criterion could be that the infants were already more fatigued due to the additional stimulation just before the habituation phase. The infants saw five additional flickering trials of three seconds each, which the infants in Study 1 did not see. Therefore, infants in Study 2 may already be exhausted. However, this is contradicted by the fact that the infants in Study 2 generally had significantly longer looking times than in Study 1 (see Figure 10). Thus, the stimulation phase may have actually increased infants' attention, leading them to look at the stimuli longer during the habituation and test phase as well. These are just speculative thoughts, and we will have to wait for the results of the EEG analysis from the PhD study to get a clearer picture of the possible effects of the stimulation phase.

Effects Sizes

We reported our effect sizes as Cohen's *d* and as rank-bias correlation. Although our effect sizes were sometimes high, with medium to large effects and a Cohen's *d* of 0.6 or 0.7 in the main analyses, Oakes (2017) cautions that with small sample sizes, effect sizes often overestimate the true effect. Another option would have been to use Hedges' *g* instead, which is a less biased version of Cohen's *d* and therefore more appropriate for small sample sizes (Bergmann et al., 2019). We had some effect sizes with negative values, which usually means that the effect is in the opposite direction. However, Bergmann et al. (2019) suggest that so-called "non-dominant preferences" - i.e., preferences that are not typically reported (e.g., a familiarity preference) - may represent a "sign error”. Extremely low statistical power increases the likelihood of such sign errors, where the statistical effect size is in the opposite direction of the true underlying effect. Therefore, we recommend further research with larger sample sizes and more robust statistical methods to ensure the accuracy and reliability of our findings.

Classification of Bilingualism

In our study, we determined bilingual status based on whether parents reported speaking more than one language to their child in a demographic questionnaire. However, we did not collect information on the proportion of time parents spent speaking each language to their child, nor did we use a specific language exposure questionnaire. For example, other studies use instruments such as the Language Exposure Questionnaire (LEQ, e.g., D'Souza et al., 2020), which quantifies the amount of exposure to each language in a typical week, or the Language Background Questionnaire (e.g., Singh et al., 2023), which details language exposure from birth to the time of the study. Many studies define a child as bilingual if he or she receives at least 25 to 75 percent exposure to each language (Pearson et al., 1997). Without this detailed information, we cannot rule out the possibility that some infants in our bilingual sample may not meet the 25 to 75 percent exposure criteria. This limitation could potentially affect the results and interpretations of our analyses.

Other

It should also be noted that we had many more dropouts in Study 2 (21 out of 62) than in Study 1 (4 out of 50) because one of the inclusion criteria was that the children had to see at least two complete phases of one of the frequencies (6 Hz versus 4 Hz) during stimulation. This inclusion criterion was necessary for the PhD dissertation in which this master's thesis takes place. Theoretically, we could have kept the data sets with only one phase for the master's thesis analyses. Although not all of these dropouts were exclusively due to this inclusion criterion, at least 11 participants were excluded because they completed only one full phase of the experiment. We cannot completely rule out the possibility that the increased dropout rate in Study 2 may have affected the generalizability of our findings. However, the inclusion criterion was essential to ensure that the data collected would meet the requirements of the dissertation and was designed to ensure that participants had sufficient exposure to the experimental conditions to provide meaningful data. We decided to exclude the data sets of participants who completed only one phase also for the master's thesis. We made this decision to avoid the potential bias that could result from some children being exposed to theta only while others were exposed to alpha only. This was done to ensure the comparability of the data between the groups and to prevent the results of the analysis from being influenced by unequal conditions.

Summary

Our study has several limitations that may affect the interpretation of the results. The small sample size and unequal group sizes between bilingual and monolingual infants may limit statistical power and introduce bias. In addition, many infants did not meet the habituation criteria, raising concerns about the reliability of these criteria and their sensitivity to detect true habituation. Effect sizes may be overestimated due to small sample sizes. In addition, the classification of bilingualism was based on parental reports without detailed exposure data. Finally, a higher dropout rate in Study 2 may affect the generalizability of the results. Addressing these issues in future research will be critical to improving the accuracy and robustness of the findings.

Future Research

Future research in the area of early bilingualism and habituation studies should focus on refining methodological approaches and addressing existing gaps, such as incorporating continuous variables for bilingualism, increasing sample size and diversity, and re-evaluating habituation paradigms to gain deeper insights into infant cognitive processes.

In the introduction to this paper, we addressed the ongoing debate about the bilingual advantage, highlighting the inconsistent evidence for its presence in both children and adults.

As noted above, there is an urgent need for meta-analyses and systematic reviews that focus specifically on early bilingualism and infant studies. To gain a clearer understanding of the early bilingual advantage, it is essential to replicate existing studies and explore this phenomenon in greater depth. Recent research has argued for treating bilingualism as a continuous variable rather than a categorical one (De Cat et al., 2018; Gunnerud et al., 2020), which may provide more nuanced insights into its effects. In addition, it is critical that bilingual and monolingual samples are matched on socioeconomic status, educational level, and cultural dimensions to ensure valid comparisons (Sebastian-Galles & Santolin, 2020). Furthermore, research suggests that multilingual individuals may exhibit different cognitive patterns than bilinguals and monolinguals, highlighting the need for a more nuanced approach in future studies (Bailey et al., 2020).

Letting Go of Infant-Controlled Procedures

Infant-controlled designs are often used in habituation-dishabituation paradigms because they allow for early termination of the experiment and prevent infants from becoming fatigued prematurely. However, Kucharský et al. (2022) suggest in their recent review that using a fixed number of trials would provide a better insight into the habituation process itself. First, this is because simulation studies can model habituation processes. However, this is only possible with a fixed number of trials design, as data in infant controlled experimental designs are not missing at random and would therefore bias the analysis. Second, and more importantly, the inclusion of all participants in the analyses provides the opportunity to fully investigate the habituation process and its heterogeneity, as participants are not required to reach any criterion. This allows us to gain a deeper understanding of the factors that prevent infants from becoming habituated.

Habituators versus Non-Habituators

Typically, habituation studies attempt to habituate infants to a stimulus to test their preference for a novel stimulus over the familiar stimulus in the test phase. Therefore, infants who do not reach a criterion are excluded from the analyses, as we can only expect a novelty effect for infants who have been previously habituated (Hunter and Ames, 1988; Oakes, 2010). But how can we be sure that the infants who did not reach a criterion really did not habituate, since the criteria used are not optimal for detecting habituation? In addition, Poli et al. (2024) showed that infants who fail to habituate because they are slower at processing information can still show a preference for a novel stimulus if they are also more curious. Therefore, it would be interesting for future studies to include non-habituating infants in the analyses. In addition, more than 40 years ago, Cohen and Menten (1981) proposed the use of multi-population models to compare "habituators" and "non-habituators". With this approach, it is possible to answer the question of who shows a general decline in attention, who does not, and what the exact level of attention is at any given time. This could provide new insights into the population of "non-habituators" that is not addressed in most habituation-dishabituation studies (Kucharský et al., 2022).

