



Visual attention in infancy: Using online webcam recordings as a proof of concept

**M. H. I. Jansen, J. C. van Elst, M. L. Villeneuve, D. J. H. Capel, C. L. Koevoets,
MabyBabies-AtHome, F. N. K. Wijnen, C. M .M. Junge**

Neuroscience and Cognition, Major Research Project

Student: M. H. I. Jansen

Supervisor: Dr. C. M. M. Junge

Date: 2021, October 17

Words: 6322

Abstract. Many infant studies have small and homogenous samples, which leads to low statistical power, low statistical conclusion validity, and low external validity; hence, the conclusions drawn in these studies might not be representative to the general population. To resolve such limitations, the ManyBabies Consortium aims to develop universal methodological frameworks to replicate key findings in developmental studies with large, heterogeneous samples. This experimental pilot study ($n = 16$) of the ManyBabies-AtHome project examined whether it is feasible to track 8- to 10-months-old infants' visual attention in their home environment using remote webcam recordings, to replicate findings of infants' preference for dynamic over static stimuli. Additionally, it was explored whether one of the two elaborated preferential looking designs, subsequent or side-by-side, suits this remote preferential looking testing framework better. In accordance with the hypotheses, this study showed that infants' preference for dynamic stimuli could be replicated using remote webcam recordings. Moreover, no clear distinction in suitability for one of the two designs was found, with good and excellent inter-rater reliability scores for the designs, and spatial location of stimuli on the screen could be accurately determined by observers based on average total looking-times. Thus, it seems that visual attention can be examined using remote visual preference paradigms, which can possibly enhance the validity of future infant studies through increase of sample sizes and heterogeneity of samples.

Key words: *Visual attention, preferential looking, infancy, online experiment, webcam recordings, proof of concept*

Acknowledgements

I would like to thank my supervisor Dr. Caroline Junge for her extensive feedback and providing me the opportunity to further develop my research skills as an independent researcher. Furthermore, I thank the ManyBabies-AtHome team for allowing me to work alongside them on their project, and providing access to their concepts and stimuli. In addition, I would like to thank Jacco van Elst, Michael Villeneuve, and the remaining technical-support staff of the Utrecht Institute of Linguistics (UiL) OTS labs, without their effort coding the experiment, this study could not have been completed. I also thank Dr. Desiree Capel for her persistent support and involvement in this project, and for enabling access to the UiL OTS Babylab database. Furthermore, I thank Charlotte Koevoets for her assistance as second observer and her guidance in the practical skills necessary for this project. Moreover, I thank Prof. Dr. Frank Wijnen and Dr. Karin Wanrooij for reviewing the manuscript. Additionally, I thank the UiL OTS Babylab research group for their feedback. Lastly, I want to thank all the parents and infants that participated in the experiment.

Layman's summary

Usually, to test what infants can and cannot yet perceive, infants are tested in well-controlled infant laboratory settings at universities. However, not everyone lives close to a university or has the time to go there during weekdays, which amongst other things, results in most experiments only testing small groups of babies, often with highly educated parents. This brings up another issue, namely that researchers draw conclusions based on results from small, specific groups of babies, to say something about larger, more diverse groups of babies. Even when, in reality, the results from the small group actually differ from the results from the larger group of babies. To test larger groups of babies, the burden of the necessity of parents coming into the lab should be reduced, in addition to increasing diversity through worldwide collaborations, such as the ManyBabies Consortium. So, with the current testing-phase study we investigated

if it is possible for parents to participate with their baby in an online experiment, using webcam recordings from their own computer, in their own home, at their own time. This way, also parents that do not live nearby a university, or do not have time during working-hours, can still participate in the experiment.

To run this testing-phase study, we used webcam recordings of sixteen 8- to 10-months-old babies. After parents agreed to participate, we sent them an email with an information letter and a link to the experiment. After starting the experiment, parents received a step-by-step instruction on creating the correct setup and what to expect during the task.

The task for babies took about five-and-a-half-minutes. On the screen, one or two bullseye images were presented on the left and/or right side. In total, there were 4 different bullseye images used in the task: a simple and complex static image, and a simple and complex moving image. The moving bullseye images were moving in a spiraling motion the entire time they were shown on the screen.

The results of the experiment showed that when we observed the movements of babies' eyes, we could clearly see, and score by hand, where babies were looking at on the screen. Also, we were able to find very similar results to what was previously found in controlled laboratory experiments. For example that babies looked longer at moving images than at static images. Furthermore, we found that we could use and study the webcam recordings in an accurate and trustworthy way.

In conclusion, it seems that visual attention can be studied using webcam recordings of an online, remote visual preference experiment, that were scored by hand. Thus, if the findings of this testing-phase study hold after more babies are tested, it could mean that future baby research could use online experiments to lower the burden for parents to participate with their baby. And thereby, the size and the diversity of the group of babies that participate in

experiments can be increased, which leads to more correct findings about babies and their development.

Introduction

Visual attention in infancy is studied in many babylabs around the world. It is often assessed through visual preference paradigms in which looking-times are the primary dependent variable (Oakes, 2017; Spelke, 1985; Teller, 1979). Research shows that the development of visual attention starts very early in life (i.e., Aslin, 1985; Spelke, 1985). Thus, from a very young age, infants are able to exhibit their preference for looking at an object or stimulus, as reflected by longer looking-times, from which researchers conclude that infants can distinguish differences between these stimuli or objects. This early development makes preferential looking a practical, stable, and reliable behavioral measure to study preferences in infants (Spelke, 1985). Hence, visual preference paradigms are found in studies of, for instance, language development (i.e., Imafuku et al., 2019; Junge et al., 2012; Tan & Burnham, 2019; The ManyBabies Consortium, 2020), categorization (i.e., Rennels et al., 2016), recognition (i.e., Reynolds, 2015), and novelty preference (i.e., Fantz, 1964; Rose et al., 2004), but are also used to examine early biomarkers for autism (i.e., Pierce et al., 2016; Vacas et al., 2021). However, multiple limitations occur in infant studies (DeBolt et al., 2020).

One of the main limitations are small sample sizes, ranging between 8 and 16 infants (Byers-Heinlein et al., 2021; Frank et al., 2017; Oakes, 2017), which leads to low statistical power (Oakes, 2017). Hence, the conclusions drawn in infant studies might be problematic and findings might not be reproducible (Oakes, 2017), due to an increased possibility of false positive results and a reduction of sensitivity to detect true differences (Button et al., 2013; Cohen, 1962; Fraley & Vazire, 2014; Oakes, 2017). Furthermore, by drawing subsamples from larger samples in previous studies, Oakes (2017) showed that having small sample sizes increases the possibility of ambiguous or false results, as opposed to larger sample sizes with

20 to 30 infants. In sum, small sample sizes lead to low statistical power, which in its turn reduces the accuracy and reasonability of research conclusions, also known as statistical conclusion validity (Cook and Campbell, 1979 in García-Pérez, 2012).

In addition to small sample sizes and low statistical power, poor generalizability is another limitation occurring in infant studies. Most infant studies rely on participants from middle-class families (Fernald, 2010). Fernald states that only parents with time, resources, motivation, and who are living nearby a university will make the effort to participate with their infant in developmental studies at a university; hence, parents with high and homogenous social economic status (SES) are represented in these studies (Fernald, 2010). SES, in its turn, is strongly related to cognitive development in infants and children (i.e., Clearfield & Niman, 2012; Farah et al., 2008; Herrmann & Guadagno, 1997). Clearfield and Niman (2012) tested high- and low-SES infants longitudinally at 6, 9, and 12 months of age. They found that high- and low-SES infants have different cognitive developmental trajectories, with low-SES infants already being delayed by 6 months in contrast to high-SES infants (Clearfield & Niman, 2012). In conclusion, to truly capture cognitive development in a broad infant population, it is important that infant studies ensure a diverse, heterogenic sample, in order to increase generalizability, and thus external validity, of findings.

To overcome limitations in infant research, many babylabs across the globe are working together in the ManyBabies (MB) Consortium. MB aims to develop universal methodological frameworks to replicate existing paradigms to increase: the validity of key findings in infancy studies, the understanding of infant development, and the diversity of samples and researchers (Frank et al., 2017; Visser et al., 2021). To achieve these goals, MB initiated multiple projects, one of which is the ManyBabies-AtHome (MBAH) project (Zaadnoordijk et al., 2021). MBAH is a methodological project in which an online, remote testing framework is developed to test infants' looking-times in their home environment at times best suited for the individual infant

and their parent(s) (Zaadnoordijk et al., 2021) using the online platform LookIt (Scott & Schulz, 2017). LookIt is an online developmental laboratory platform, developed by Massachusetts Institute of Technology (MIT), that enables parents to participate with their child in developmental psychological experiments from their own home (Scott & Schulz, 2017). By using LookIt instead of a traditional university laboratory, MB aims to alleviate parent's burden; hence, accomplish better accessibility of participation in developmental studies, and thereby enhancing heterogeneity and increasing sizes of samples (Zaadnoordijk et al., 2021). Based on previous studies (Scott et al., 2017; Scott & Schulz, 2017), it is expected that using online methods for infant studies could lower the threshold for parents to participate with their child, thus enabling both parents in rural areas and full-time working parents to participate.

The current study

The current study originated as a pilot for the MBAH project. This study explored the feasibility to examine visual attention of infants, and replicate previous findings, using online, remote webcam recordings at their own home. Visual attention in infants can be studied with multiple preferential looking designs. In this pilot study, two designs are evaluated as a proof of concept. First, a One-Stimulus (1Stim) design, in which one stimulus is presented at either the left or right side of the screen (Teller, 1979), is applied. Subsequently, a Two-Stimuli (2Stim) design, in which two stimuli are presented simultaneously on the left and right side of the screen (Fanz, 1965), is used.

