Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Empty Looks or Paying Attention?

Exploring Infants' Visual Behavior during Encoding of an Elicited Imitation Task

Trine Sonne*

Osman S. Kingo*

Peter Krøjgaard*

*Center on Autobiographical Memory Research, Department of Psychology and Behavioral Sciences, Aarhus University.

Acknowledgements

This article was supported by the Danish National Research Foundation (DNRF89),Mind*Lab* and VELUX FONDEN (#10386). We would like to give a special thanks to Patricia Bauer for suggesting the idea of employing eye-tracking within this paradigm and for access to props from her lab. We would like to thank Bettina Damsgaard Hald and Sidsel Tornhøj Sørensen for their help booking the participants, coding and entering of data. Furthermore, we would like to thank Rasmus Jølberg for recording the videos, Omar Sinué Salgado Leyva for assisting with the pupil dilation data, and Jonathan Mark Koppel for providing us with constructive comments on the article. Finally, we would like to thank all participating infants and adults for making this study possible.

Corresponding Author:

Trine Sonne

Bartholins Allé 9, 1350-424

Center on Autobiographical Memory Research,

Department of Psychology and Behavioral Sciences, Aarhus University

DK-8000 Aarhus C, Denmark

Phone: +45 87 16 53 61, E-mail: trines@psy.au.dk

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Empty Looks or Paying Attention?

Exploring Infants' Visual Behavior during Encoding of an Elicited Imitation Task

Abstract

Recently there has been an increased interest in the relationship between looking time during encoding and subsequent memory performance in imitation tasks. Hitherto the results have been inconclusive: one line of research supporting the link between looking time and performance, another line finding no relation. The existing studies may however, have been restricted by using small samples, limited looking time measures, and short retention intervals. We here examined the relationship between the encoding process by means of looking time as well as pupil dilation (by means of eye-tracking technology) in sixty-eight 20-month-old infants participating in an elicited imitation task and their subsequent memory performance (at an encoding test and at a 2-week delayed recall test). Additional twenty-two infants provided baseline measures. Simple looking time (assessed as fixation duration) did not correlate consistently with subsequent memory measures. In some cases however, looking time correlated *negatively* with imitation scores. In contrast, positive correlations were found between pupil dilation and some of the memory measures, suggesting that pupil dilation may be a more sensitive tool compared to looking time measures.

Keywords: elicited imitation, eye-tracking, infant memory, pupil dilation, looking time.

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Empty Looks or Paying Attention?

Exploring Infants' Visual Behavior during Encoding of an Elicited Imitation Task

Introduction

The imitation paradigm is one of the most influential non-verbal tasks for assessing infant memory. The paradigm builds on the "monkey see, monkey do" principle taking advantage of the natural amusement infants find in imitation (e.g., Bauer, 2007; Meltzoff, 1988). In a typical imitation task the experimenter demonstrates an action sequence to the child, and then after a delay (lasting seconds to months) the child is asked to imitate the action sequence. Despite the non-verbal nature of this task, it has been acknowledged as a measure of explicit or declarative memory (e.g., Carver, Bauer, & Nelson, 2000; McDonough, Mandler, McKee, & Squire, 1995; Meltzoff, 1995). The results from numerous studies employing this paradigm have shown that with age infants encode information faster, remember more actions and sequences over longer delays, and become more proficient at employing different cues (for reviews see Bauer, 2007; Hayne, 2004).

Despite the well-documented developmental changes in infant memory capabilities we still only have limited knowledge regarding the memory *processes* behind these memory behaviors (Pathman & Bauer, 2013). This could be due to the fact that most of the experiments conducted within this paradigm so far primarily have focused on the *product* of a memory test (e.g., assessing the number of remembered actions and sequences). This gap in our knowledge has recently led researchers to employ methods allowing a more *process*-oriented focus, for instance by comparing electrophysiological measures of brain activity, such as ERP, with behavioral memory measures (e.g., Bauer, Wiebe, Carver, Waters, & Nelson, 2003; Bauer, Wiebe, Carver, Lukowski, Haight, Waters, & Nelson, 2006; Nordqvist, Rudner, Johansson, Lindgren, & Heimann, 2015), or by investigating looking time or visual behavior in relation to imitation skills (e.g., Elsner, Pfeifer, Parker, & Hauf, 2013; Óturai, Kolling, & Knopf, 2013). Considering the importance of the latter approach, Taylor and Herbert (2013a, p. 14) recently wrote that "Knowing how images and dynamic events are scanned by infants provides valuable information about what has been perceived and is, therefore, potentially available for encoding and subsequent recall".

The assumption that looking time during encoding relates to learning and subsequent memory is far from new. Classical theories and models concerning adult short-term or working memory acknowledge the importance of rehearsal for memory (e.g., the modal model of memory, Atkinson & Shiffrin, 1968; Baddeley, Eysenck, & Anderson, 2009). Within the infant literature, several paradigms for testing infant memory are based on the assumption that looking time is necessary for, as well as an indicator of, memory (e.g., the Visual Paired Comparison paradigm or Habituation tasks; for a review see Hayne, 2004). Accordingly, an infant's or a child's gaze has typically been assumed to reflect the focus of his or her attention (Elsner et al., 2013; Esseily & Fagard, 2013).

When considering the few studies that have been conducted focusing on visual behavior during the encoding of an imitation task, the field is, however, characterized by inconclusive results: One line of research stressing the importance of looking time for later memory, and the other showing that looking time may not always predict memory or imitation. In the following, studies from these two lines of research will be outlined. As will become evident the inconclusive results could potentially be explained by differences in procedures used as well as methodological limitations, since the studies combined are characterized by using small samples, limited looking time measures, as well as only one time-point imitation test. The aim of the present study was therefore to carefully examine the relationship between visual behavior and imitation performance by investigating 20-month-olds' encoding process during an imitation task and to relate this to performance at an encoding test and a delayed test by use of a more comprehensive methodological approach.