A New Way to Use Habituation

As mentioned above, most habituation-dishabituation studies are designed to test for a novelty or familiarity effect, to assess whether a group of infants can discriminate between two stimuli. However, instead of testing for a novelty preference, one could present one stimulus for a fixed number of trials and then repeat the habituation phase with a different stimulus. If an infant is fully habituated to the first stimulus and has fully memorized it, a second stimulus would elicit a different orienting response in the form of a second habituation curve. This would indicate that the infant can discriminate between the two stimuli. If the first stimulus was not fully explored and the infant could not discriminate between the two stimuli, there would be only one habituation curve as the infant would continue to habituate. There could be four different outcomes: First, as mentioned above, the infant has habituated to both stimuli and can discriminate between them. This would result in two different habituation curves for each stimulus. Second, the infant has habituated to the first stimulus but cannot discriminate between the two stimuli. Therefore, the looks in the habituation curve would continue to decrease. Third, the infant has not habituated but can discriminate between the two stimuli. This would not result in a habituation curve, but the second stimulus would still elicit higher looking times. Finally, the infant is not habituated and cannot discriminate between the two stimuli. There is no systematic decrease in looking times for either stimulus (see Kucharský et al., 2022, Figure 5).

Summary

Future research on early bilingualism and habituation could treat bilingualism as a continuous variable, ensuring sample diversity and improving comparisons between groups. There is a need to re-evaluate habituation paradigms by incorporating fixed experimental designs that could better capture the habituation process, including the analysis of non-habituating infants. In addition, the use of novel approaches in habituation studies, such as the assessment of multiple habituation curves, could provide deeper insights into infants' cognitive processes and their ability to discriminate between stimuli.

Conclusion

The purpose of this master's thesis was to investigate the potential effects of bilingualism on habituation and dishabituation processes in infants. Our primary goal was to replicate the findings of Singh et al. (2015), who hypothesized that bilingual infants would show faster and stronger habituation, as well as a stronger dishabituation response. However, our results did not support the existence of a significant early bilingual advantage. The results of our study are consistent with recent discussions in the literature that question the universality of such an advantage. While early theories suggested that bilingualism might yield cognitive advantages through the constant need to switch between languages and inhibit one language while using the other (Green, 1998), more recent perspectives emphasize that any potential cognitive differences may instead arise from the different environmental experiences of bilingual infants (Bialystok, 2017; Singh et al., 2024). The lack of significant differences in our study may indicate that, despite being exposed to a more complex linguistic environment, bilingual infants do not necessarily process information more efficiently than their monolingual peers. Nevertheless, the cognitive and sensorimotor development of bilingual and monolingual children appears to progress at similar rates. Bilingual children may compensate for inefficient information processing by using other cognitive strategies, such as increased cognitive flexibility and greater reliance on contextual information. This cognitive flexibility may help compensate for developmental differences and may indicate a level of cognitive adaptation that allows for similar developmental outcomes.

While our study adds valuable data to the ongoing debate about the cognitive benefits of bilingualism, it also underscores the complexities and challenges (power, habituation criteria, classification of bilingualism, etc) involved in replicating these effects. The findings highlight the need for larger, more methodologically rigorous studies. Future research should focus on refining experimental designs, improving sample diversity, and adopting more nuanced approaches to better understand the cognitive processes underlying early bilingualism.

Endnote

During the writing of this text, we used both “DeepL” (DeepL SE, 2024) and “ChatGPT” (OpenAI, 2024). Specifically, “DeepL” was used for translation and linguistic correction, while “ChatGPT” was used for summarizing and understanding during literature research and for linguistic correction of text passages. After using these tools, the text was thoroughly reviewed and edited by the author of the thesis.

References

Anwari, F., Kievit, R., Lakens, D., Przybylski, A., Tiokhin, L., Wiernik, B., & Orben, A. (2021). *Evaluating the practical relevance of observed effect sizes in psychological research*. https://doi.org/10.31234/osf.io/g3vtr

Ashmead, D. H., & Davis, D. L. (1996). Measuring Habituation in Infants: An Approach Using Regression Analysis. *Child Development*, *67*(6), 2677–2690. https://doi.org/10.1111/j.1467-8624.1996.tb01882.x

Begus, K., Southgate, V., & Gliga, T. (2015). Neural mechanisms of infant learning: Differences in frontal theta activity during object exploration modulate subsequent object recognition. *Biology Letters*, *11*(5), 20150041. https://doi.org/10.1098/rsbl.2015.0041

Bergmann, C., Rabagliati, H., & Tsuji, S. (2019). *What's in a looking time preference?* https://doi.org/10.31234/osf.io/6u453

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, *143*(3), 233–262. https://doi.org/10.1037/bul0000099

Cohen, L. B., & Menten, T. G. (1981). The rise and fall of infant habituation. *Infant Behavior and Development*, *4*, 269–280.

Colombo, J., & Mitchell, D. W. (2009). Infant Visual Habituation. *Neurobiology of learning and memory*, *92*(2), 225–234. https://doi.org/10.1016/j.nlm.2008.06.002

Colombo, J., Richman, W. A., Shaddy, D. J., Follmer Greenhoot, A., & Maikranz, J. M. (2001). Heart Rate-Defined Phases of Attention, Look Duration, and Infant Performance in the Paired-Comparison Paradigm. *Child Development*, *72*(6), 1605–1616. https://doi.org/10.1111/1467-8624.00368

Dannemiller, J. L. (1984). Infant habituation criteria: I. A Monte Carlo study of the 50 % decrement criterion. *Infant Behavior and Development*, *7*(2), 147–166. https://doi.org/10.1016/S0163-6383(84)80055-7

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

DeepL SE. (2024). DeepL translator. https://www.deepl.com/

Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test. *International Review of Social Psychology*, *30*(1), 92–101. https://doi.org/10.5334/irsp.82

Dmitrienko, A., & D'Agostino Sr, R. (2013). Traditional multiplicity adjustment methods in clinical trials. *Statistics in Medicine*, *32*(29), 5172–5218. https://doi.org/10.1002/sim.5990

Donnelly, S., Brooks, P. J., & Homer, B. D. (2015). *Examining the Bilingual Advantage on Conflict Resolution Tasks: A Meta-Analysis.* CogSci.

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

D'Souza, D., & D'Souza, H. (2021). Bilingual adaptations in early development. *Trends in Cognitive Sciences*, *25*(9), 727–729. https://doi.org/10.1016/j.tics.2021.06.002

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. Cambridge Core. https://doi.org/10.1017/S1366728998000133

Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological review*, *77*(5), 419.