The current study focusses on four goals. The first goal is to replicate previous findings (Cohen, 1969 in Cohen, 1972; Cohen, 1972; Courage et al., 2006; Shaddy & Colombo, 2004) using a different experimental method. Thus, it is examined which stimulus-type relatively attracts the most attention, by comparing looking-times from trials with different stimuli (MBAH, personal communication). However, in contrast to previous studies, this is explored using relatively small notebook or computer screens in home environments, instead of larger tv-screens that are

typically used in controlled lab settings. In the current study, 2 types of manipulations (movement and complexity; Figure 1) are examined using 4 bullseye stimuli (see Appendix B): Static Simple (SS), Static Complex (SC), Dynamic Simple (DS), and Dynamic Complex (DC). For movement, infants are expected to have longer looking-times at dynamic than at static stimuli (Courage et al., 2006; Shaddy & Colombo, 2004). For complexity, a larger number of components in a stimulus, such as more circles in a bullseye stimulus, is believed to increase the duration of looking-times (Cohen, 1969 in Cohen, 1972; Cohen, 1972). However, it is important to note that looking-times change as a function of age, in a u-shaped fashion rather than a linear fashion (Richards, 2010). Thus, a decrease in looking-times in general is found from 3 to 6 months of age (Colombo et al., in Richards, 2010; Courage et al., 2006; Shaddy & Colombo, 2004). Yet, at 6 to 12 months of age, infants show increased looking-times at Sesame Street stimuli and facial stimuli (Courage et al., 2006; Richards, 2010). However, for geometric patterns, the decrease is plateauing and no increase in looking-times was found (Courage et al., 2006; Richards, 2010). In sum, it is expected that infants look longer at dynamic stimuli than at static stimuli; yet, for complexity, no distinct difference is predicted.

The second goal is to examine, based on effect sizes, whether one of the designs enables infants to distinguish the stimuli more easily. In the 2Stim design, infants can detect distinctive features more easily because they are able to switch between the stimuli on the screen (Oakes et al., 2009), which could possibly reduce the demand on the visual short-term memory (Oakes & Ribar, 2005); thereby, enabling infants to encode more details about stimuli differences (Oakes & Ribar, 2005), allowing them to more easily increase looking-times at their preferred stimulus. Nevertheless, 6-months-old infants are able to note subtle differences and similarities between stimuli when presented in successive order (Oakes & Ribar, 2005), as is the case in the 1Stim design. In sum, no distinct advantage of either 1Stim design or 2Stim design can be determined

based on previous literature; yet, it is slightly more likely that larger differences in looking-times between stimuli are found in the 2Stim design.

The third goal is to determine whether one of the two preferential looking designs is better suited for an online, remote setting experiment with manual annotation of eye movements; therefore, differences between the designs with respect to reliability, by evaluating differences in inter-rater reliability scores, are assessed. It could be possible that inter-rater reliability scores are lower in the 2Stim design, due to potential difficulty to determine where the infants are looking at when two stimuli being presented on one relatively small screen, instead of two larger tv-screens as is common in lab settings. In contrast, in the 1Stim design, infants' looking behavior might be detected more distinctly, due to the large contrast of the stimulus on one side and the plain background on the other side of the screen. Hence, larger inter-rater reliability is expected in the 1Stim design, suggesting a better fit for manually annotated online, remote preferential looking experiments.

Not only the inter-rater reliability describes the reliability of the online, remote experimental method, but the fourth goal assesses the reliability of this method as well. Specifically, the possibility for observers to accurately, above chance level (Teller, 1979), conclude on which side of the screen the stimulus is presented in the 1Stim design based on the total looking-time by the infant, is examined. Based on previous studies using online preferential looking designs (i.e., Scott & Schulz, 2017; Semmelmann, 2017; Smith-Flores et al., 2021; Tran et al., 2017), this is expected to be possible.

In sum, in the current study it is expected that: (1) findings can be replicated, meaning that infants are expected to look longer at dynamic than at static stimuli in both 1Stim- and 2Stim design, but no significant difference in looking-times is expected with respect to complexity in both designs; (2) stimuli might be slightly better distinguishable by infants in the 2Stim design as shown by effect sizes; (3) the 1Stim design could be a better fit for manually annotated,

online, remote preferential looking experiments based on inter-rater reliability; and (4) it is possible to accurately determine the spatial location of the stimuli in the 1Stim design above chance level.

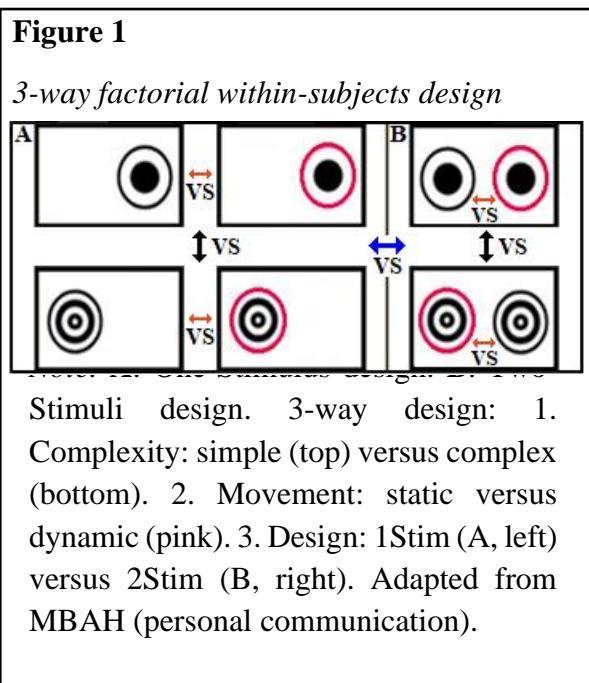
Methods

Participants

The participants in this study were infants aged between 8 months and 0 days and 10 months and 31 days. All infants were born full term, i.e., after 37 weeks of gestation. Data from 10 infants were excluded due to technical issues, such as parents not being able to upload the webcam recording ($n = 4$), no sound in the webcam recording ($n = 5$), or infant making loud noises masking the sound of the experiment ($n = 1$). Hence, in total there were 16 participants, (9 female), with a mean age of 9 months and 12 days ($M_{age_in_days} = 286.75$; $SD_{age_in_days} = 17.98$). Ethical approval for the study was obtained from the Faculty Ethics Assessment Committee of Humanities from Utrecht University. All infants participated with the informed consent of their parent(s).

Design

The 3-way factorial within-subjects design (Figure 1), started with a brief baseline event, in which the Static Simple (SS) stimulus was presented for 2 seconds on either side of the screen repeatedly, ending with an attention grabber in the center of the screen. This provided the coders an image of where the infant would look when stimuli were presented on the screen. Subsequently, the 1Stim design



was presented, followed by the 2Stim design. In the 1Stim design, infants saw all four possible bullseye stimuli (see Appendix B) presented four times, alternating between the left and right side of the screen. Thereafter, in the 2Stim design, either two complex or two simple stimuli were presented on the screen simultaneously; hence, the level of complexity of stimuli was matched within trials. The level of movement, i.e., static or dynamic, differed between stimuli within trials.

There were two versions of the experiment counterbalancing the order of the presentation of stimuli within the 1Stim- and 2Stim designs, to ensure blind coding by the experimenters. Participants were randomly assigned to either of the versions; $n = 8$ in both versions.

Materials

Due to the European General Data Protection Regulation (GDPR) it is, currently, impossible to use LookIt in the Netherlands; hence, the experiment was built in-home using JsPsych and JavaScript, and ran on servers hosted by Utrecht University. All data were collected and stored on the university servers. Moreover, the experiment script was built open source, meaning that other experimenters could eventually use and adjust the experiment by downloading it from GitHub.¹ The visual stimuli (see Appendix B) used in the experiment were created by the MBAH Consortium (MBAH, personal communication), which we adjusted to 200 by 200 pixels and exported as .mp4 file. All visual stimuli were presented for 12 seconds accompanied by the same musical sound (a musical chimes) after the pilot-trials showed that the suggested 20-second-trials (MBAH, personal communication) felt rather long. Visual attention grabbers (a rotating alarm clock in the 1Stim design, and a green spiral in the 2Stim design) were presented for 2.5 seconds, simultaneously with a different musical chimes. The musical sounds helped focusing infants' attention to the screen, in addition to enabling researchers to

¹ <https://github.com/Uil-OTS-labs/jspysch-cam-rec>

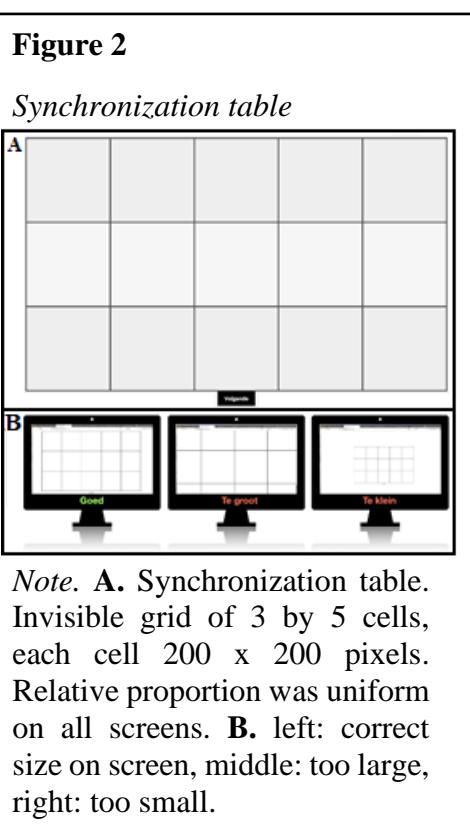
distinguish the trials from the attention grabbers, to determine trial durations, and to determine the duration of the experimental task in general.

Apparatus

Parents were provided with a step-by-step instruction for setting up and running the experiment (see Appendix A). For instance, they were asked to adjust the volume of their computer to their preference, to place their child in a highchair around 60 centimeters away from the screen, and to confirm that the webcam was placed at the center above the screen, at the height of the child's face, to ensure eye movements were visible.