Looking time during encoding as predictor of imitation performance

Only few studies have explicitly documented a relationship between looking time during encoding and subsequent recall. Some of the evidence documenting a link between looking time and imitation performance comes from studies conducted by Vivanti and colleagues on children with autism and typically developing children (e.g., Vivanti & Dissanayake, 2014; Vivanti, Nadig, Ozonoff, & Rogers, 2008). By means of eye-tracking Vivanti et al. (2008) assessed the percentage of time a group of 8-15 year-old children with autism as well as typically developing children spent fixating on either the face of the experimenter, the actions performed, or outside of these areas in two imitation tasks (one involving meaningful actions, the other involving non-meaningful gestures). Imitation performance was, as expected, better in the typically developing children compared to the children with autism. Overall, both groups looked at the actions for the same amount of time. However, the duration of fixation to the action area was positively related to more accurate reproductions in the autism group for the non-meaningful gestures-to-be-imitated. Based on these results, Vivanti and colleagues (2008; Vivanti & Dissanayake, 2014) suggest that differences in attention during encoding are related to imitation performance in children with autism. Note, however, that the studies conducted by Vivanti and colleagues only used simple actions or gestures-to-be-imitated, and did not focus on the children's memory across a delay.

Another study supporting the claim that looking time during encoding is related to later imitation was conducted by Taylor and Herbert (2013b). They eye-tracked a group of 6-, 9-, and 12-month-old infants as well as an adult comparison group while the participants watched a videotaped deferred imitation task involving an experimenter modeling actions with a puppet. More specifically, Taylor and Herbert (2013b) focused on fixation duration at three different areas of interest: puppet, person, and background. The results revealed that there were significant differences in the looking time scores provided by the adult group compared to the infants. Overall, the adults looked longer at the demonstration compared to the infants, whereas no differences were found between the three groups of infants. Importantly, a difference in overall patterns of looking time was also found when dividing the infants into two groups based on whether they were capable of imitating the actions or not. The infants who imitated the actions focused more at the person than non-imitators, and they did not look as long at the background compared to non-imitators. Taylor and Herbert (2013b, p. 778) concluded by stating that differences in attentional focus were related to subsequent recall. One potential limitation to the study by Taylor and Herbert (2013b) seem to be that despite using three-step action sequences, only few infants were capable of imitating one or more of the actions (only 14 out of 44 infants). Thus, there was a floor-effect in the imitation results which may have made it difficult to assess and capture a 'true' effect of differences in looking time on the individual imitation scores. Furthermore, the main differences in results were found between infants and adults, and as such no clear pattern seems to be evident when solely focusing on infants. Investigating older infants could potentially surpass this limitation.

Despite the limitations of the abovementioned studies, the overall assumption that looking time is related to imitation performance receives indirect support by another common finding within the imitation paradigm, namely that multiple experiences with props to-be-remembered facilitate long-term memory regardless of the age of the child (e.g., Bauer, Hertsgaard, & Wewerka, 1995; Bauer, Wenner, Dropik, & Wewerka, 2000; Taylor & Herbert, 2013b). One mechanism behind the facilitative effect of multiple experiences may be that several presentations, all things equal, allows for prolonged looking at the props. As suggested by Taylor and Herbert (2013b, p. 771) "manipulations to the amount of learning, or to the cues available during learning, may enhance infant memory by altering what or how infants attend to the events". To summarize, classical theory as well as some empirical evidence provide support to the claim that looking time during encoding is related to subsequent memory performance.

Imitation studies suggesting that looking time does not predict performance

The results from a number of recent studies have, however, casted doubt on the assumption that looking time during encoding predicts subsequent memory performance in imitation tasks (Esseily & Fagard, 2013; Kirkorian, 2012; Kolling, Óturai, & Knopf, 2014; Óturai, Kolling, & Knopf, 2013).

In a recent study Esseily and Fagard (2013) investigated the significance of ostensive cues for infants' attention during an imitation task in order to explain the enhancing effect ostensive cues have been found to have on infants' imitation performance. A group of 10-month-olds were presented with a video of a person modeling a target action while being eye-tracked. Half of the infants watched a version with ostensive cues (i.e., where the experimenter would look directly at the infant and use infant-directed speech before performing the target action), whereas the other half watched a version without ostensive cues. The infants' performance was assessed in a baseline trial and in three subsequent trials each preceded by a demonstration. Despite succeeding in manipulating the gaze of the infants in the ostensive condition to the important parts of the demonstration, it was, somewhat surprisingly, the performance of the infants in the non-ostensive condition that improved the fastest (i.e., they improved performance following the first demonstration, whereas the infants in the ostensive condition did not improve until after the third demonstration). This suggests that enhanced looking time to important parts of an action demonstration does not necessarily lead to improved memory performance.

These findings were supported by a recent study in which Óturai et al. (2013) investigated 18-month-olds' gaze pattern in relation to their performance on a deferred imitation task. The infants were presented with videos of a model demonstrating either arbitrary or functional target actions while being eye-tracked. After a 30 minute delay the infants' imitation behavior was assessed. Overall, the infants' imitation performance was selective, that is, they were more likely to reproduce the functional target actions as opposed to the arbitrary target actions. However, no fixation time differences were found in relation to the two different types of actions, as well as no correlations between gaze behavior and overall imitation performance. They did, however, find an association between looking time and the acquisition of arbitrary target actions. Furthermore, the infants, who were "exact" imitators, focused more at the experimenter while the experimenter talked and had more saccades between the experimenter's face and the props as well as longer fixation duration at the props during the demonstration of the target actions compared to the more "selective" imitators. Óturai et al. (2013) concluded by stating that selective visual attention cannot explain selective imitation, and as such no consistent link between visual behavior and imitation was found. One potential limitation to this study was that eye-tracking and the imitation task were conducted in two separate rooms, potentially increasing the cognitive demands on the infants' memory capabilities.