Gunnerud, H. L., Ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, *146*(12), 1059–1083. https://doi.org/10.1037/bul0000301

Hartkopf, J., Moser, J., Schleger, F., Preissl, H., & Keune, J. (2019). Changes in event-related brain responses and habituation during child development – A systematic literature review. *Clinical Neurophysiology*, *130*(12), 2238–2254. https://doi.org/10.1016/j.clinph.2019.08.029

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*(4), 341–348. https://doi.org/10.1002/icd.364

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research*.

Kalashnikova, M., Pejovic, J., & Carreiras, M. (2020). The Effects of Bilingualism on Attentional Processes in the First Year of Life. *Developmental Science*, *24*. https://doi.org/10.1111/desc.13011

Kavšek, M. (2013). The comparator model of infant visual habituation and dishabituation: Recent insights. *Developmental Psychobiology*, *55*(8), 793–808. https://doi.org/10.1002/dev.21081

Kosie, J. E., Zettersten, M., Abu-Zhaya, R., Amso, D., Babineau, M., Baumgartner, H. A., Bazhydai, M., Belia, M., Benavides-Varela, S., Bergmann, C., Berteletti, I., Black, A. K., Borges, P., Borovsky, A., Byers-Heinlein, K., Cabrera, L., Calignano, G., Cao, A., Chijiiwa, H., … Lew-Williams, C. (2023). *ManyBabies 5: A large-scale investigation of the proposed shift from familiarity preference to novelty preference in infant looking time*. https://doi.org/10.31234/osf.io/ck3vd

Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*, *106*(16), 6556–6560.

Kucharský, Š., Zaharieva, M., Raijmakers, M., & Visser, I. (2022). Habituation, part II. Rethinking the habituation paradigm. *Infant and Child Development*, e2383. https://doi.org/10.1002/icd.2383

Lavoie, C., & Desrochers, S. (2002). Visual Habituation at Five Months: Short-Term Reliability of Measures Obtained With a New Polynomial Regression Criterion. *The*

Journal of Genetic Psychology, *163*(3), 261–271. https://doi.org/10.1080/00221320209598682

Lowe, C. J., Cho, I., Goldsmith, S. F., & Morton, J. B. (2021). The Bilingual Advantage in Children's Executive Functioning Is Not Related to Language Status: A Meta-Analytic Review. *Psychological Science*, *32*(7), 1115–1146. https://doi.org/10.1177/0956797621993108

Mandler, J. M. (2004). Thought before language. *Trends in Cognitive Sciences*, *8*(11), 508–513. https://doi.org/10.1016/j.tics.2004.09.004

Mangold, P. (2018). *Das Unsichtbare entdecken durch werkzeuggestützte wissenschaftliche Beobachtung*.

Oakes, L. M. (2010). Using Habituation of Looking Time to Assess Mental Processes in Infancy. *Journal of cognition and development : official journal of the Cognitive Development Society*, *11*(3), 255–268. https://doi.org/10.1080/15248371003699977

Oakes, L. M. (2017). Sample Size, Statistical Power, and False Conclusions in Infant Looking-Time Research. *Infancy*, *22*(4), 436–469. https://doi.org/10.1111/infa.12186

OpenAI. (2024). [Large language model]. ChatGPT-4o. https://www.openai.com/chatgpt

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

Paap, K., Schwieter, J., & Paradis, M. (2019). *The Bilingual Advantage Debate*. 701–735. https://doi.org/10.1002/9781119387725.ch34

Pahnke, J. (2007). *VISUELLE HABITUATION UND DISHABITUATION ALS MAßE KOGNITIVER FÄHIGKEITEN IM SÄUGLINGSALTER: Individuelle Differenzen in Habituationsaufgaben zur Unterscheidung einzelner vs. Kategorialer Stimuli*. Ruprecht-Karls-Universität Heidelberg.

Paulus, M. (2022). Should infant psychology rely on the violation-of-expectation method? Not anymore. *Infant and Child Development*, *31*(1), e2306. https://doi.org/10.1002/icd.2306

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*(1), 41–58. https://doi.org/10.1017/S0142716400009863

Poli, F., Ghilardi, T., Beijers, R., de Weerth, C., Hinne, M., Mars, R. B., & Hunnius, S. (2024). Individual differences in processing speed and curiosity explain infant habituation and dishabituation performance. *Developmental Science*, *27*(3), e13460. https://doi.org/10.1111/desc.13460

Schöner, G., & Thelen, E. (2006). Using Dynamic Field Theory to Rethink Infant Habituation. *Psychological review*, *113*, 273–299. https://doi.org/10.1037/0033-295X.113.2.273

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A Bilingual Advantage in Visual Language Discrimination in Infancy. *Psychological Science*, *23*(9), 994–999. https://doi.org/10.1177/0956797612436817

Singh, L. (2021). Evidence for an Early Novelty Orientation in Bilingual Learners. *Child Development Perspectives*, *Early View*, 1–7. https://doi.org/10.1111/cdep.12407

Singh, L., Barr, R., Quinn, P. C., Kalashnikova, M., Rocha-Hidalgo, J., Freda, K., & D'Souza, D. (2024). Effects of environmental diversity on exploration and learning: The case of bilingualism. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001562

Singh, L., Fu, C. S. L., Rahman, A. A., Hameed, W. B., Sanmugam, S., Agarwal, P., Jiang, B., Chong, Y. S., Meaney, M. J., Rifkin-Graboi, A., & Team, the G. R. (2015). Back to Basics: A Bilingual Advantage in Infant Visual Habituation. *Child Development*, *86*(1), 294–302. https://doi.org/10.1111/cdev.12271

Singh, L., Kalashnikova, M., & Quinn, P. (2023). Bilingual infants readily orient to novel visual stimuli. *Journal of Experimental Psychology: General*, *152*(11), 3218–3228. https://doi.org/10.1037/xge0001444

Sokolov, E. N., Waydenfeld, S. W., Worters, R., & Clarke, A. (1963). Perception and the conditioned reflex. *(No Title)*.

Spit, S., Geambașu, A., Renswoude, D. van, 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*, *26*(6), e13377. https://doi.org/10.1111/desc.13377

Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological review*, *73*(1), 16.

Tsui, A., & Fennell, C. (2019). *Do Bilingual Infants Possess Enhanced Cognitive Skills?* (PPR332765). https://doi.org/10.31234/osf.io/dcn7y

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). https://doi.org/10.3390/bs9030027

Zaharieva, M. S., Kucharský, Š., Colonnesi, C., Gu, T., Jo, S., Luttenbacher, I., Mannsdörfer, L., Matetovici, M., Mickute, U., & Raijmakers, M. (2021). Habituation, Part I. Design choices in the infant habituation paradigm: A pre-registered crowd-sourced systematic review and meta-analysis. *Infant and Child Development*.