The size of the notebook or computer screen could vary per participant; however, using a synchronization table (Figure 2), it was ensured that the experiment was shown on the screen as large as possible. The stimuli were presented on an invisible grid of 3 by 5 cells on the screen, and each cell was 200 x 200 pixels. Hence the grid was 1000 pixels wide and 600 pixels high. Stimuli were presented in the leftmost cell and the rightmost cell of the middle row. However, due to differences in screen sizes, parents were asked to either zoom in or zoom out, so that the test grid fitted their screen perfectly. As a consequence, the actual size of the pixels changed.

Nonetheless, the relative dimension of the stimulus display remained the same. In other words, there were always 3 cells between the two stimuli presentation locations, which equals the size



of 3 stimuli. Due to unknown and differing screen sizes, visual angles associated with the stimulus displays cannot be calculated for this experiment.²

Procedure

Data were collected from infants who had been registered in the UiL OTS Babylab database by their parent(s). In a telephone conversation the experimenter provided parents with information on, for instance, the use of their webcam, the stimuli, the duration of the experiment, and the reimbursement being a children's book. After the telephone conversation, participating parents received an email with a personal code, the link to the experiment, and the information letter attached (see Appendix C). In their own time parents could start the experiment, first receiving instructions (see Appendix A) followed by the experimental task for their child. From the start to finish, the duration of the experiment, including the setup, was approximately 20 minutes for parents, of which their child was actively involved for approximately 5 minutes. The personal code which parents received through email was used to link the raw webcam-recordings to the looking-time data. Additionally, this personal code was entered as participant number into the Babylab database. This arrangement allows us to delete webcam recordings on parents' request without losing critical data.

Video annotation

² Visual angle influences the perceived speed of motion of dynamic stimuli (Kaufmann, 1995; MBAH, personal communication). Hence, depending on screen size and the exact distance of the infant to the screen, the speed of the dynamic stimuli lay between 5 deg/sec and 10 deg/sec (MBAH, personal communication).

Prior to data analysis the webcam recordings of infants' looking behavior were manually annotated off-line using ELAN, version 6.0 (Sloetjes & Wittenburg, 2008). The experimenter and a second observer were blind to the stimuli that were presented on the screen, merely the duration of the stimuli was known based on the cooccurring sounds. The second observer annotated a third of the webcam recordings to realize inter-rater reliability.

In order to analyze the looking-times as coded using ELAN, average total looking-times were used. The average total looking-times were calculated by summing the times spent looking at a stimulus from the onset to the end of the trial, and dividing it by the 4 times a stimulus was presented (provided that no data was missing). When data from a trial was missing completely (e.g. due to distraction of the child by their environment), that trial was excluded from the calculation.

Analysis

IBM SPSS Statistics 26 (IBM Corp., 2017) was used to perform most statistical analyses; all with an alpha level of .05 to determine significance and all analyses were two-tailed. Two Repeated Measures Analysis of Variance (ANOVA) were conducted to determine whether we could replicate findings (first goal), thus, whether there were significant differences in average looking-times between stimuli per trial type. All assumptions were met. In addition, to examine differences in looking-times for complexity in the 2Stim design, average proportion looking-times were computed³ and a non-parametric Wilcoxon Signed Rank test was performed, due to non-normality of average proportion looking-times and difference scores.

³ Average proportion looking-times in the 2Stim design were computed using total looking-times per stimulus type. For example, proportion static simple = [total Static Simple looking-time]/[total Static Simple looking-time + total Dynamic Simple looking-time].

Furthermore, to test whether infants can distinguish stimuli more clearly in either of the two designs (second goal), the partial eta squared effect sizes of the Repeated Measures ANOVAs for the 1Stim and 2Stim design for movement were compared.⁴

Additionally, to examine whether there is a difference in inter-rater reliability in favor of the 1Stim design (third goal), reliability analyses were performed using total annotated looking-times, for both the 1Stim design and 2Stim design, individual webcam recordings, and accuracy of spatial locations of stimuli.

Lastly, to assess whether the observer's coding could accurately, above chance level, reveal on which side the stimuli in the 1Stim design were presented (fourth goal), a permutation test with 100,000 bootstraps, for a *p*-value of .05 in 256 trials (16 infants x 16 1Stim trials), was performed in MATLAB (MATLAB, 2021). Subsequently, using Microsoft Excel, the percentage of inaccurate and accurate decisions by observers were calculated.

Results

Replication

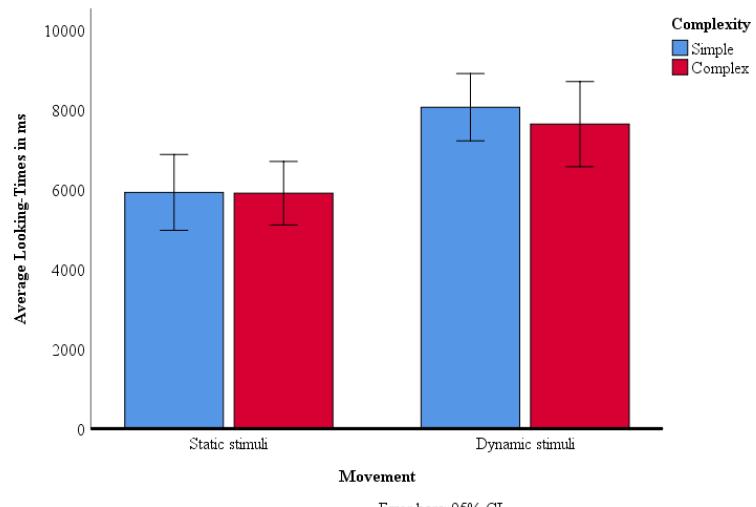
One-Stimulus Design

⁴ Effect sizes for complexity could not be compared, since stimuli in the 2Stim design were matched on the level of complexity within trials, and only proportion looking-times could be compared between trials. In addition, this also led to the inability to use Cohen's *d* effect sizes, which was planned originally.

A 2 x 2 Repeated Measures ANOVA (Figure 3; Table 1) was used to investigate preferential looking behavior of infants in the 1Stim design. A significant main effect for movement was obtained $F(1, 15) = 183.78, p < .001$, partial $\eta^2 = .93$. However, no significant main effect for complexity was found $F(1, 15) = 0.45, p = .513$, partial $\eta^2 = .03$. In addition, no

Figure 3

Repeated Measures ANOVA for Average Looking-Times in ms for the One-Stimulus Design



Note. 2 x 2 Repeated Measures ANOVA, with movement (static and dynamic) and complexity (simple and complex).

significant interaction between movement and complexity was found $F(1, 15) = 0.74, p = .404$, partial $\eta^2 = .05$. Examination of the means indicated that infants looked longer at the Dynamic Simple ($M = 8049.63, SD = 1581.04$) and Dynamic Complex ($M = 7627.88, SD = 2000.13$) stimuli, than at the Static Simple ($M = 5917.69, SD = 1783.79$) and Static Complex ($M = 5898.00, SD = 1489.59$) stimuli. In sum, for movement, infants looked significantly longer at the dynamic stimuli than at the static stimuli, both for the simple and complex stimuli. Yet, no significant difference in looking-times was found with respect to complexity, in the 1Stim design.

Two-stimuli design

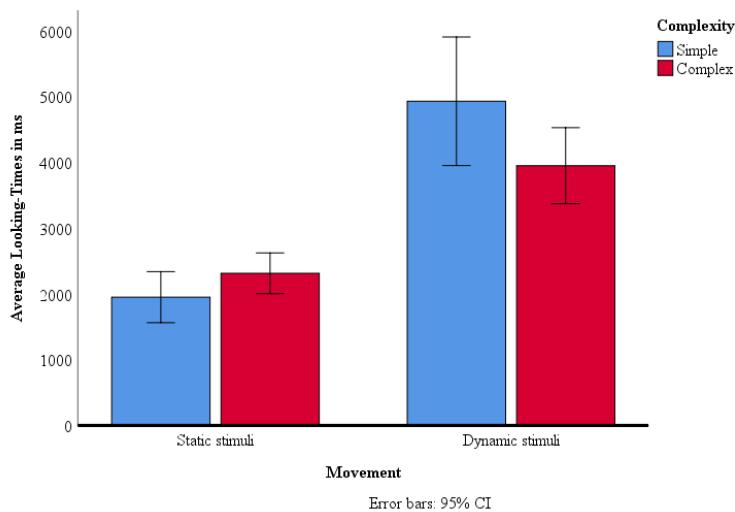
Movement. A 2 x 2

Repeated Measures ANOVA (Figure 4; Table 2) was used to investigate preferential looking behavior of infants for movement in the 2Stim design.

A significant main effect for movement was obtained $F(1, 15) = 48.11, p < .001$, partial $\eta^2 = .76$. Examination of the means indicated that infants looked longer at the Dynamic Simple ($M = 4930.63, SD = 1833.86$)

Figure 4

Repeated Measures ANOVA for Average Looking-Times in ms for Movement in the Two-Stimuli Design

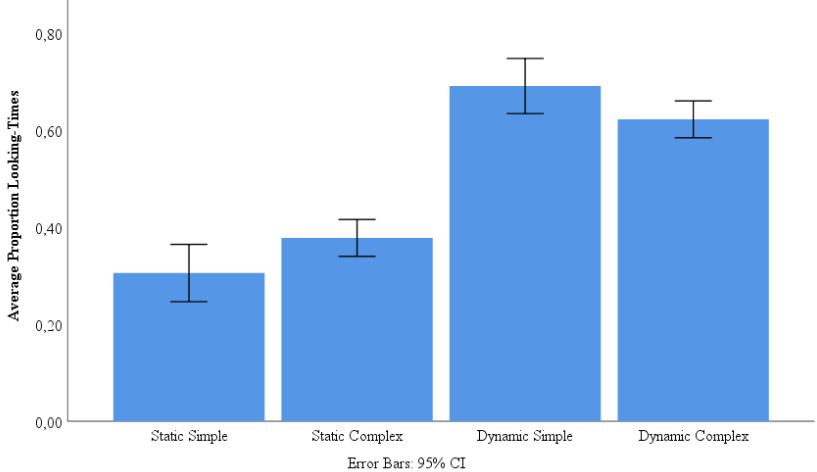


Note. 2 x 2 Repeated Measures ANOVA, with movement (static and dynamic) and complexity (simple and complex).

and Dynamic Complex ($M = 3951.63, SD = 1083.27$) stimuli, than at the Static Simple ($M = 1952.12, SD = 728.18$) and Static Complex ($M = 2315.56, SD = 583.00$) stimuli. In sum, for movement, infants looked significantly longer at the dynamic stimuli than at the static stimuli, both for the simple and complex stimuli.