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Summary

Although an understanding and a description of the relationship between looking time during encoding and later memory performance in an imitation task are theoretically warranted, only few studies have actually examined this relationship and the results from these studies are inconclusive. The relationship between looking time, attention, and memory therefore remains unclear (e.g., Taylor & Herbert, 2013a, 2013b). The inconclusive results could be stemming from methodological differences, differences in the investigated populations, or lack of statistical power. As should be evident from the introduction, several different age groups have been the target of investigation (i.e., from the range of 6 months to the age of 15 years). In addition, different populations have been investigated (infants, children, adults, as well as children with autism), and the samples have generally been relatively small (e.g. 20 infants, Oturai, Kolling, & Knopf, 2013). Furthermore, the studies seem to be characterized by differences in the imitation tasks with regard to (a) the used props, (b) whether gestures were employed, and (c) the number of steps involved in the to-be-remembered action sequences -- mainly caused by differences in overall aim of the studies. Finally, to the best of our knowledge none of the existing studies focusing on looking time in the imitation paradigm have examined the relationship between looking time during encoding of an event and memory performance at more than one time point and across extended retention intervals (e.g., employing an immediate as well as a delayed recall test). For comparison, in the 'conventional' imitation paradigm (i.e., using in vivo demonstrations and involving no attempt to assess the encoding process), researchers are not only interested in immediate recall, but consider infants' memory performance across delays as an important aspect of infants' memory (e.g., Bauer et al., 2000). In order to shed further light on the possible importance of the encoding process, studies involving memory assessments across extended delays are needed.

In an attempt of gaining more insight and a better understanding of visual behavior during the encoding of an elicited imitation task and its possible relation to later imitation scores, we here present a study with a design carefully aimed at examining just this. To the best of our knowledge, this is the first study to investigate the predictive value of visual behavior during encoding within the imitation paradigm employing *multiple* measures of visual behavior (looking time *and* pupil dilation), and their possible relationship to subsequent memory performance assessed at *two* time points (an immediate encoding test and a delayed test across a two-week retention interval). Furthermore, in order to avoid Type II errors we employed a large sample size (N = 90).

The encoding process in the present study was examined by means of eye-tracking technology. Eye-tracking is a non-invasive technology and has been suggested as a central tool for investigating the microstructures of infants' visual behavior (Aslin, 2007; Aslin & McMurray, 2004; Hepach, Vaish, & Tomasello, 2015). Instead of solely investigating looking time (assessed as fixation duration), our hope was that adding a measure of pupil dilation could help us understand the heretofore inconclusive results represented by the two lines of research described above. Pupil size data are automatically captured when employing eye-tracking technology, and have recently been described as a sensitive measure of the underlying cognitive processes inherent in a task (Aslin, 2012; Jackson, & Sirois, 2009; Sirois, & Jackson, 2012). Another advantage is that pupil size data are highly temporally sensitive and as such make it possible to investigate exactly at what point in the demonstration of the task that the change of the pupil size occurs (also referred to as phasic changes, Jackson & Sirois, 2009). According to Hepach et al. (2015, p. 2) this measure is "particularly interesting for developmental research because whereas eye movements can reflect the distribution of attention, changes in pupil dilation may provide a measure of the degree of psychological involvement in pre-verbal and just-verbal populations." Finally, Jackson and Sirois (2009) suggested that pupil dilation might be a superior measure, avoiding some of the interpretational problems often inherent in looking time data, while it at the same time can be directly compared to looking time.

The Present Study

In the present study 20-month-old infants participated in an elicited imitation task. In order to eye-track the infants' eye-movements each demonstration was video-filmed to be displayed on a monitor. Movies were preferred over live demonstrations since this allowed (a) the eye-tracking to be more controlled and accurate, (b) increased control of the stimuli in general, and (c) ensured that all of the infants received the same demonstration as well as the same amount of experimenter contact during encoding. All infants were presented with three different movies each illustrating a three-step action sequence performed using three unique props (see Fig. 1).

Following each demonstration the infants were asked to imitate this sequence (encoding test). Two weeks later the infants returned for an additional test (delayed test). Two weeks were chosen as retention in order to make sure to distinguish performance at encoding test and at the memory test, but also to induce some forgetting. Further, this retention interval is to our knowledge the longest retention that has been investigated focusing on visual behavior in relation to imitation scores. In the imitation tests we assessed both the number of produced actions and the number of correct sequences. Piloting suggested that the task was too difficult for the originally chosen age group of 16 month-olds, since they showed close to no imitation, probably due to a media deficit effect (see Barr, 2010; Kirkorian et al., 2015). We decided to test a group of 20-month-olds instead, since this age group would be less affected by the media deficit effect found to level off around 15 months of age (Barr, 2010; see also Barr, & Hayne, 1999).

Our overall expectation was that, despite the risk of a video deficit effect, the infants would still participate in the imitation task and hence show some sign of memory of the action sequences. Furthermore, we wanted to explore exactly how looking time (fixation duration) was related to later imitation scores. Despite the literature's inconclusive results with regard to looking time and memory performance, we expected that looking time across both demonstrations (assessed as fixation duration) would be positively correlated with the infants' imitation. We made this prediction based on the fact that imitation is dependent on observation and that some (but not all studies) have documented this relationship empirically.

We were also interested in the relationship between visual behavior and subsequent memory focusing on specific demonstrations and specific parts of the action sequence. First, we expected that looking time during the 1st demonstration would be especially important for later memory. Our reasoning was based on previous research showing that infants' looking preferences often shift during a test if the exposure to a given stimulus is increased, as well as the common finding that infants habituate to visual stimuli when presented with several demonstrations (DeLoache, 1976; Houston-Price & Nakai, 2004; Kingo & Krøjgaard, 2015; Snyder, Blank, & Marsolek, 2008). In order not to have the results affected by these possible confounds, we chose to primarily focus on visual behaviour during the 1st demonstration.