List of Figures

List of Tables

Appendix

A. Abstract

There is an ongoing debate about whether bilingualism provides cognitive benefits, particularly in terms of attention, memory, and learning. However, the evidence is mixed, and more research is needed, especially for young children. This study examines the effects of early bilingualism on habituation and dishabituation processes in infants. Habituation refers to the reduction of attention to a repeated stimulus, whereas dishabituation occurs when a novel stimulus regains attention. Two experiments were conducted with 7- to 8-month-old monolingual and bilingual infants using an infant-controlled habituation-dishabituation paradigm. The study aimed to replicate and extend previous findings by Singh et al. (2015), which suggested that bilingual infants may process information more efficiently, leading to faster habituation and stronger preference for novel stimuli. However, we did not find significant differences between bilingual and monolingual infants in the speed of habituation, strength of habituation, or dishabituation response. The results suggest that the expected cognitive advantages of bilingualism were not observed, possibly due to limitations such as small sample size and low statistical power. This study highlights the need for further research with larger sample sizes to better understand the impact of bilingualism on early cognitive development.

Keywords: *early bilingualism, information processing, habituation, dishabituation, bilingual advantage, infant cognition*

B. Zusammenfassung

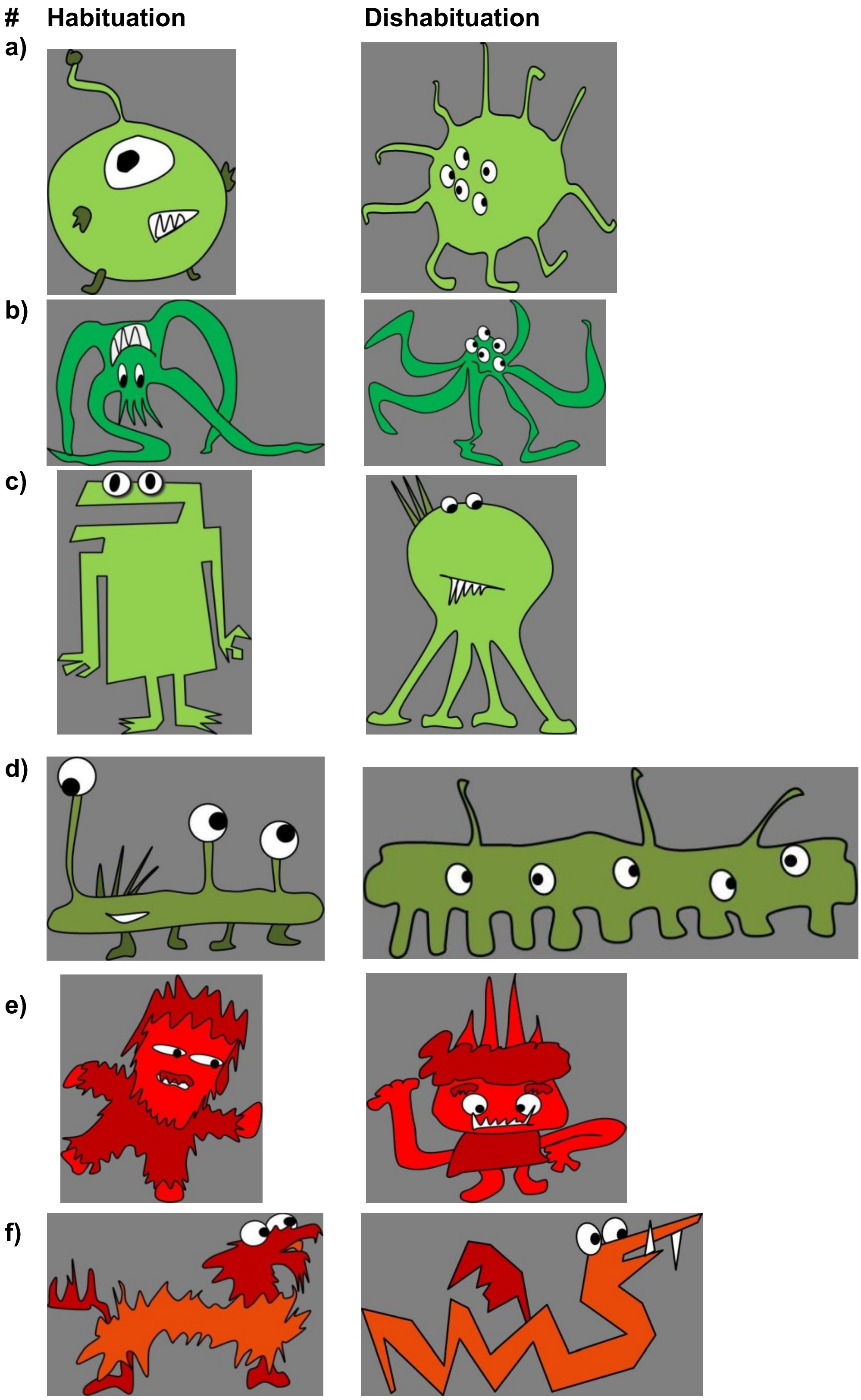
Die Frage, ob Zweisprachigkeit kognitive Vorteile wie verbesserte Aufmerksamkeit, besseres Gedächtnis und effizienteres Lernen bietet, ist Gegenstand einer anhaltenden wissenschaftlichen Debatte. Insbesondere bei Kleinkindern ist die Beweislage uneinheitlich, was den Bedarf an weiterer Forschung unterstreicht. Diese Studie untersucht die Auswirkungen früher Zweisprachigkeit auf Habituations- und Dishabituationsprozesse bei Säuglingen im Alter von 7 bis 8 Monaten. Habituation bezieht sich auf die Abnahme der Aufmerksamkeit gegenüber einem wiederholt dargebotenen Reiz, während Dishabituation auftritt, wenn die Aufmerksamkeit gegenüber einem neuen Reiz wieder zunimmt. Es wurden zwei Experimente durchgeführt, in denen ein „infant-controlled" Habituation-Dishabituation-Paradigma verwendet wurde. Die Stichprobe setzte sich aus monolingualen und bilingualen Säuglingen zusammen. Ziel der Studie war es, frühere Ergebnisse von Singh et al. (2015) zu replizieren und zu erweitern, die nahelegten, dass bilinguale Säuglinge Informationen effizienter verarbeiten können, was sich in einer schnelleren Habituation und einer stärkeren Präferenz für neue Stimuli manifestiert. Im Gegensatz zu den erwarteten Ergebnissen fanden wir jedoch keine signifikanten Unterschiede zwischen bilingualen und monolingualen Säuglingen hinsichtlich der Habituationsgeschwindigkeit, der Stärke der Habituation oder der Dishabituationsreaktion. Diese Ergebnisse könnten durch methodische Einschränkungen wie eine kleine Stichprobengröße und eine daraus resultierende geringe statistische Aussagekraft erklärt werden. Die Studie unterstreicht die Notwendigkeit weiterer Forschung mit größeren und vielfältigeren Stichproben, um die möglichen Auswirkungen der Zweisprachigkeit auf die kognitive Entwicklung in der frühen Kindheit genauer zu untersuchen.