Complexity. For complexity in the 2Stim design, a non-parametric Wilcoxon Signed Rank test (Table 3) using average proportion looking-times per stimulus type (Figure 5) was performed, which indicated that there is a significant difference in average proportion looking-times between the Static Simple ($Mdn = .302$) versus Static Complex ($Mdn = .358$) stimuli, with larger average proportion looking-times at the Static Complex stimuli, $T = 28.0, z = -2.07$ (corrected for ties), $N - \text{Ties} = 16, p = .039$, two tailed, $r = .52$, which can be considered a large effect (Cohen, 1988). Furthermore, for the Dynamic Simple ($Mdn = .698$) versus the Dynamic

Complex ($Mdn = .642$) stimuli, a significant difference in average proportion looking-times was found, $T = 28.0$, $z = -2.07$ (corrected for ties), $N - \text{Ties} = 16$, $p = .039$, two tailed, $r = .52$, which can be considered a large effect (Cohen, 1988). In sum, for complexity in the 2Stim

Figure 5*Average Proportion Looking-Times in the Two-Stimuli Design*

Note. Average proportion looking-times to indirectly examine complexity in the 2Stim design.

design, infants had larger average proportion looking-times at the Static Complex stimulus than at the Static Simple stimulus. Nevertheless, for dynamic stimuli, they had larger looking-times at the Dynamic Simple stimulus than at the Dynamic Complex stimulus.

Infants' preference

The second goal was to examine whether one of the designs enabled infants to distinguish the stimuli more clearly. The comparison between the effect sizes of the Repeated Measure ANOVAs for movement in the 1Stim design and 2Stim design are displayed in Table 4, showing that 93 % of variance in looking-time can be explained by movement in the 1Stim design, whereas 76 % of variance in looking-time can be explained by movement in the 2Stim design. Thus, the effect sizes for movement can be considered large (Cohen, 1988) in both 1Stim and 2Stim designs.

Inter-rater reliability

To assess suitability of the designs with respect to manual annotation of the webcam recordings, inter-rater reliability coefficients are compared. As can be seen from Table 5, an inter-rater reliability score of $r = .949$ was found for the 1Stim design, which can be interpreted as excellent (Koo & Li, 2016). However, for the 2Stim design, a somewhat smaller inter-rater reliability score of $r = .814$ was found, which can be interpreted as good (Koo & Li, 2016). In addition, accuracy scores were identical, resulting in an inter-rater reliability score of $r = 1.00$. In sum, all inter-rater reliability scores are high; yet, scores for the 1Stim design are higher than for the 2Stim design.

Accuracy

The fourth goal of this study was to examine whether it is possible to accurately determine the spatial location of the stimuli in the 1Stim design above chance level, using infants' looking-times. The permutation test showed that the chance level for accuracy for the 256 1Stim trials was 55.1 %. Calculations in Excel showed that in 6 of the 256 trials, infants had longer looking-times at the opposite side of the screen than where the stimulus was presented. Thus, $6 * 100 / 256 = 2.34\%$ of the trials were inaccurate. Hence, in 97.66 % ($100 - 2.34 = 97.66\%$) of the 1Stim trials, the spatial location of the stimuli was accurately detected by the observers.

Discussion

The current study examined the feasibility of using online, remote webcam recordings as a method to study visual attention in infants through a visual preference paradigm, to provide a possible solution to the current absence of large sample sizes and heterogenic samples in infant studies. This contributes to the validity of future infant studies, since it could increase statistical power, statistical conclusion validity, and external validity in infant studies.

A notion of caution is due with respect to interpreting the current results, since this study only examined 16 participants, leading to low statistical power and external validity. A power analysis showed that, assuming large effects, 23 participants should have been analyzed; hence,

statistical findings in the current study might be invalid. Nevertheless, given the large effect sizes that were found, practical significance can be assumed (Sullivan & Feinn, 2012). Furthermore, due to the current study being a pilot sampling from the UiL OTS Babylab database, in which parents who expressed willingness to visit the babylab at the university are signed up, the likelihood of infants with a high SES is considerable. However, merely the possibility of the methodology was examined, in contrary to developmental trajectories. Thus, in the current paper, homogeneity of the sample might not have profoundly influenced conclusions. Additionally, the degrees of visual angle differed per participant, through differences in screen size and exact distance of the infants to the screen. Nonetheless, since infants from 1 month of age can perceive motion between 1.4 and 118 degrees of visual angle per second (Kaufmann, 1995; MBAH, personal communication), the motion of the dynamic stimuli (between 5 and 10 deg/sec) should be perceived by all participants. Moreover, since mainly the fact that the dynamic stimuli are not static is relevant for this study, instead of the explicit velocity of the dynamic stimuli, the differences in degrees of visual angle are not likely to profoundly influence the conclusions drawn in the current study.

Replication

The first goal of this pilot study was to replicate previous findings (Cohen, 1969 in Cohen, 1972; Cohen, 1972; Courage et al., 2006; Shaddy & Colombo, 2004) using a different experimental method.

Movement

In accordance with previous literature (Courage et al., 2006; Shaddy & Colombo, 2004), it was found that infants have longer looking-times at dynamic than at static stimuli, in both 1Stim design and 2Stim design. This could mean that, even when infants are in a distracting home environment, and the scale of the stimuli is notably smaller than in the lab, conducting a

preferential looking experiment could still be possible. Furthermore, even when infants did not look at the screen, which was reported as missing data, it did not purely occur due to distraction or loss of interest. Often, infants tried to seek contact with their parent and subsequently looking or even pointing at the computer screen. Hence, it could be possible that, instead of a loss of interest, infants might have been hyper focused on the experiment, wanting to share their visual attention with their parent, which could be reasonable since joint attention is continuing to develop at 9 months of age (Mundy & Newell, 2007; Scaife & Bruner, 1975; Vaughan et al., 2003).

Complexity

Additionally, it was hypothesized that infants between 8 and 10 months old do not differ in looking-time with respect to complexity. Congruently with previous studies (Colombo et al., 1999, in Richards, 2010), no significant effect for complexity was found in the 1Stim design. However, in the 2Stim design, infants did have significantly larger average proportion looking-times at the Static Complex stimulus than the Static Simple stimulus, which could be explained by the theory of Cohen (1969 in Cohen, 1972; Cohen, 1972) stating that a larger number of components in a stimulus increases the duration of looking-times. Nevertheless, for dynamic stimuli, infants seemed to prefer looking at the Dynamic Simple stimulus over the Dynamic Complex stimulus, as shown by the significant difference in average proportion looking-times. The distinction between these stimuli in the 2Stim design could possibly be explained by the Goldilocks effect (Kidd et al., 2012). The Goldilocks effect proposes that “infants avoid spending time examining stimuli that are either too simple (highly predictable) or too complex (highly unexpected)” (Kidd et al., 2012, p.1). Hence, infants focus their attention to stimuli that are interesting enough to look at, but not too complex to be understood (Kidd et al., 2012). When looking at the results in the current study, it is likely that the static stimuli are experienced as too simple or too predictable in both designs, resulting in shorter looking-times. On the

contrary, in general, the dynamic stimuli seem just complex enough for infants to focus their attention to. However, when two complex stimuli (one static, one dynamic) are presented simultaneously on the screen, results showed that infants spend less time looking at the dynamic stimulus, as opposed to when two simple stimuli (one static, one dynamic) are presented on the screen. In other words, the Dynamic Complex stimulus might be slightly too complex when there is another source of information on the screen; whereas the Dynamic Simple stimulus seemed to be just complex enough in that situation. Hence, a significant difference was found when comparing the looking-times at the Dynamic Simple and Dynamic Complex stimuli in the 2Stim design. Nevertheless, do note that the sample in the current study is small, thus, the statistical power is low; hence, the MBAH project, which is likely to have a large and diverse sample size, will provide the clarity that is necessary.

Infants' preference

With the second goal, it was intended to examine whether one of the designs enabled infants to distinguish the stimuli more easily. Originally, Paired Samples *t* tests, thus Cohen's *d* effect sizes, were planned to determine the difference between looking-time means. However, due to the current experimental design, the levels of complexity could only be examined indirectly in the 2Stim design since they are not presented on the screen simultaneously. As a result, this question could not be assessed with a Paired Samples *t* tests and thus Cohen's *d* effect sizes and, therefore, the question could not be answered completely. Nevertheless, partial eta squared effect sizes were compared. The results showed that 93 percent of the variance in the 1Stim design could be explained by movement, in contrast to the 2Stim design, where the variance explained by movement was 76 percent. Hence, it seems that the 1Stim design might be better suitable to examine the factor movement, since there is hardly any variance that is not explained by this factor. However, another manner to explain these scores is that infants might potentially be influenced greater by the factor complexity in the 2Stim design, since the variance explained

by movement is lower and significant differences for complexity were found using the proportions of looking-times. Thus, it could be plausible that when two stimuli are presented side-by-side, infants might be able to better comprehend multiple aspects and details of stimuli, than when stimuli are presented one-by-one on the screen, which was also seen in previous literature (Oakes et al., 2009; Oakes & Ribar, 2005). However, the order in which the designs were presented (1Stim followed by 2Stim) might have influenced the ability to focus on other details in stimuli as well, since infants were familiarized with the stimuli in the 2Stim design. Nonetheless, in order to provide more clearance on this question, further research should compare complexity directly in the 2Stim design.