Furthermore, we speculated that looking time during the third and final action of the action sequences (during the 1st demonstration) would be particularly important, since this allows the infant to see the end-state or the goal of the action sequence (cf., Woodward, Sommerville, & Guajardo, 2001). Finally, we predicted that fixating specifically at the props of the specific actions during the 1st demonstration would increase the probability of encoding the to-be-remembered actions and sequences and for allowing the props to become effective retrieval cues (see Taylor & Herbert, 2013b, for a similar argument). Hence, we expected positive correlations between looking time within these areas of interest in the predefined time segments and later imitation scores. On the other hand we predicted that looking time at the experimenter during both demonstrations could potentially attenuate the infants' encoding of the to-be-remembered actions (cf., Esseily & Fagard, 2013; Topál, Gergely, Miklósi, Erdőhegyi, & Csibra, 2008), and we thus expected to find negative correlations between this area of interest and the imitation scores.

Besides general looking time, we also examined pupil dilation. There is general consensus that dilation of the pupil is not only to be taken as sign of arousal, but indeed can reflect load on memory, increased mental workload or even be an indicator of the intensity of the processing demands (e.g., Bailey & Iqbal, 2008; Hess & Polt, 1964; Just & Carpenter, 1993; Laeng, Sirois, & Gredebäck, 2012). A study by Gardner, Philip, and Radacy (1978) extended previous studies on adults by showing that dilation of the pupil during encoding as well as retrieval of information is also seen in children (7-9 years of age). Combined, this led us to believe that dilation of the pupil could be a valid indicator of the level of the processing during encoding. Based on the assumption that the final action of the different action sequences in particular would provide the infants with important information regarding the completion or the end-state of the action sequences, we expected that dilation of the infants' pupils during the third and final action of each prop (during the 1st demonstration) would be positively correlated with the infants' behaviour at the two memory tests.

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Method

Participants

Ninety healthy and full-term 20 month-olds (M_{age}= 20.19 months, *SD* = .15, 43 girls) were recruited from birth registers from the BLINDED National Board of Health. The infants came from BLINDED and the near surroundings and had an Apgar score of at least seven. They were predominantly Scandinavian Caucasian, living in families with middle to higher SES. All of the infants received a teddy for participating. Sixteen infants were excluded from the analysis: One due to experimental error, nine due to fussiness, and six infants were excluded because their total sampled looking-time to the stimuli were more than two standard deviations below the overall group mean. The infants were randomized into two groups: a test group (*n*=68) and a baseline group (*n*=22), allowing us to assess what the infants spontaneously would do with the props without initial demonstrations.

Materials

The Test Setting

The infants were seated in an adjustable chair in front of the eye-tracker booth with a parent nearby. Besides being adjustable for the eye-tracker setup, the chair could also be turned around, allowing the infants to swiftly change position from sitting in front of the TV screen (demonstration) to be positioned in front of the experimenter (encoding test) without leaving the chair. This arrangement allowed the encoding test to be run immediately after each prop had been demonstrated on the video screen. An adjustable table was used in order to make the height of the table fit the height of the chair. The props were presented on a plastic tray allowing the infants to be close to the props, reducing potential motor difficulties. Taken together these arrangements made it possible to test the infants' imitation in the same setting as the video demonstrations were given, as opposed to other studies using video-recordings for demonstration, in which baseline and testing were conducted in one room and demonstration including eye-tracking in another (e.g., Kolling et al., 2014; Óturai et al., 2013). In addition this arrangement reduced the delay from demonstration to encoding test to approximately 12s, thereby closely resembling the parallel length of delay in conventional in-vivo demonstrations. A HDD camera recorded the sessions for later re-coding.

Props

Three different three-step props were used: A Jumping Jack (JJ), a Shaker (SH), and a Spinner (SP) (see Fig. 2). All of the props had been designed for and used in a previous study from our lab (BLINDED), and were not otherwise commercially available. Although identical to the props used in the previous study, the Spinner had been created in a smaller version making it easier to handle for the infants.

Videos

In order to eye-track the infants during the demonstration of the actions sequences, each of the demonstrations was recorded on video. This also provided the infants with identical demonstrations. All videos depicted the same experimenter (the first author), who also did all of the testing. The videos were kept as simple and similar as possible, i.e., controlling the amount of eye-contact, the narration, clothing of the experimenter, and luminance. Furthermore the background displayed a completely grey area with no salient features. The duration of the videos varied from 45 seconds to 48 seconds. During each demonstration the narration was kept simple: The experimenter would start out by saying (in native language):*"You can play with all of these things. Look how I play with all of these things"*. During each demonstration the experimenter only said: *I do like this!* [Action 1], *Like this!*

[Action 2], *And look!* [Action 3]. *That is how I play with all of these things"*. Only the experimenter's upper body was visible on the video, indicating that the infants could see everything from the table and up (see Fig. 1 and 2).

Eye-tracker and Booth

In order to record the infants' eye-movements we used a Tobii X120 eye-tracker to assess fixations at 120 Hz (with 0.5° accuracy) on a 30" LCD widescreen. The total visual angle of the screen was 49.1° (width) x 31.9° (height), while the visual angle of the stimuli area was 42.2° (width) x 21.0° (height)°. The eyes of the infants were approximately 70 centimeters from the eye-tracker, and at level with the center of the screen. The Tobii Fixation Filter (default) was used. This filter detects quick changes in the gaze point signal using sliding averaging and can thus distinguish between fixations and saccades. Interpolation of samples was used when data samples were missing (e.g., during a blink). Before the actual data collection, a 5-point calibration (with four points in the corners and one in the middle of the screen) was conducted using Tobii Studio calibration for infants, allowing for post-calibration check of the actual calibration. The stimulus material was presented by the use of E-prime software. The light in the testing room was held constant during all trials in order to avoid creating differences in lumination, hence affecting the pupil dilation data.

Design and Procedure

The infants in the test group visited the laboratory twice. The elicited imitation task was chosen in order to include an encoding test immediately after the demonstration, allowing us to assess whether potential forgetting at the delayed test would be due to forgetting or inferior encoding of the given action sequence.