Schlagwörter: *frühe Zweisprachigkeit, Informationsverarbeitung, Habituation, Dishabituation, bilingualer Vorteil, frühkindliche Kognition*

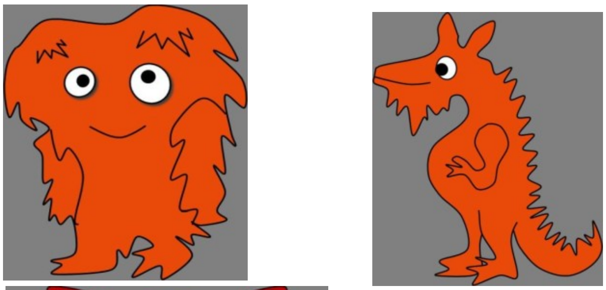
C. Stimuli

Figure C1

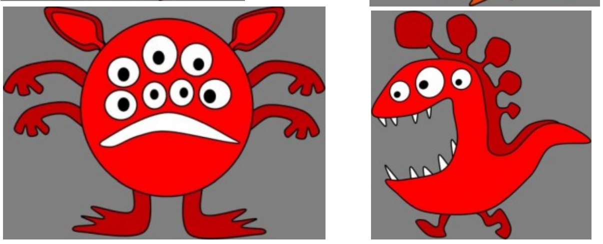
Habituation and Dishabituation Stimuli



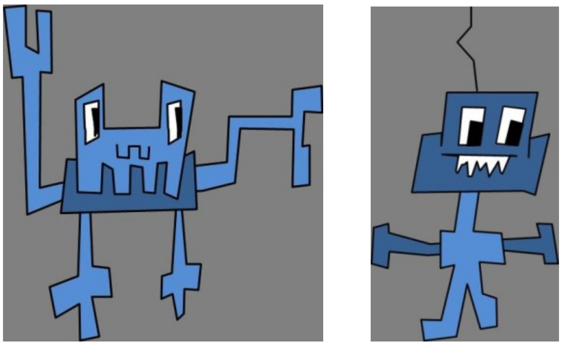
g)



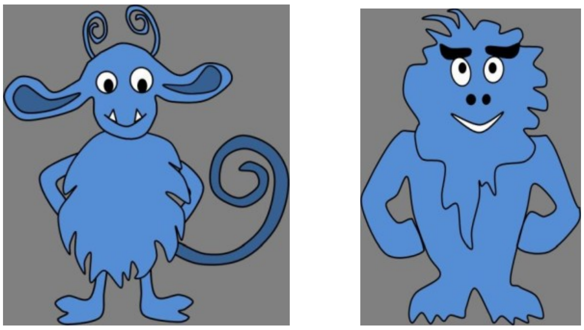
h)



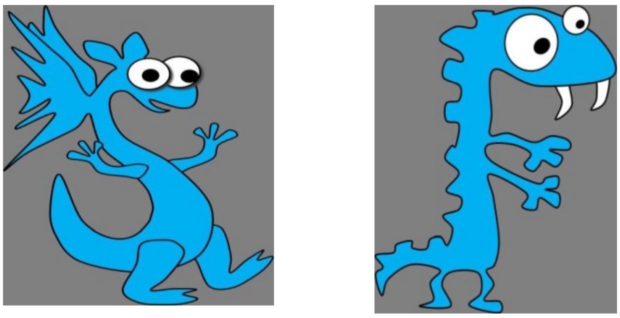
i)



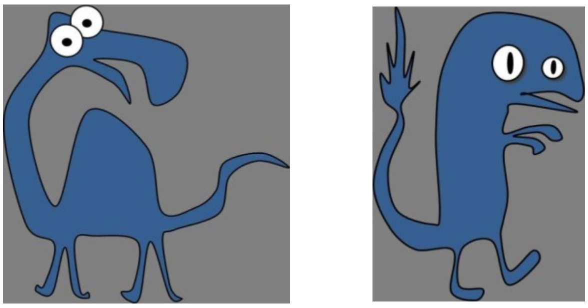
j)



k)



l)