Inter-rater reliability

In addition to exploring differences between designs with respect to infants' ability to distinguish stimuli more easily, the third goal of this study was to examine whether observers manually annotated the looking behavior of the infants more reliably in either of the designs. Due to the large contrast of the stimulus on one side and the plain background on the other side of the screen; it was suggested that the inter-rater reliability would be higher in the 1Stim design, since infants' looking behavior could be detected more distinctly. Results showed that, indeed, a higher inter-rater reliability score was found in the 1Stim design. Nevertheless, since both designs indicated good (2Stim; Koo & Li, 2016) and excellent (1Stim; Koo & Li, 2016) inter-rater reliability scores, both preferential looking design could be used for manually annotated online, remote preferential looking experiments, provided that findings hold with larger sample sizes.

Accuracy

Spatial location stimuli

Lastly, the fourth goal of this study assessed the possibility for observers to accurately, above chance level, conclude on which side of the screen the stimuli were presented in the 1Stim design based on the total looking-times of infants. Results showed that, indeed, spatial location of stimuli on either the left or right side of the screen in the 1Stim design, could be determined almost perfectly. Similar experiences are seen in other studies; however, often, highly attractive stimuli such as short clips from popular infant-tv-programs or pictures of infants' own parents' faces are used (Scott & Schulz, 2017; Semmelmann, 2017; Smith-Flores et al., 2021; Tran et al., 2017), which could attract attention more easily, possibly resulting in more stable looking behavior throughout the experiment. Nevertheless, since spatial location could be determined nearly excellent in the current study, it could probably be concluded that online, remote webcam recordings in home environments can be used in visual preference designs to study visual attention in infancy, even when stimuli are not highly attractive.

Trial duration

However, determining the duration of trials based on the musical chimes, recorded concurrently with the webcam video recordings, tuned out to be somewhat inaccurate. It was found that the encoded trial durations deviated from the 12 seconds presentation time of the stimuli ($M = 11412,38$, $SD = 151,17$). This difference could potentially be explained by a fadeout effect of around 500 ms at the end of the musical chimes that were used in the experiment. Due to this fadeout of volume, participants' webcams appeared to be unable to detect the last 500 ms of the chimes, causing the observed difference. Furthermore, the deviation between participants could originate from the different specifications of webcam's that were used. As a result, the total durations of the experiments also deviated, ranging between 05:35:616 (min:sec:ms), and 05:37:156, with a mean duration of 05:36:685. This difference could be problematic when timing is crucial, such as in language studies when words are presented alongside.

Another limitation of the dependency on the recorded musical chimes is that when infants made loud noises, the sounds of the stimuli were masked, and could hardly or no longer be distinguished. Additionally, in several webcam recordings no audio was recoded, as the local webcam recorder muted the laptop's own sounds. Since 16 out of 26 webcam recordings could be analyzed, a margin of error of approximately 38 percent would be convenient, which is roughly similar to findings by Scott and Schulz (2017) using LookIt. Due to these issues, exclusively using musical chimes to determine trial duration cannot be recommended as preferred method. Instead, parents could be asked to place a mirror behind the infant to observe both infant and the screen simultaneously in the same webcam recording; however, this does add an extra participation criterium and would cost parents more time to setup. Therefore, it could be recommended to create a screen recording overlaying the webcam recording. In sum, encoded trial durations deviated around 500 ms from the 12 second stimulus presentation, probably due to a fadeout effect of around 500 ms in the musical chimes that were used, and the inability of webcams to record this. Furthermore, due to differences in specifications of webcams, thus the volume threshold for audio recording, deviation between participants occurred. A possible solution for this problem, with the least burden for parents, is to create a screen recording overlying the webcam recording.

Conclusion

In conclusion, this experimental pilot study sheds light on the possibility of online, remote infant testing in their home environment. In accordance with the hypotheses, 8- to 10-months-old infants tend to look longer to dynamic stimuli, as opposed to static stimuli. Moreover, as expected, for complexity, infants showed no strong preference for either simple or complex stimuli in the one-stimulus design; however, in the two-stimuli design infants looked, proportionally, significantly longer at the Static Complex stimulus and Dynamic Simple stimulus. Furthermore, as expected, this study shows that inter-relater reliability of looking

behavior is highest in the One-Stimulus design but both designs seem well suited. In addition, spatial location can be accurately determined by observers. Thus, it seems that visual attention can be examined using manually annotated webcam recordings of online, remote visual preference experiments; hence, if these findings hold after a larger and diverse sample is examined, future infant research could use these experiments to increase sample sizes and heterogeneity of samples, and thereby enhancing the validity of findings in infant studies.

References

- Allen, P., Bennett, K., & Heritage, B. (2014). *SPSS Statistics version 22: A practical guide*. (3 ed.) Cengage Learning Australia Pty Limited.
- Aslin, R. N. (1985). Oculomotor measures of visual development. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 391–417). Ablex Publishing.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). *Six solutions for more reliable infant research* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/u37fy>
- Clearfield, M. W., & Niman, L. C. (2012). SES affects infant cognitive flexibility. *Infant Behavior and Development*, 35(1), 29–35. <https://doi.org/10.1016/j.infbeh.2011.09.007>
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, 65(3), 145.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, L. B. (1972). Attention-Getting and Attention-Holding Processes of Infant Visual Preferences. *Child Development*, 43(3), 869–879. <https://doi.org/10.2307/1127638>
- Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, 52(1), 337–367.
- Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseman, L. J. (1991). Individual Differences in Infant Visual Attention: Are Short Lookers Faster Processors or Feature Processors? *Child Development*, 62(6), 1247–1257. <https://doi.org/10.1111/j.1467-8624.1991.tb01603.x>
- Courage, M. L., Reynolds, G. D., & Richards, J. E. (2006). Infants' Attention to Patterned Stimuli: Developmental Change From 3 to 12 Months of Age. *Child Development*, 77(3), 680–695. <https://doi.org/10.1111/j.1467-8624.2006.00897.x>
- DeBolt, M. C., Rhemtulla, M., & Oakes, L. M. (2020). Robust data and power in infant research: A case study of the effect of number of infants and number of trials in visual preference procedures. *Infancy*, 25(4), 393–419. <https://doi.org/10.1111/infa.12337>
- Fantz, R. L. (1964). Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Ones. *Science*, 146(3644), 668–670.
- Farah, M. J., Betancourt, L., Shera, D. M., Savage, J. H., Giannetta, J. M., Brodsky, N. L., Malmud, E. K., & Hurt, H. (2008). Environmental stimulation, parental nurturance and cognitive development in humans. *Developmental Science*, 11(5), 793–801. <https://doi.org/10.1111/j.1467-7687.2008.00688.x>

- Fernald, A. (2010). Getting beyond the “convenience sample” in research on early cognitive development. *The Behavioral and Brain Sciences*, 33(2–3), 91–92. <https://doi.org/10.1017/S0140525X10000294>
- Fraley, R. C., & Vazire, S. (2014). The N-Pact Factor: Evaluating the Quality of Empirical Journals with Respect to Sample Size and Statistical Power. *PLOS ONE*, 9(10), e109019. <https://doi.org/10.1371/journal.pone.0109019>
- Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., & Yurovsky, D. (2017). A Collaborative Approach to Infant Research: Promoting Reproducibility, Best Practices, and Theory-Building. *Infancy*, 22(4), 421–435. <https://doi.org/10.1111/infa.12182>
- García-Pérez, M. (2012). Statistical Conclusion Validity: Some Common Threats and Simple Remedies. *Frontiers in Psychology*, 3, 325. <https://doi.org/10.3389/fpsyg.2012.00325>
- Herrmann, D., & Guadagno, M. A. (1997). Memory Performance and Socio-Economic Status. *Applied Cognitive Psychology*, 11(2), 113–120. [https://doi.org/10.1002/\(SICI\)1099-0720\(199704\)11:2<113::AID-ACP424>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1099-0720(199704)11:2<113::AID-ACP424>3.0.CO;2-F)
- IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.
- Imafuku, M., Kawai, M., Niwa, F., Shinya, Y., & Myowa, M. (2019). Audiovisual speech perception and language acquisition in preterm infants: A longitudinal study. *Early Human Development*, 128, 93–100. <https://doi.org/10.1016/j.earlhumdev.2018.11.001>

- Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science*, 15(4), 463–473. <https://doi.org/10.1111/j.1467-7687.2012.1144.x>
- Jarosz, A. F., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*, 7(1). <https://doi.org/10.7771/1932-6246.1167>
- Kaufmann, F. (1995). Development of motion perception in early infancy. *European Journal of Pediatrics*, 154(S4), S48–S53. <https://doi.org/10.1007/BF02191506>
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS ONE*, 7(5), e36399. <https://doi.org/10.1371/journal.pone.0036399>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, 3(1), 24-52.
- MATLAB. (2021). *version 9.10.0 (R2021a)*. Natick, Massachusetts: The MathWorks Inc.
- Mundy, P., & Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, 16(5), 269–274. <https://doi.org/10.1111/j.1467-8721.2007.00518.x>
- 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>

Oakes, L. M., & Ribar, R. J. (2005). A Comparison of Infants' Categorization in Paired and Successive Presentation Familiarization Tasks. *Infancy*, 7(1), 85–98.

https://doi.org/10.1207/s15327078in0701_7

Pierce, K., Marinero, S., Hazin, R., McKenna, B., Barnes, C. C., & Malige, A. (2016). Eye Tracking Reveals Abnormal Visual Preference for Geometric Images as an Early Biomarker of an Autism Spectrum Disorder Subtype Associated With Increased Symptom Severity. *Biological Psychiatry*, 79(8), 657–666.

<https://doi.org/10.1016/j.biopsych.2015.03.032>

Rennels, J. L., Kayl, A. J., Langlois, J. H., Davis, R. E., & Orlewicz, M. (2016). Asymmetries in infants' attention toward and categorization of male faces: The potential role of experience. *Journal of Experimental Child Psychology*, 142, 137–157.