At the first visit the infants played a brief warm-up game with the experimenter allowing the infant to get familiar with the experimenter and to get used to the test setting. After this the infants watched the demonstration of each prop (repeated twice) on a TV-screen while being eye-tracked. To avoid potential confusion, the experimenter, who was also the person on the video, was never present in the same room as the infants while the demonstrations were presented on the screen. The experimenter thus left the parent and the infant alone in the testing room while playing the video demonstration. Parents had been carefully instructed not to interfere during testing. Immediately after the video demonstration for each individual prop (only separated by the approximately 12s it took for the experimenter to walk from the control room and enter the testing room), the infants participated in an encoding test for that prop (see Fig. 1). The experimenter thus re-entered the room, placed the props in front of the infant and asked (in native language): *"Can you show me how to play with all of these things – just like I did?"* Immediately after the posed question, the experimenter started the timer and handed the props to the infant. The trial lasted 90 seconds or until the infants had produced all of the actions, or engaged in non-task behavior (e.g., continuously pushing the tray away, or dropping the props on the ground after actively having played with the props). No corrections were carried out, indicating that our main interest was to see how much they had encoded, but not to correct them until they could produce the sequences. An advantage of not correcting the infants during the encoding test was that the results from the encoding test became directly comparable to the results obtained from the delayed recall test (see below). The order of the props was counterbalanced across infants using two specific fixed orders: Spinner, Shaker, Jumping Jack, or Jumping Jack, Shaker, Spinner.

At the second visit, about two weeks later (*M* = 14.03 days, *SD* = .85 days), the infants participated in the delayed test, where they were asked to re-enact without further demonstration or help, what they recalled from the original demonstration. The experimenter thus asked: “Can you show me how to play with all of these things?” The order of the props and time frame to perform the target sequences was identical to the first visit (see Fig. 1).

The infants in the control group were tested in order to get a baseline measure of what the infants spontaneously would do with the props without the initial demonstrations. With no preceding demonstrations, they only needed to visit the lab once. The narration was completely identical to the narration from the delayed test.

Coding and Data Reduction

Behavioral Memory Data

The infants' behavior was coded on-line by the experimenter. Twenty percent of the infants' behavior was re-coded by a second independent coder. Overall, the interrater agreement was high: 92%, Pearson's r =.97, Cohen's kappa = .92. Two different aspects of the infants' behavior were coded: 1) the number of correct target actions, and 2) the number of correct target sequences, i.e., the number of correctly ordered pairs of target actions:

1) Each correctly produced target action resulted in one point. Each target action could only be coded once. Since all of the props consisted of three-step sequences, the infants could receive a minimum score of 0 or a maximum score of 3 for each prop.
2) In order to code the target sequences we followed the manual employed by Bauer (1992): for each correctly ordered pairs of target actions the infants would receive one point. Following this, the infants could get a minimum score of 0 or a maximum score of 2 if all of the target actions were produced in the correct order.

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

These scores were calculated for each prop, as well as for all props combined.

Eye-tracking Data

Looking Time.

The following AoIs (Areas of Interest) were determined (see Fig. 3): *Overall AoI* (fixation duration at the whole movie), *Face AoI* (fixation duration at the experimenter's face), *Action 1 AoI* (one for each prop; fixation duration at the prop during the three seconds it took to demonstrate the first action), *Action 2 AoI* (same as before), and finally *Action 3 AoI* (same as before). The three action AoIs allowed us to compute fixation duration at the three props across all three actions combined, but also to assess fixation duration specifically at the third and final action of each prop.

The AoIs were drawn manually, and the sizes of the AoIs were made to fit potential smaller variations in position during the demonstration. Furthermore, due to the different sizes of the three props, the sizes of the AoIs varied slightly across props (see Fig. 3).

Besides focusing on fixation duration within specific AoIs, we also focused on time segments allowing us to assess total fixation duration (across both consecutive demonstrations) as well as fixations within the first of the two demonstrations only. Due to the fact that the demonstrations were of slightly different lengths (varying from 45 to 48 seconds), proportional looking time was calculated each time we assessed fixations within the demonstration(s). It was thus calculated how long the infants looked at the demonstration out of the possible maximum of looking time for each prop respectively.

Pupil Dilation.

Pupil size data were recorded simultaneously from the left and right eye at each sample (120 samples per second) and these two numbers were averaged and the following analyses are based on this mean. Following the procedure from other studies using pupil dilation (e.g., Geangu, Hauf, Bhardwaj, & Bentz, 2011), missing samples from one eye were replaced or interpolated with data from the other eye. Only infants with at least some data at the specified point of interest (Action 3 and baseline from the 1st demonstration– see below) were included in the analyses. For JJ this ended up with 63 infants, for SH 67 infants, and for SP 60 infants.

In order to obtain a measure of the degree of dilation we did a baseline-correction by subtracting a baseline measure from the point of interest (in this case Action 3 from each prop during the 1st demonstration). According to Jackson and Sirois (2009) a baseline diameter is usually thought of as a function of luminance. The baseline was identified as a point in time close to the point of interest (to ensure comparable display luminance), but when no actions were performed. This was done in order to avoid placing the baseline at a time point that could interfere with the infants' processing of the previous steps as well as the risk of documenting changes in pupil size simply due to changes in lumination (which could be the case in the very beginning of the movie by not providing the infants enough time to get used to the specific luminance of the movie). Baseline was thus calculated separately for each prop and was measured immediately after the completion of the first three actions, just before the second demonstration was about to begin. The duration of the baseline for each prop was 1 second. The baseline correction was thus done for each prop separately. An increase in pupil size (comparing baseline with the point of interest) would be accounted for as dilation (see Results section).

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Results and Discussion

The following results will be presented in the same order as they were considered, and after each set of results a brief discussion will be provided.

Behavioral Memory Data

Our first objective was to investigate whether the basic elicited imitation paradigm worked as expected. Preliminary analyses revealed that no order effects regarding the presentation of props were to be found (all *p*'s> .133), and in the following analyses data from the two orders were collapsed. We first considered the infants' production of target actions (see Fig. 4). An independent samples *t*-test revealed that the infants in the test group produced a significantly larger number of actions at the encoding test compared to the baseline group, $t(88) = 9.97$, $p<.001$, $r = .73$. Furthermore, the infants in the test group produced a significantly larger number of actions at the delayed test relative to the baseline group, $t(88) = 7.86$, $p<.001$, $r = .64$. Finally, a Paired-Samples *t*-test revealed that the infants in the test group also displayed forgetting across the two weeks, $t(67) = 2.10$, $p = .04$, $r = .25$.