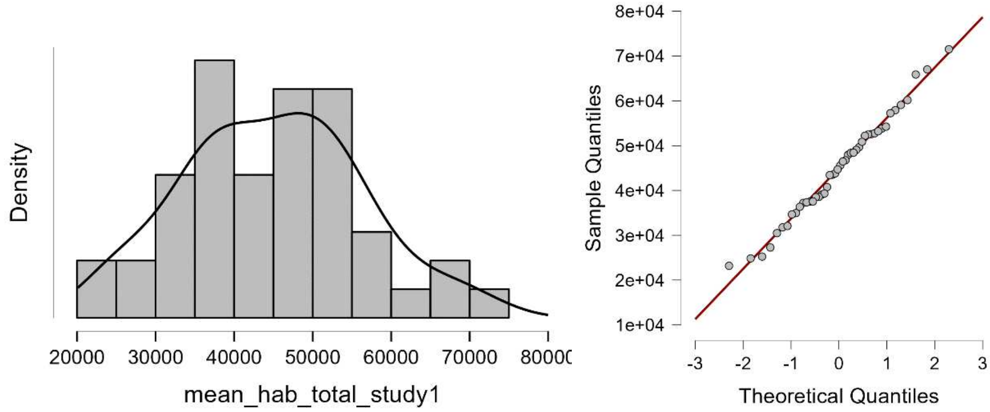


D. Distribution Plots

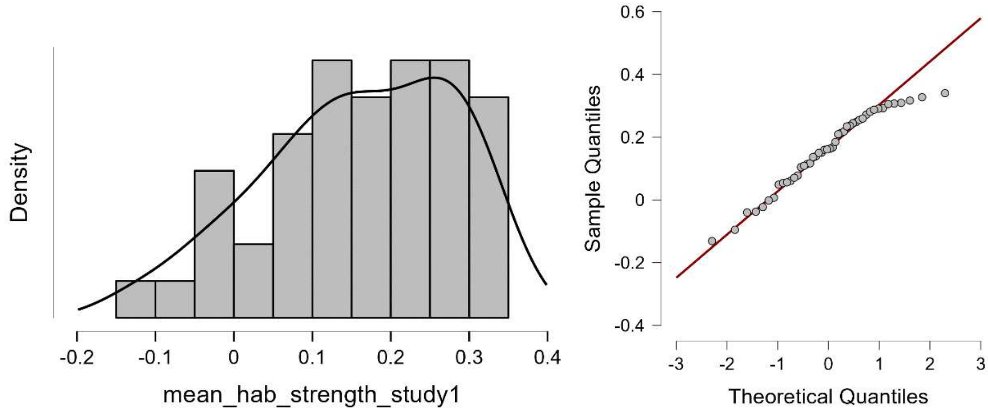
Figure D1

Distribution Plots Study 1

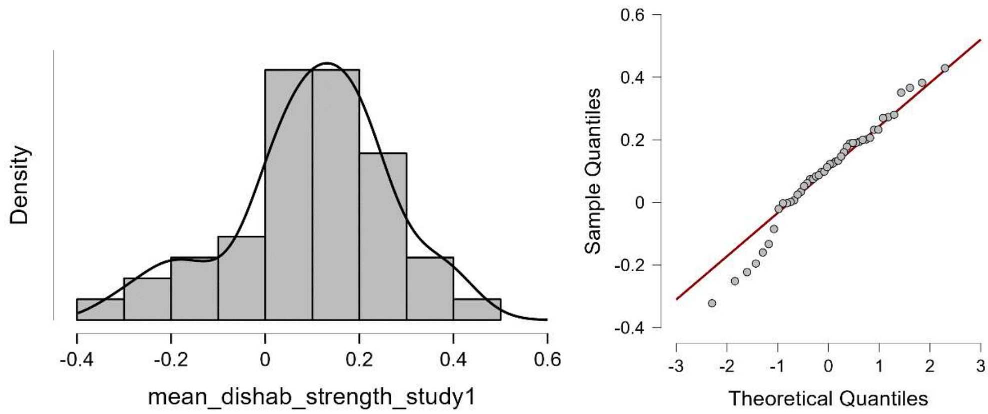
A) Habituation Speed



B) Habituation Strength



C) Dishabituation Response

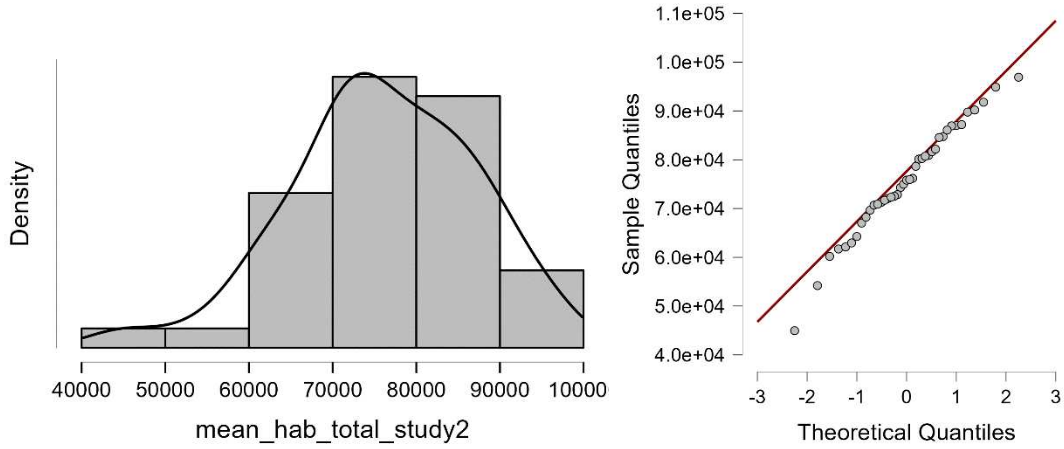


Note. These Figures show the distribution and corresponding Q-Q plots of the dependent variables from Study 1, indicating that a normal distribution is not completely given.

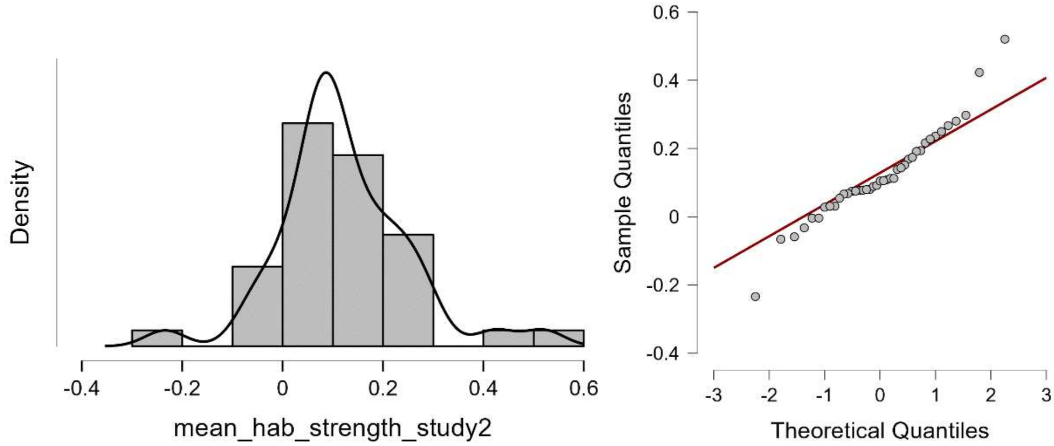
Figure D2

Distribution Plots Study 2

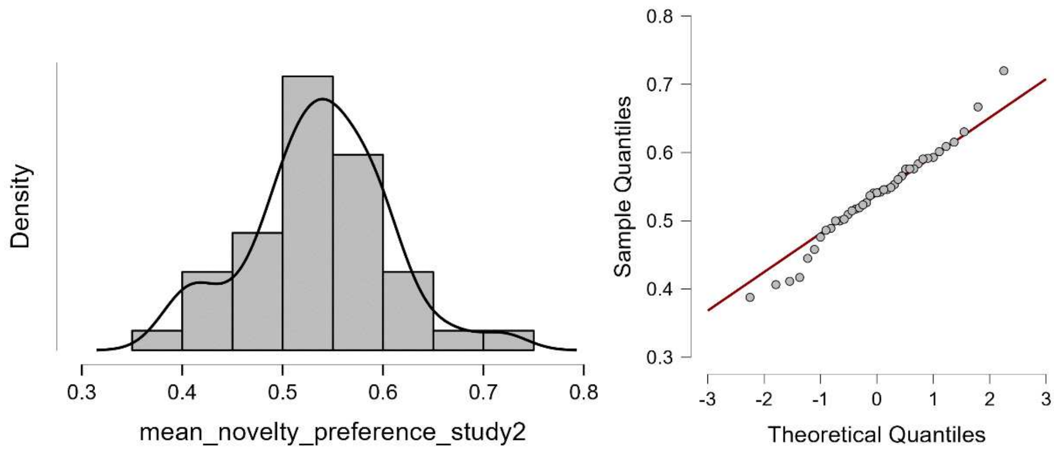
A) Habituation Speed



B) Habituation Strength



C) Dishabituation Response



Note. These Figures show the distribution and the corresponding Q-Q plots of the dependent variables from Study 2, indicating that a normal distribution is not completely given.