<https://doi.org/10.1016/j.jecp.2015.09.026>

Reynolds, G. D. (2015). Infant visual attention and object recognition. *Behavioural Brain Research*, 285, 34–43. <https://doi.org/10.1016/j.bbr.2015.01.015>

Richards, J. E. (2010). The development of attention to simple and complex visual stimuli in infants: Behavioral and psychophysiological measures. *Developmental Review*, 30(2), 203–219. <https://doi.org/10.1016/j.dr.2010.03.005>

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2004). Infant visual recognition memory. *Developmental Review*, 24(1), 74–100. <https://doi.org/10.1016/j.dr.2003.09.004>

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, 253(5489), 265-266. <https://doi.org/10.1038/253265a0>

- Scott, K., Chu, J., & Schulz, L. (2017). Lookit (Part 2): Assessing the Viability of Online Developmental Research, Results From Three Case Studies. *Open Mind*, 1(1), 15–29. https://doi.org/10.1162/OPMI_a_00001
- Scott, K., & Schulz, L. (2017). Lookit (Part 1): A New Online Platform for Developmental Research. *Open Mind*, 1(1), 4–14. https://doi.org/10.1162/OPMI_a_00002
- Semmelmann, K., Hönekopp, A., & Weigelt, S. (2017). Looking Tasks Online: Utilizing Webcams to Collect Video Data from Home. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01582>
- Shaddy, D. J., & Colombo, J. (2004). Developmental changes in infant attention to dynamic and static stimuli. *Infancy*, 355–365.
- Sloetjes, H., & Wittenburg, P. (2008). Annotation by category-ELAN and ISO DCR. *6th international Conference on Language Resources and Evaluation (LREC 2008)*.
- Smith-Flores, A. S., Perez, J., Zhang, M. H., & Feigenson, L. (2021). *Online measures of looking and learning in infancy*. PsyArXiv. <https://doi.org/10.31234/osf.io/tdbnh>
- Spelke, E. S. (1985). Preferential-looking methods as tools for the study of cognition in infancy. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 323–363). Ablex Publishing.
- Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—Or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, 4(3), 279–282. <https://doi.org/10.4300/JGME-D-12-00156.1>
- Tan, S. H. J., & Burnham, D. (2019). Auditory-Visual Speech Segmentation in Infants. *The 15th International Conference on Auditory-Visual Speech Processing*, 43–46. <https://doi.org/10.21437/AVSP.2019-9>

- Teller, D. Y. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development*, 2, 135–153.
[https://doi.org/10.1016/S0163-6383\(79\)80016-8](https://doi.org/10.1016/S0163-6383(79)80016-8)
- Tran, M., Cabral, L., Patel, R., & Cusack, R. (2017). Online recruitment and testing of infants with Mechanical Turk. *Journal of Experimental Child Psychology*, 156, 168–178.
<https://doi.org/10.1016/j.jecp.2016.12.003>
- Vacas, J., Antolí, A., Sánchez-Raya, A., Pérez-Dueñas, C., & Cuadrado, F. (2021). Visual preference for social vs. Non-social images in young children with autism spectrum disorders. An eye tracking study. *PLOS ONE*, 16(6), e0252795.
<https://doi.org/10.1371/journal.pone.0252795>
- Vaughan, A., Mundy, P., Block, J., Burnette, C., Delgado, C., Gomez, Y., Meyer, J., Neal, A. R., & Pomares, Y. (2003). Child, Caregiver, and Temperament Contributions to Infant Joint Attention. *Infancy*, 4(4), 603–616. https://doi.org/10.1207/S15327078IN0404_11
- Visser, I., Bergmann, C., Byers-Heinlein, K., Dal Ben, R., Duch, W., Forbes, S., Franchin, L., Frank, M., Geraci, A., & Hamlin, J. K. (2021). Improving the generalizability of infant psychological research: The ManyBabies model. *Behavioral and Brain Sciences*.
- Zaadnoordijk, L., Buckler, H., Cusack, R., Tsuji, S., & Bergmann, C. (2021). A Global Perspective on Testing Infants Online: Introducing ManyBabies-AtHome. *Frontiers in Psychology*, 12, 703234. <https://doi.org/10.3389/fpsyg.2021.703234>

Tables

Table 1

Repeated Measures ANOVA for Average Looking-Times in ms for the One-Stimulus Design

	<i>df</i>	<i>F</i>	<i>p</i>	partial η^2
Movement	1	183.78	< .001*	.93
Error (Movement)	15			
Complexity	1	0.45	.513	.03
Error (Complexity)	15			
Movement x Complexity	1	0.74	.404	.05
Error (Movement x Complexity)	15			

Note. $p = .05$, two tailed.

Table 2

*Repeated Measures ANOVA for Average Looking-Times in ms for Movement in the Two-
Stimuli Design*

	<i>df</i>	<i>F</i>	<i>p</i>	partial η^2
Movement	1	48.11	< .001*	.76
Error (Movement)	15			

Note. $p = .05$, two tailed.

Table 3

Wilcoxon Signed Ranks Test for Average Proportion Looking-Times for Complexity in the Two-Stimuli Design

		Negative		Positive Ranks					
		Ranks							
		<i>n</i>	<i>Mean</i>	<i>n</i>	<i>Mean</i>	<i>Median</i>	<i>t</i>	<i>z</i>	<i>p (r)</i>
			<i>Rank</i>		<i>Rank</i>				
Static		5	5.60	11	9.82		28.00	-2.068	.039* (0.52)
Simple						.302			
Complex						.358			
Dynamic		11	9.82	5	5.60		28.00	-2.068	.039* (0.52)
Simple						.698			
Complex						.563			

Note. * $p = .05$, two tailed.

Table 4

Effect Sizes for Stimuli Factor Movement in the One-Stimulus Design and Two-Stimuli Design Inferring Ability to Detect Differences and Exhibit Preference

	One-Stimulus design	Two-Stimuli design
	partial η^2	partial η^2
Movement	.93	.76

Note. Small = 0.01, medium = 0.06, large = 0.14 (Cohen, 1988).

Table 5*Inter-rater reliability scores*

Recording	<i>M</i>	Range	<i>r</i>	95% CI [LL, UL]
One-Stimulus design	3208.19	3157.91	.949	[.913, .973]
1	4540.02	136.34	.964	[.926, .982]
2	1461.17	21.78	.999	[.998, .999]
3	3121.59	38.00	.999	[.999, 1.00]
4	4237.08	173.84	.992	[.984, .996]
5	2681.11	23.91	.988	[.976, .994]
Two-Stimuli design	2672.76	2343.75	.815	[.650, .924]
1	3872.28	802.563	.939	[.784, .980]
2	2009.00	158.36	.990	[.968, .997]
3	2124.94	94.63	.994	[.983, .998]
4	2214.94	413.63	.966	[.892, .988]
5	2142.66	49.69	.993	[.981, .998]
Accuracy	23.80	0.00	1.00	

Note. M = item means. Range = item means range. *r* = intraclass correlation coefficient.

95% CI = intraclass confidence interval, LL = lower limit, UL = upper limit.

Appendix A

Instructions to parents (in Dutch)

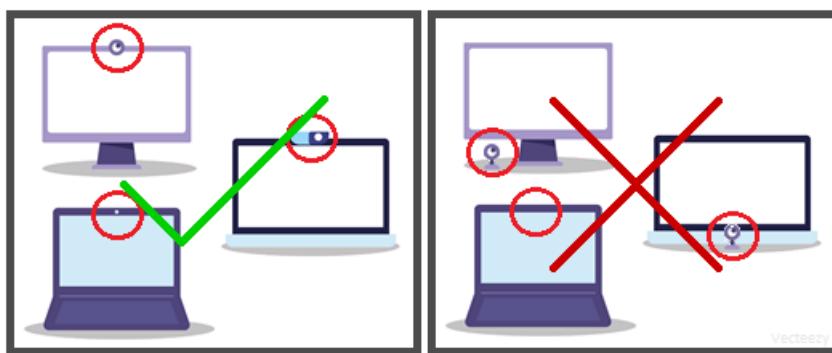
De meest wenselijke opstelling van uw laptop of computer ziet er uit zoals in de afbeeldingen. Volg de onderstaande stappen om deze opstelling te maken.



Of



1. De kinderstoel staat op ongeveer een (volwassen) armlengte afstand van het scherm, zodat er ongeveer **60 centimeter** tussen uw kind en het scherm zit.
2. De webcam is op **hoofdhoogte** van uw kind. Indien uw kind hoger zit, plaats dan bijvoorbeeld enkele boeken of een doos onder de laptop, zodat de webcam op hoofdhoogte van uw kind staat.



3. De webcam is **bovenin, in het midden van het scherm** geplaatst.
4. Er is **voldoende (dag)licht** in de omgeving zodat de ogen van uw kind goed zichtbaar zijn.
 - Indien dit niet het geval is plaatst u een **lamp hoog achter het scherm**, zodat er *geen schaduw* van uw scherm op het gezicht van uw kind valt.
 - Zorg ervoor dat er **niet te veel licht achter uw kind** vandaan komt om tegenlicht te voorkomen; zet uw kind dus idealiter niet met diens rug naar een raam.
5. Zorg voor zo **min mogelijk afleiding**

- Zet telefoon, tv, radio, en alle andere apparatuur wat kan storen **uit** of op **stil**, zodat er gedurende het onderzoek zo min mogelijk geluiden of andere afleiding is in de omgeving.
 - Zorg ervoor dat er geen speelgoed dichtbij of in het zicht van uw kind ligt.
6. Blijf **zelf achter uw kind** staan of zitten en kijk zo veel mogelijk naar beneden gedurende het onderzoek. U mag wel contact maken met uw kind als uw kind hiernaar opzoek is.
- Om het kijkgedrag van uw kind niet te beïnvloeden is het van belang dat u **achter uw kind blijft**, zodat uw kind maximale aandacht heeft voor het scherm en niet wordt afgeleid.
 - Verder vragen wij u om naar **beneden te kijken** (zodat u niet naar het scherm kijkt), **niet te praten** over het scherm of wat u hoort, en **niet te wijzen** naar het scherm.
 - Houd gedurende het onderzoek het handje van uw kind vast of **leg uw hand** op de **schouder** van uw kind, zodat uw kind weet dat u nog **aanwezig** bent zonder achterom te hoeven kijken.