Taken together, these analyses show, that (a) the infants in the test condition learned from the demonstrations, (b) that the infants remembered the demonstrated actions across the two-week delay, but (c) that they also showed some forgetting (see Fig. 4).

A similar pattern of results was found when looking at the target sequences, i.e., whether the infants reproduced the target actions in the correct order (see Fig. 4): An independent samples *t*-test revealed that the infants in the test group produced a significantly larger number of sequences at the encoding test compared to the baseline group, $t(88) = 11.71$, $p<.001$, $r = .78$. Furthermore, the infants in the test group produced a significantly larger number of sequences at the delayed test relative to the baseline group, $t(88) = 8.57$, $p<.001$, *r*

=.67. Finally, a Paired-Samples *t*-test revealed that the infants in the test group also displayed forgetting across the two weeks, $t(67) = 2.25$, $p = .03$, $r = .27$.

Taken together, the obtained pattern of results with regard to both actions and sequences was clear and systematic, as well as quite typical for infants in this age group when participating in elicited imitations tasks (see for instance Bauer & Mandler, 1989 for a live comparison), confirming that the basic design worked as expected. Furthermore, despite having the action sequences demonstrated on video, the infants performed very well, indicating that a possible video deficit effect (e.g., Barr, 2010) was negligible. This may be due, at least in part, to the fact that the experimental procedure used in the present study closely resembled the procedure typically employed in the elicited imitation paradigm (e.g., the encoding test took place in the same setting *and* immediately after the demonstrations) -- except for the fact that demonstrations were based on video recordings. Note that these results serve as a necessary pre-requisite to examine possible relationships between visual behavior and behavioral memory data.

Looking Time and Behavioral Memory Data

Overall looking Time

First, we analyzed whether there would be a relationship between overall looking time across the two demonstrations and the imitation scores from the two memory tests. Thus, we correlated fixations within the Overall AoI during the two demonstrations with the infants' imitation performance (target actions and sequences) at the encoding test and at the delayed test.

As evident in Table 1, surprisingly, all of the correlations (two-tailed) were in a negative direction, thus going against our expectations. Although we mainly found non-significant negative correlations, we did find a significant negative correlation between the infants' looking time at the two demonstrations and their target action performance at the encoding test.

Overall looking time during the 1st demonstration

As argued earlier, the 1st demonstration might provide us with cleaner data with regard to looking time. The correlations however, were still in a negative direction (although not significant; see Table 1).

Looking time at the three actions (accumulated) during the 1st demonstration

We then examined whether there were correlations between accumulated looking time for the three target actions combined for the three props (accumulating looking time at each of the action AoIs) during the 1st demonstration and the infants' imitation scores. The results again revealed negative, albeit non-significant, correlations (see Table 1).

Looking time at the final action (accumulated) during the 1st demonstration

As argued earlier we expected the third and final action of the action sequence to be essential for comprehending the whole action sequence, as well as in order for the props to become effective retrieval cues. We thus ran a correlation between looking time at Action 3 AoI (accumulated for all three props) during the 1st demonstration and the behavioral results. The results again revealed negative correlations (see Table 1). A significant negative correlation was found in relation to the infants' looking time at the final action and their target action performance at the encoding test.

Looking time at the final action during the 1st demonstration for each prop

Based on these results, we decided to investigate if the same pattern would be found when looking at the data prop-wise. Referring to the study by Elsner et al. (2013, p. 434) it might make more sense to look at the props separately due to the differences between the three tasks, such as the different objects, movements, and size of AoIs. We thus ran a correlation between looking time at Action 3 during the 1st demonstration and imitation performance for each prop respectively. However, as can be seen in Table 1, no clear patterns were to be found in these results.

Looking time at the experimenter during both demonstrations

Due to the fact that we primarily found negative correlations between looking time and the behavioral results in the previous analyses, our original expectations of a negative correlation between looking time across the two demonstrations and the behavioral measures that was *specific* to looking at the experimenters' face instead of at the actions, did not seem to have any additional explanatory power by itself. In order to control for the general negative influence of longer looks, we calculated proportional looking at the experimenter (dividing looking time at the Face AOI by overall looking time at the actions) during both demonstrations. This would then provide us with data on how much time the infants spent looking at the experimenter relative to the props. The results from these analyses yielded negative correlations, although none were significant (see Table 1).

Summarizing the looking time data

Overall, we found no strong relationship between looking time data and imitation scores. In some cases we even found significant *negative* correlations between looking time and imitation scores. It should be mentioned that although some of the correlations were significant, they were generally weak.

As previously described, the field investigating looking time during encoding in imitation tasks is characterized by inconclusive results. When considering the instances in our study in which significant *negative* correlations between looking time during encoding and subsequent imitation measures were obtained, the results are thought-provoking: On the face of it, they suggest that the *less* the infants looked during encoding, the better they fared. Obviously, this cannot be the whole story, since zero looking time during encoding can hardly lead to superior memory. In most cases, however, we simply found no correlations at all, which seem to indicate that our results were in accordance with the results obtained in more recent studies suggesting that looking time does not always predict memory (e.g., Kirkorian, 2012; Kolling et al, 2014; Óturai et al., 2013).

Another pattern that emerged from the results was that looking time seemed to be primarily related to performance at the encoding test, thus creating the possibility that other memory processes, such as consolidation, are more important for explaining the performance at the delayed test.

Habituation

Studies focusing on infant visual attention have shown that 'short lookers' might actually process or encode the material more efficiently than 'long lookers' (e.g., Chong et al., 2015; Colombo, 1995; Colombo, Mitchell, Coldren, & Freeseman, 1991; Courage, Howe, & Squires, 2004). In line with this, the habituation paradigm suggests that a decline in attention (drop in fixation duration) is merely a sign of an internal representation having been created (e.g., DeLoache, 1976). Following this argument it has been suggested that fast habituators are indeed more efficient at processing information (DeLoache, 1976).