E. Supplementary tables

Table E1

Frequencies for fully watched phases

	phase	Frequency	Percent	Valid Percent	Cumulative Percent
Study 1	1	4	8.696	8.696	8.696
OREP	2	6	13.043	13.043	21.739
	3	10	21.739	21.739	43.478
	4	8	17.391	17.391	60.870
	5	8	17.391	17.391	78.261
	6	5	10.870	10.870	89.130
	7	5	10.870	10.870	100.000
	Missing	0	0.000		
	Total	46	100.000		
Study 2	2	13	31.707	31.707	31.707
FLOREP	3	15	36.585	36.585	68.293
	4	11	26.829	26.829	95.122
	5	2	4.878	4.878	100.000
	Missing	0	0.000		
	Total	41	100.000		

Note. In Study 1 (OREP), between 1 and 7 phases were observed. In Study 2 (FLOREP), a minimum of 2 phases could be observed. In addition, phases could be skipped to ensure that the stimulation was viewed in its entirety.

Table E2

Descriptive Statistics for Study 1 and Study 2

	Valid	Mean	Std. Error of Mean	95% Confidence Interval Mean		Std. Deviation	Minimum	Maximum
				Upper	Lower			
Study 1 - OREP								
Standard Habituation Criterion								
Habituation Speed	28	32775.952	2158.883	37205.614	28346.290	11423.735	12800.000	59760.000
Habituation Strength	28	0.443	0.016	0.476	0.410	0.085	0.342	0.631
Dishabituation Response	28	0.371	0.031	0.435	0.306	0.167	0.036	0.692
Peak Habituation Criterion								
Habituation Speed	34	34773.686	1758.423	38351.224	31196.149	10253.278	12800.000	59760.000
Habituation Strength	34	0.367	0.018	0.405	0.330	0.106	0.106	0.556
Dishabituation Response	34	0.286	0.037	0.361	0.212	0.213	-0.273	0.678
Study 2 - FLOREP								
Standard Habituation Criterion								
Habituation Speed	15	62455.633	3803.060	70612.386	54298.881	14729.188	22000.000	88107.000
Habituation Strength	15	0.465	0.033	0.536	0.394	0.129	0.346	0.769
Dishabituation Response	15	0.563	0.030	0.627	0.499	0.116	0.288	0.752
Peak Habituation Criterion								
Habituation Speed	22	60465.659	2464.768	65591.425	55339.893	11560.787	34267.000	88107.000
Habituation Strength	21	0.369	0.039	0.449	0.288	0.177	0.092	0.769
Dishabituation Response	22	0.529	0.029	0.590	0.468	0.138	0.205	0.752

Note. Descriptive statistics for all three dependent variables for both habituation criteria.

Table E3

Descriptive Statistics Study 1 OREP - split by Linguality

		Valid	Mean	Std. Error of Mean	95% Confidence Interval Mean Upper	95% Confidence Interval Mean Lower	Std. Deviation	Minimum	Maximum
Study 1 - OREP									
Standard Habituation Criterion									
Habituation Speed	monolingual	18	35647.407	2160.380	40205.412	31089.403	9165.718	20160.000	49360.000
	bilingual	10	27607.333	4320.248	37380.414	17834.253	13661.825	12800.000	59760.000
Habituation Strength	monolingual	18	0.441	0.021	0.485	0.397	0.089	0.342	0.631
	bilingual	10	0.446	0.026	0.505	0.387	0.082	0.356	0.567
Dishabituation Response	monolingual	18	0.412	0.036	0.489	0.335	0.155	0.229	0.692
	bilingual	10	0.296	0.053	0.417	0.175	0.169	0.036	0.577
Peak Habituation Criterion									
Habituation Speed	monolingual	23	36226.174	2009.711	40394.060	32058.288	9638.236	14080.000	49360.000
	bilingual	11	31736.667	3405.337	39324.231	24149.103	11294.225	12800.000	59760.000
Habituation Strength	monolingual	23	0.368	0.023	0.416	0.320	0.112	0.106	0.556
	bilingual	11	0.367	0.030	0.433	0.300	0.099	0.190	0.498
Dishabituation Response	monolingual	23	0.293	0.048	0.392	0.194	0.229	-0.273	0.678
	bilingual	11	0.272	0.056	0.396	0.148	0.185	-0.017	0.577

Note. Descriptive statistics for all three dependent variables for both habituation criteria split by linguality. Habituation speed (sum of trials 1 to 12) is reported in milliseconds. Habituation strength is measured as $\frac{(\text{first 3 trials} - \text{last 3 trials})}{(\text{first 3 trials} + \text{last 3 trials})}$, yielding values between -1 and +1. Dishabituation response is measured in Study 1 as $\frac{test\ trials - last\ 2\ trials}{test\ trials + last\ 2\ trials}$, yielding values between -1 and +1 - positive values indicate a preference for the novel stimulus.

Table E4

Descriptive Statistics Study 2 FLOREP - split by Linguality

					95% Confidence Interval Mean				
		Valid	**Mean**	**Std. Error of Mean**	**Upper**	**Lower**	**Std. Deviation**	**Minimum**	**Maximum**
Study 2 - FLOREP									
Standard Habituation Criterion									
Habituation Speed	monolingual	10	65759.750	3450.216	73564.680	57954.820	10910.540	50160.000	88107.000
	bilingual	5	55847.400	9050.998	80976.999	30717.801	20238.647	22000.000	71202.000
Habituation Strength	monolingual	10	0.445	0.034	0.523	0.367	0.109	0.346	0.691
	bilingual	5	0.505	0.075	0.714	0.297	0.168	0.368	0.769
Dishabituation Response	monolingual	10	0.567	0.044	0.665	0.468	0.138	0.288	0.752
	bilingual	5	0.556	0.030	0.638	0.473	0.066	0.480	0.663
Peak Habituation Criterion									
Habituation Speed	monolingual	15	62774.900	2659.351	68478.642	57071.158	10299.624	50160.000	88107.000
	bilingual	7	55517.286	5047.730	67868.637	43165.935	13355.039	34267.000	71202.000
Habituation Strength	monolingual	15	0.343	0.042	0.433	0.253	0.163	0.092	0.691
	bilingual	6	0.433	0.086	0.654	0.212	0.210	0.178	0.769
Dishabituation Response	monolingual	15	0.513	0.042	0.603	0.424	0.162	0.205	0.752
	bilingual	7	0.563	0.022	0.618	0.508	0.059	0.480	0.675

Note. Descriptive statistics for all three dependent variables for both habituation criteria split by linguality. Habituation speed (sum of trials 1 to 15) is reported in milliseconds. Habituation strength is measured as $\frac{(\text{first 3 trials} - \text{last 3 trials})}{(\text{first 3 trials} + \text{last 3 trials})}$, yielding values between -1 and +1. Dishabituation response is measured in Study 2 as $\frac{novel\ looking\ time}{novel\ looking\ time + familiar\ looking\ time}$, yielding values between -1 and +1 - positive values indicating a preference for the novel stimulus.