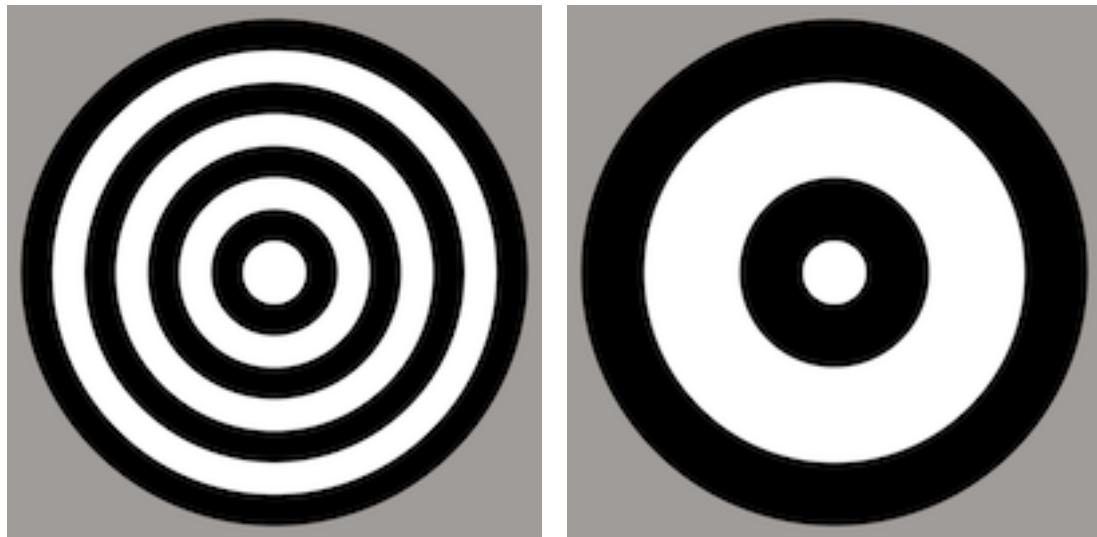
Als uw kind erg onrustig is kunt u uw kind **geruststellen**. Let wel op dat u **niet** met uw kind **praat** over of **wijst** naar het scherm.

Als uw kind niet langer meer in de kinderstoel wil zitten, dan kunt u het onderzoek stoppen.

Let op: Wij vragen u de webcamopname te stoppen maar de **opname wel te uploaden**. Alle kleine beetjes helpen. Ook is het belangrijk om na het taakje op ‘volgende’ te blijven klikken en klik tot slot op “einde” om het **onderzoek volledig af te ronden**.

Appendix B

Stimuli

Figure 6*Stimuli*

Note. Left: complex stimulus. Right: simple stimulus. Original size of 200 x 200 pixels. Presented size depended on scale of the experiment on the notebook or computer screen. Dynamic stimuli were continuously moving in a spiraling motion for 12 seconds. Adapted from MBAH (personal communication).

Appendix C

Information letter to parents (in Dutch)

**Informatie over deelname aan het
Online onderzoek naar het kijkgedrag van baby's**

1. Inleiding

Het huidige onderzoek is een wetenschappelijk onderzoek van Universiteit Utrecht, dat volledig online plaats vindt; uw kind doet dus mee aan het onderzoek vanuit uw eigen huis. Het onderzoek is getoetst door de Facultaire Ethische ToetsingsCommissie – Geesteswetenschappen van Universiteit Utrecht. Door deel te nemen aan het onderzoek gaat u akkoord met de voorwaarden zoals vermeld in de Toestemmingsverklaring (zie Bijlage 1 Toestemmingsverklaring, online in te vullen). Mocht u op dit punt besluiten niet deel te willen nemen aan het onderzoek, dan hoeft u verder niets te doen en kunt u de door u ontvangen mail als niet-verzonden beschouwen.

2. Wat is de achtergrond en het doel van het onderzoek?

Onderzoekers hebben al veel geleerd over het kijkgedrag van baby's. Dit soort onderzoek doen we tot nu toe altijd in de labs van de universiteit; echter, nu kijken we of we het onderzoek ook bij ouders thuis uit kunnen voeren, zonder dat de onderzoeker daar bij aanwezig is. Door baby's thuis, online, deel te laten nemen aan het onderzoek is het mogelijk om beter aan te sluiten bij hun ritme en dat van hun ouders.

In dit onderzoek bestuderen we de vraag of we kunnen bepalen naar welke afbeeldingen baby's het langst kijken, en of we kunnen onderscheiden of baby's naar links, rechts of helemaal niet op het scherm kijken.

3. Hoe wordt het onderzoek uitgevoerd?

Het onderzoek zal dus volledig online plaatsvinden. Dit vereist wel enige betrokkenheid van u als ouder. U heeft een internetlink ontvangen via de mail. Als u daarop klikt zult u worden doorgelied naar het onderzoek. Eerst zult u zelf enkele voorbereidingen moeten treffen, voordat u uw kind actief bij het onderzoek hoeft te betrekken. De voorbereidingen zullen gemiddeld niet meer dan 10 minuten duren (zie Bijlage 2 Ouderinstructies).

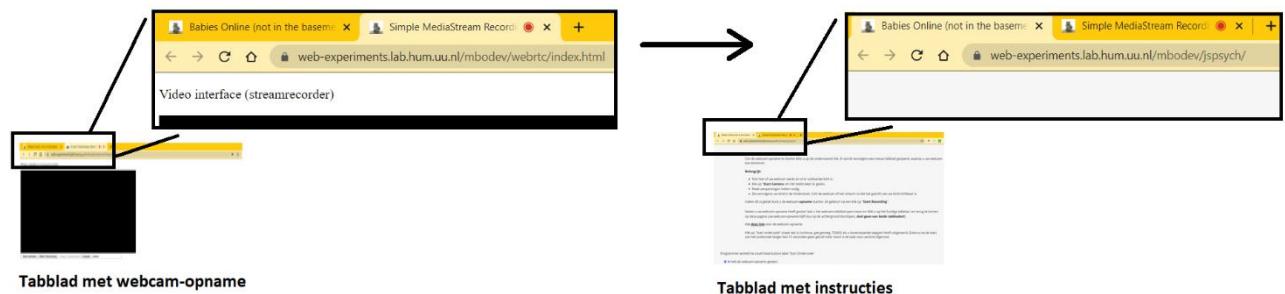
Voorbereiding

Nadat u op de link heeft geklikt krijgt u allereerst een mededeling over het gebruik van een laptop of desktop computer met een beveiligd (thuis)wifi netwerk, en de vraag uw browserscherm te maximaliseren. Vervolgens wordt er gevraagd om uw toestemming voor deelname aan het onderzoek, uw unieke persoonlijke code in te voeren die u heeft ontvangen in de mail, en krijgt u uitleg over de gewenste opstelling van uw laptop/computer en kinderstoel. Daarna vragen wij u uw geluid te testen, vragen we u of de lay-out van het onderzoek past op het beeldscherm, en vragen we u om uw webcam te testen.

Deelname van uw kind aan het taakje

Pas nadat u ziet dat de webcam werkt hoeft u uw kind in de kinderstoel plaatsen.

Vervolgens **start** u de **webcam-opname** en gaat u terug naar het tabblad met de instructies (er staan nu dus twee tabbladen open, zie afbeelding 1). Daar kunt u het taakje voor uw kind starten. Dit taakje duurt maximaal 10 minuten. Dit lijkt misschien lang, maar we verwachten niet dat baby's hun aandacht de hele tijd richten op het scherm.



Afbeelding 1. Wisseling van tabbladen. Geen van beide tabbladen afsluiten.

Tijdens het taakje krijgt uw kind stilstaande en bewegende aandacht trekende afbeeldingen te zien en zullen er geluiden te horen zijn. Om het kijkgedrag van uw kind te kunnen onderzoeken wordt er gedurende het hele taakje een video-opname gemaakt via uw webcam van uw kind. Ook wordt het geluid opgenomen; dit is ter controle voor de onderzoekers, zodat zij weten welk deel van het onderzoek er op dat moment gaande is. Wees u ervan bewust dat omgevingsgeluiden tijdens de webcam-opname dus ook worden opgenomen.

Na afloop van het taakje

Na afloop van het taakje stopt u de webcam en klikt u op '**uploaden**'. De beelden worden dan geupload naar de **beveiligde opslag** van de universiteit.

Daarna wordt u na afloop van het onderzoek gevraagd of u open staat voor het hergebruik van de webcam- en geluidsopnames door andere onderzoekers voor nieuw onderzoek en/of voor wetenschappelijke lezingen of lessen. U kunt hier wel of niet mee akkoord gaan. [Uw keuze hierin heeft geen enkele invloed op uw deelname aan dit huidige onderzoek](#) (zie Bijlage 1 Toestemmingsverklaring, online in te vullen, deel 2). Tot slot is het belangrijk om op "**einde**" te klikken. Pas nadat u op einde geklikt heeft wordt de data **geslagen** op de server van de universiteit en is het onderzoek afgerond. U kunt daarna de tabbladen sluiten.

4. Wat wordt er van u en uw kind verwacht?

Wij vragen u om de aanwijzingen zo goed mogelijk op te volgen. Daarnaast is het belangrijk dat u tijdens het onderzoek uw kind zo min mogelijk afleidt, omdat wij geïnteresseerd zijn in het natuurlijke kijkgedrag van baby's. Om het kijkgedrag van uw kind niet te beïnvloeden, raden wij aan dat u achter uw kind blijft staan of zitten, zodat uw kind maximale aandacht heeft voor het scherm en niet wordt afgeleid. Wel mag u naar uw kind glimlachen als uw kind contact met u zoekt. Het is prima als uw kind voelt dat u er bent, bijvoorbeeld door uw hand op de schouder van uw kind te leggen, of een vinger vast te houden. Daarnaast vragen wij u om niet te praten over het scherm of over wat u hoort, en niet

te wijzen naar het scherm. Het is dus belangrijk dat u de aandacht van uw kind niet stuurt (zie Bijlage 2 Ouderinstructies).