Due to the fact that the infants in the present study watched the demonstration of each prop twice, they may have habituated to the demonstrations. Hence, we would expect a possible habituation measure to be correlated with the imitation scores. In order to test whether the infants habituated to the demonstrations, we first ran a Paired-Samples *t*-test comparing proportional looking time at the two demonstrations ($M_{\text{1st demo}} = .76$, $SD = .22$; $M_{\text{2nd demo}} = .69$, SD = .28), $t(67) = 5.02$, $p < .001$, $r = .52$, revealing a significant drop in looking time across the two demonstrations. The habituation measure was thus calculated as the drop in looking time from 1st to 2nd demonstration (i.e., computing a change score). We then ran a correlation between the habituation measure and the infants' imitation scores. The analysis revealed a positive correlation between the habituation measure and subsequent performed target actions and target sequences at the encoding test (see Table 1). Thus, the infants with the largest change score from the 1st to the 2nd demonstration, the better the performance at the encoding test.

At this point it was evidenced that the looking time measures correlated primarily with imitation scores at the encoding test. The results seem to suggest that other factors, *not* captured by measuring looking time, were affecting memory performance – especially after a two-week delay. Note that these generally negative results cannot be discarded by reference to atypical or poor memory performance. As specified earlier, the infants did indeed remember what they had seen across a two-week retention interval with approximately the same magnitude of recall as is typically found in similar studies (e.g., Bauer & Mandler, 1989; Kingo & Krøjgaard, 2013). Further, the overall negative results cannot be explained by lack of statistical power either, as the experimental cell is substantially larger (*n*=68) than is typically used in elicited imitation studies (see for instance Bauer, 1992; Bauer, Hertsgaard, & Wewerka, 1995). Next we turn to the pupil dilation data, since this measure might, as described earlier, be more sensitive to the underlying processes in the task.

Pupil Dilation

As reported in the coding section, all of the analyses on pupil dilation are based on baseline-corrected numbers. We found a significant dilation of the pupils at the final action of each sequence during the 1st demonstration, as evidenced by *t*-tests comparing our baseline measure and the size of the pupils during the final action in the 1st demonstration of each prop respectively (see Fig. 5 for *M*s and *SD*s). More specifically, for the Jumping Jack we found a significant difference between the baseline and the third step of the 1st demonstration illustrated by a *t*-test : $t(62) = 5.35$, $p<.001$, $r = .56$. The same was found for the Shaker: $t(66) = 2.54$, $p=.014$, $r = .30$, as well as for the Spinner: $t(59) = 6.58$, $p<.001$, $r = .65$.

As previously described, pupil dilation is often considered a sign of mental work load (e.g., Bailey & Iqbal, 2008; Hess & Polt, 1964). As can be seen in Table 1, we found that pupil dilation during the final action of the 1st demonstration was positively correlated with some of the behavioral measures. More specifically, significant positive correlations were found between pupil dilation at the final action of the Spinner and the Shaker and the sequences produced at the delayed test (see Table 1). In addition, there was a significant positive correlation between dilation at the Spinner and performance of target sequences at the encoding test.

The fact that we did not find significant correlations for pupil dilation during the 1st demonstration of the Jumping Jack and the behavioral measures may have a straightforward explanation. In general, we observed that the infants had a hard time completing the final action of the Jumping Jack on their own, probably due to the fact that the string they were supposed to pull may have been difficult to identify on video. Performance on this specific prop may thus have been affected by a greater degree of 'perceptual impoverishment' (see e.g., Kirkorian et al., 2015) due to the demonstration being presented on video.

However, focusing on the Spinner and the Shaker we obtained significant results in relation to the delayed test. The fact that our significant results were primarily confined to sequences at the delayed test, whereas the significant associations (albeit negative) we attained between looking time and memory performance were in relation to actions at the initial encoding test, suggests that pupil dilation captures other parts of the memory process than looking time measures. Furthermore, the fact that we found positive correlations with the use of pupil dilation and not in relation to looking time further questions the predictive value of simple looking time (assessed as fixation duration) within the elicited imitation paradigm. Additionally, we analysed whether any correlations were to be found regarding pupil dilation and looking time data (both taken from third step of the 1st iteration), and no correlations were found (all *p*'s >.49.)

General Discussion

In this study we investigated possible relations between visual behavior during encoding and imitation. Relative to the few existing studies exploring this relationship, the present study employed a methodologically more comprehensive approach by expanding both the measures on visual behavior as well as the number of dependent variables. By assessing visual behavior by means of both looking time and pupil dilation and by using both an immediate encoding test as well as a delayed test in a large data set, we hoped to obtain a more elabortated understanding of possible ways in which visual behavior could predict later

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

imitation. Overall, our results suggest that the link between visual behavior during encoding and imitation is highly complex and not as simple as predicted.

Surprisingly, the relationship between visual behavior and imitation largely depended on the kind of measures used. When focusing on looking time during encoding and the relation to subsequent imitation scores, we mainly found no correlations, and the few significant correlations we found were (to our surprise) negative and related to the encoding test. Conversely, when focusing on pupil dilation, we primarily found positive correlations (although not all significant), mainly for the delayed test. These results suggest that different methods for investigating visual behavior may tap into different aspects of the memory process. Thus, whereas simple looking time measures may primarily, if at all, provide information regarding the short-term consequences of the encoding process, as evidenced in the results obtained in the encoding tests, the pupil dilation measures may predominantly inform us about whether the consolidation of these memories will be successful, as evidenced in the delayed memory tests. Note that this pattern in results would have been beyond scientific scrutiny had we not employed a comprehensive methodological approach in which (1) looking time measures were accompanied by pupil dilation measures, and (2) subsequent imitation tests were expanded to include both encoding as well as delayed tests. Another difference worth mentioning is that the looking time data primarily seem to tap into actions, whereas pupil dilation data seem to be a better predictor of sequences. Future studies could be designed to investigate this further.