Table E5

Exploratory Analyses Independent Samples T-test comparing Study 1 (OREP) and Study 2 (FLOREP)

							95% CI for Effect Size	
	Test	**Statistic**	**df**	**p**	**Effect Size**	**SE Effect Size**	**Lower**	**Upper**
Mean Habituation	Welch	-12.875	84.192	**< .001***	-2.764	0.359	-3.352	-2.167
	Mann-Whitney	52.000		**< .001***	-0.945	0.124	-0.966	-0.912
Mean Dishabituation	Welch	-4.666	84.939	**< .001***	-1.000	0.239	-1.444	-0.550
	Mann-Whitney	460.500		**< .001***	-0.512	0.124	-0.669	-0.310
Mean First 3	Welch	-5.132	84.685	**< .001***	-1.097	0.243	-1.546	-0.642
	Mann-Whitney	394.500		**< .001***	-0.582	0.124	-0.721	-0.397
Standard Habituation Criterion	Welch	3.527	71.120	**< .001***	0.746	0.228	0.305	1.182
	Mann-Whitney	1261.500		**0.003***	0.338	0.124	0.107	0.534
Peak Habituation Criterion	Welch	3.966	67.222	**< .001***	0.837	0.232	0.390	1.278
	Mann-Whitney	1299.000		**0.001***	0.378	0.124	0.152	0.566

Note. For the Welch t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

* $p < .05$

Table E6

Pearson's Correlations Study 1 OREP

Variable		sex	age	linguality	phases	study	hab speed	hab strength	dishab strength	first 3	last 3	dishab
1. sex	Pearson's r	—										
	p-value	—										
2. age	Pearson's r	**0.223** *	—									
	p-value	0.038	—									
3. linguality	Pearson's r	-0.119	-0.013	—								
	p-value	0.274	0.902	—								
4. phases	Pearson's r	0.071	-0.108	0.110	—							
	p-value	0.512	0.322	0.310	—							
5. study	Pearson's r	0.142	**0.248** *	0.139	**-0.311** **	—						
	p-value	0.190	0.021	0.200	0.003	—						
6. hab speed	Pearson's r	0.187	-0.094	-0.231	-0.032	NaN	—					
	p-value	0.214	0.534	0.123	0.835	NaN	—					
7. hab strength	Pearson's r	-0.084	0.235	0.105	**0.383** **	NaN	-0.279	—				
	p-value	0.579	0.116	0.486	0.009	NaN	0.060	—				
8. dishab strength	Pearson's r	0.036	0.270	0.090	**0.406** **	NaN	-0.047	**0.605** ***	—			

Variable		sex	age	linguality	phases	study	hab speed		hab strength		dishab strength		first 3		last 3		dishab
	p-value	0.812	0.070	0.554	0.005	NaN	0.754		< .001		—						
9. first 3	Pearson's r	0.068	0.034	-0.194	0.176	NaN	**0.812**	***	0.189		0.209		—				
	p-value	0.655	0.821	0.195	0.243	NaN	< .001		0.209		0.164		—				
10. last 3	Pearson's r	0.202	-0.122	-0.204	-0.149	NaN	**0.833**	***	**-0.667**	***	**-0.350**	*	**0.576**	***	—		
	p-value	0.179	0.420	0.174	0.321	NaN	< .001		< .001		0.017		< .001		—		
11. dishab	Pearson's r	0.204	0.085	-0.174	0.146	NaN	**0.680**	***	-0.067		**0.508**	***	**0.687**	***	**0.562**	***	—
	p-value	0.173	0.574	0.248	0.332	NaN	< .001		0.657		< .001		< .001		< .001		—

Note. The variance in study is equal to 0. * p < .05, ** p < .01, *** p < .001

Table E7

Pearson's Correlations Study 2 FLOREP

Variable		sex	age	ling.	phases	study	hab speed	hab strength	dishab response	stim	dishab	test	first 3	Last 3
1. sex	Pearson's r	—												
	p-value	—												
2. age	Pearson's r	**0.223** *	—											
	p-value	0.038	—											
3. linguality	Pearson's r	-0.119	-0.013	—										
	p-value	0.274	0.902	—										
4. phases	Pearson's r	0.071	-0.108	0.110	—									
	p-value	0.512	0.322	0.310	—									
5. study	Pearson's r	0.142	**0.248** *	0.139	**-0.311** **	—								
	p-value	0.190	0.021	0.200	0.003	—								
6. hab speed	Pearson's r	-0.136	-0.075	0.109	0.286	NaN	—							
	p-value	0.395	0.639	0.496	0.070	NaN	—							
7. hab strength	Pearson's r	0.068	0.283	**-0.317** *	-0.163	NaN	-0.256	—						
	p-value	0.672	0.072	0.043	0.308	NaN	0.106	—						
8. dishab response	Pearson's r	0.039	-0.006	-0.039	0.038	NaN	-0.003	0.251	—					

Variable		sex	age	ling.	phases	study	hab speed	hab strength	dishab response	stim	dishab	test	first 3	Last 3
	p-value	0.811	0.972	0.807	0.812	NaN	0.983	0.114	—					
9. stimulation	Pearson's r	-0.066	-0.085	-0.223	0.159	NaN	**0.419** **	-0.135	0.159	—				
	p-value	0.682	0.599	0.161	0.322	NaN	0.006	0.399	0.319	—				
10. dishab	Pearson's r	-0.035	-0.032	0.100	0.292	NaN	**0.525** ***	-0.244	**0.413** **	**0.437** **	—			
	p-value	0.827	0.841	0.535	0.064	NaN	< .001	0.125	0.007	0.004	—			
11. test	Pearson's r	-0.127	-0.129	0.191	0.236	NaN	**0.549** ***	**-0.440** **	**-0.496** ***	0.287	**0.544** ***	—		
	p-value	0.428	0.422	0.232	0.137	NaN	< .001	0.004	< .001	0.069	< .001	—		
12. first 3	Pearson's r	-0.121	0.211	-0.158	0.116	NaN	**0.680** ***	**0.350** *	0.162	**0.329** *	**0.448** **	0.297	—	
	p-value	0.450	0.186	0.323	0.470	NaN	< .001	0.025	0.313	0.036	0.003	0.059	—	
13. last 3	Pearson's r	-0.079	-0.125	0.285	0.305	NaN	**0.762** ***	**-0.739** ***	-0.087	**0.377** *	**0.576** ***	**0.622** ***	**0.335** *	—
	p-value	0.625	0.438	0.071	0.053	NaN	< .001	< .001	0.590	0.015	< .001	< .001	0.032	—

Note. The variance in study is equal to 0. * p < .05, ** p < .01, *** p < .001