Van uw kind wordt verwacht dat hij/zij naar het scherm zal kijken. Kijkt uw kind zelden of niet naar het scherm, dan is ook dat bruikbare informatie over het kijkgedrag van baby's.

Het onderzoek zal voor uw kind maximaal 10 minuten duren. Totaal zult uzelf gemiddeld niet langer dan 30 minuten bezig zijn met het onderzoek. Deelname aan het onderzoek is eenmalig.

5. Wat zijn mogelijke voor- en nadelen van deelname aan dit onderzoek?

U en uw kind hebben zelf geen voordeel van deelname aan dit onderzoek. Echter, voor de toekomst kan het onderzoek wel nuttige gegevens opleveren. Zo kan het onderzoek naar mogelijkheden om baby's thuis, online, te testen helpen om tijdens babyonderzoek beter aan te sluiten bij het ritme van de baby en ouders. Daarnaast wordt er verwacht dat we een meer diverse groep baby's kunnen onderzoeken, die een betere afspiegeling vormen van de maatschappij.

Een mogelijk nadeel voor uw kind is dat uw kind tijdens het onderzoek maximaal 10 minuten lang redelijk stil zal zitten, er verveling op kan treden of uw kind vermoeid kan raken. Hoewel wij denken dat het onwaarschijnlijk is dat uw kind gestrest of overstuur zal raken van het onderzoek, willen we benadrukken dat het onderzoek in een dergelijke situatie door u gestopt kan worden als u denkt dat dit beter is voor uw kind.

6. Wat gebeurt er als uw kind niet (meer) mee wil werken tijdens het onderzoek?

Het kan zijn dat uw kind tijdens het onderzoek niet of zelden naar het scherm kijkt. Dit is geen enkel probleem en het blijft waardevolle informatie over het kijkgedrag van baby's. Echter, als uw kind te onrustig wordt en uit de kinderstoel wil mag u zelf de inschatting maken of het beter is om het onderzoek te stoppen, of dat het nog voortgezet kan worden na een paar kalmerende woorden. Wanneer uw kind echt niet meer mee wil doen, dan willen wij liever dat u het onderzoek stopt. Wij vinden het welzijn van het kind belangrijker dan ons onderzoek. Wij willen u wel vragen om de webcam-opname te **uploaden** en het **onderzoek af te ronden**, omdat iedere bijdrage bruikbaar is voor het onderzoek.

7. Vrijwillige deelname

Deelname is altijd vrijwillig. Als u toch besluit dat uw kind niet mee mag doen, dan hoeft u verder niets te doen. U hoeft niets te tekenen. U hoeft ook niet te zeggen waarom u uw kind niet wilt laten meedoen. Als uw kind wel meedoet, dan kunt u zich altijd bedenken en stoppen op ieder gewenst moment — ook tijdens het onderzoek. Bovendien kunt u nadat uw kind heeft meegedaan nog uw toestemming intrekken. Indien u daarvoor kiest, zullen de onderzoeksgegevens van uw kind niet meegenomen worden in de analyses. De onderzoeksgegevens van uw kind kunnen echter niet meer verwijderd worden als deze al zijn geanalyseerd, of als de onderzoeksgegevens niet meer tot uw kind te herleiden zijn, zie hieronder.

8. Wat gebeurt er met de verzamelde gegevens?

Uw persoonsgegevens en die van uw kind (naam, geslacht, leeftijd, geboortedatum, zwangerschapsduur, emailadres) worden beheerd door het Babylab van het Utrechts Instituut voor Linguïstiek (UiL OTS) van de Universiteit Utrecht. Mocht u deze persoonsgegevens willen corrigeren of laten verwijderen, dan kunt u dit doen door contact op te nemen met Desiree Capel, manager van het

Babylab (babylabutrecht@uu.nl). De persoonsgegevens zullen niet verstrekt worden aan anderen dan de onderzoekers die direct bij dit project betrokken zijn.

Tijdens deelname van uw kind aan het taakje worden gegevens verzameld en opgeslagen. De gegevens van uw kind worden opgeslagen en bewaard op een door de Universiteit Utrecht beveiligde server. In dit onderzoek worden zowel de webcam-opnames met geluid als de kijktijden apart van elkaar opgeslagen en krijgen ieder een unieke code.

Dit betekent dat de kijktijden ‘gecodeerd’ zijn en alleen via die unieke code naar uw kind te herleiden zijn. De sleutel waarmee gecodeerde gegevens naar u of uw kind persoonlijk te herleiden zijn, wordt zo spoedig mogelijk vernietigd.

De webcam-beelden zijn niet op deze manier te ‘coderen’, omdat uw kind zichtbaar moet blijven om de kijktijden te kunnen bepalen. De unieke persoonlijke code die wij aan de opnames koppelen zorgt ervoor dat wij uw opname kunnen terugvinden, mocht u deze willen laten verwijderen.

Let op: het is dus belangrijk dat u uw persoonlijke code goed bewaart. Alleen als u uw persoonlijke code aan ons door kunt geven kunnen wij de webcam-opname op uw verzoek verwijderen.

Wij zijn verplicht de ruwe onderzoeksgegevens, waaronder de webcam-opname met geluid waarin uw kind zicht- en hoorbaar is, en de gecodeerde kijktijden minimaal 10 jaar te bewaren. Daarvoor geeft u toestemming als uw kind meedoet aan dit onderzoek. Als u dat niet wilt, dan kan uw kind niet meedoen aan dit onderzoek.

De gecodeerde onderzoeksgegevens kunnen worden gedeeld met en/of worden hergebruikt door andere wetenschappers om eventueel andere onderzoeks vragen mee te beantwoorden (zie Bijlage 1 Toestemmingsverklaring, online in te vullen).

9. Is er een vergoeding wanneer u besluit om met uw kind aan dit onderzoek mee te doen?

Als dank voor uw en uw kinds deelname aan het onderzoek willen wij uw kind graag een cadeautje sturen. Om dit te kunnen doen zal er na afloop van het onderzoek gevraagd worden om uw persoonlijke code en adresgegevens te mailen naar babylabutrecht@uu.nl. Indien u geen cadeautje wilt ontvangen hoeft u uw adresgegevens niet te sturen, klik wel op “einde” om het onderzoek af te ronden. De ontvangen adresgegevens worden na verzending van het cadeautje verwijderd.

10. Goedkeuring van dit onderzoek

De Facultaire Ethische ToetsingsCommissie - Geesteswetenschappen (FETC-GW) heeft dit onderzoek goedgekeurd. Wanneer u een klacht wilt indienen over de procedure omtrent dit onderzoek, dan kunt u contact opnemen met de secretaris van de FETC-GW, e-mail: fetc-gw@uu.nl, of met de functionaris voor gegevensbescherming van de Universiteit Utrecht, e-mail: privacy@uu.nl

11. Meer informatie over dit onderzoek?

Mocht u voorafgaand, tijdens, of na het onderzoek vragen hebben, dan kunt u contact opnemen met de onderzoekers. U kunt op ieder gewenst moment mailen naar Maaike Jansen (m.h.i.jansen@uu.nl, babylabutrecht@uu.nl) of naar de eindverantwoordelijk onderzoeker dr. Caroline Junge (c.m.m.junge@uu.nl)

12. Bijlagen:

- Toestemmingsverklaring

Bijlage 1 Toestemmingsverklaring, online in te vullen

TOESTEMMINGSVERKLARING voor deelname aan:

Online onderzoek naar het kijkgedrag van baby's

Voorafgaand aan het onderzoek:

Ik, ouder of voogd van het kind dat deelneemt aan dit onderzoek, bevestig:

1. dat ik via de **informatiebrief** naar tevredenheid over het onderzoek ben ingelicht;
2. dat ik in de gelegenheid ben gesteld om **vragen** over het onderzoek te stellen en dat mijn eventuele vragen naar tevredenheid zijn beantwoord;
3. dat ik gelegenheid heb gehad om grondig **na te denken** over deelname van mijn kind aan het onderzoek;
4. dat ik uit **vrije wil** samen met mijn kind deelneem.

Ik, ouder of voogd van het kind dat deelneemt aan dit onderzoek, stem er mee in dat:

1. de verzamelde gegevens voor wetenschappelijke doelen worden verkregen en bewaard zoals in de **informatiebrief** vermeld staat;
2. de verzamelde, gecodeerde, onderzoeksgegevens door wetenschappers kunnen worden gedeeld en/of worden hergebruikt om eventueel andere onderzoeks vragen mee te beantwoorden;
3. er voor wetenschappelijke doeleinden **beeld- en geluidsopnamen** worden gemaakt. Deze opnamen zullen niet met andere onderzoekers worden gedeeld dan diegenen die direct betrokken zijn bij dit onderzoek, tenzij u hier uitdrukkelijk toestemming voor geeft aan het einde van het onderzoek.

Ik begrijp dat:

1. ik het recht heb om mijn **toestemming** voor het gebruik van de data van mijn kind **in te trekken** zoals vermeld staat in de informatiebrief

Door dit hokje aan te vinken verklaart u de Toestemmingsverklaring grondig te hebben gelezen en begrepen, en stemt u in met de deelname van uw kind aan het onderzoek.

Indien u geen toestemming geeft wordt het onderzoek beëindigd.

Na afloop van het onderzoek:

Verklaring omtrent hergebruik van gegevens

Uw keuze over onderstaande punten heeft geen enkele invloed op uw deelname aan het huidige onderzoek.

- (1) Gaat u er mee akkoord dat het verzamelde audio- en videomateriaal, dit is dus niet anoniem, van (u en) uw kind, gedeeld wordt met andere onderzoekers voor onderzoeksdoeleinden?
 Ja, daar ga ik mee akkoord. Nee, niet akkoord.
- (2) Soms worden beeld- en/of geluidsopnames ook getoond in wetenschappelijke lezingen of lessen. Gaat u ermee akkoord dat het gemaakte audio- en/of videomateriaal gebruikt wordt voor deze doeleinden?
 Ja, daar ga ik mee akkoord. Nee, niet akkoord.