The differences in results with regard to looking time and pupil dilation could potentially tap into another well-known problem when dealing with looking time paradigms or eye-tracking technology, namely the fact that one can indeed look without paying attention, or as Aslin (2012, p. 2) states, infants might at times engage in blank stares.

Furthermore, it has been found that it is possible to shift attention without actually moving one's eyes, and one can thus overtly orient attention by moving one's eyes or covertly by moving attention (Posner, 1980). Following Courage, Reynolds, and Richards (2006) visual attention is not a simple construct at all and in fact may not completely mirror information processing. Indeed they suggest that different levels of attentional engagement can be found by taking into account phasic changes in infants' heart rates as they engage in a stimulus presentation (see also Richards, 2003). Based on this, simple looking time may not be as effective in detecting increased processing. Hence, pupil dilation may be a more sensitive measure in this regard, due to the fact that the dilations happen spontaneously and for the most part are out of voluntary control (see e.g., Laeng, Sirois, & Gredebäck, 2012). The results obtained here may therefore have consequences for studies and paradigms within infancy research that build on simple looking time (e.g., Violation-of-Expectation studies; see Baillargeon, 2004).

Despite the above-mentioned results, some limitations are worth mentioning. Firstly, considering the relationship between visual measures during encoding and behavioral imitation the correlations were generally small, indicating that other factors – besides looking time and pupil dilation – may contribute to explaining the imitation scores. However, from the present results it is not possible to assess firmly what these factors might be.

Lastly, in order to validate our proposal of pupil dilation as a more sensitive method for assessing the link between visual behavior and imitation, we still need to clarify exactly what it means cognitively when pupils dilate. As previously mentioned, and leaving aside luminance-related effects, pupil dilation is often taken as evidence of mental workload (e.g., Bailey & Iqbal, 2008; Hess & Polt, 1964). However, what does this mean more precisely? At least in principle, a pupil may dilate for different cognitive reasons. For instance, having pupils dilate due to pure surprise is quite different from having them dilate as a result of trying hard to understand one's surroundings. More studies are clearly needed in order to answer this question. One potentially interesting approach in this regard would be to combine pupil dilation with recognition tests employing electrophysiological measures of brain activity, such as ERP. Because ERP measures have been used to assess recognition (e.g., Bauer et al., 2006; Richards, 2003), ERP measures may allow us to specify the cognitive processes underlying pupil dilation in more detail.

At the time of writing many aspects concerning how visual behavior relates to encoding remains uncertain. Hopefully the present study will help by pointing out methodological considerations and suggestions regarding the employment of eye-tracking technology within infant research. In conclusion, the combination of two measures of visual behavior seem to suggest that it is not how long you are looking at something during the encoding of an event that matters, it is whether you pay attention while looking.

References

Aslin, R. N. (2007). What's in a look? *Developmental Science, 10,* 1048-1053. doi: 10.1111/J.1467-7687.2007.00563.X.

Aslin, R. N. (2012). Infant Eyes: A Window on Cognitive Development. *Infancy, 17*, 126-138. doi: 10.1111/j.1532-7078.2011.00097.x.

Aslin, R. N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition. *Infancy, 6*, 155–163. http://dx.doi.org/10.1207/s15327078in0602 1.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence and J.T. Spence (Eds.) *The psychology of learning and motivation: Advances in research and theory,* Vol. II (pp. 89-195). New York: Academic Press.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Hove, England & New York. Taylor & Francis Group.

Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Trans. on Computer-Human Interaction, 14*, 1–28. doi: 10.1145/1314683.1314689.

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science, 13,* 89-94.

Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review, 30*, 128-154. doi: 10.1016/j.dr.2010.03.001.

Barr, R., & Hayne, H. (1999). Developmental changes in imitation from television during infancy. *Child Development, 70,* 1067–1081.

Bauer, P. J. (1992). Holding it all together: How enabling relations facilitate young children's event recall. *Cognitive Development, 7,* 1-28.

Bauer, P. J. (2007). *Remembering the times of our lives: Memory in infancy and beyond*. Mahwah, NJ: Laurence Erlbaum Associates, Inc.

Bauer, P. J., Hertsgaard, L. A., & Wewerka, S. S. (1995). Effects of experience and reminding on long-term recall in children: Remembering not to forget. *Journal of Experimental Child Psychology, 59,* 260–298.

Bauer, P.J., & Mandler, J. M. (1989). One thing follows another: Effects of temporal structure on 1- to 2-year-olds' recall of events. *Developmental Psychology, 25,* 197-206.

Bauer, P. J., Wenner, J. A., Dropik, P. L., & Wewerka, S. S. (2000). Parameters of remembering and forgetting in the transition from infancy to early childhood. *Monographs of the Society for Research in Child Development, 65,* 1–204. http://www.jstor.org/stable/3181580.

Bauer, P. J., Wiebe, S. A., Carver, L. J., Lukowski, A. F., Haight, J. C., Waters, J. M., & Nelson, C. A. (2006). Electrophysiological indices of encoding and behavioral indices of recall: Examining relations and developmental change late in the first year of life. *Developmental Neuropsychology, 29,* 293-320.

Bauer, P. J., Wiebe, S. A., Carver, L. J., Waters, J. M., & Nelson, C. A. (2003). Developments in Long-Term Explicit Memory late in the First Year of Life: Behavioral and Electrophysiological Indices. *Psychological Science, 14*, 629-635.

Carver, L. J., Bauer, P. J., & Nelson, C.A. (2000). Associations between infant brain activity and recall memory. *Developmental Science, 3,* 234-246. doi: 10.1111/1467-7687.00116.

Chong, H. J., Richmond, J.L., Wong, J., Qiu, A., & Rifkin-Graboi, A. (2015). Looking Behavior at Test and Relational Memory in 6-Month-Old Infants. *Infancy, 20*, 18-41. doi: 10.1111/infa.12067.

Colombo, J. (1995). On the Neural Mechanisms Underlying Developmental and Individual Differences in Visual Fixation in Infancy: Two Hypotheses. *Developmental Review, 15,* 97-135.

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*, 1247-1257.

Courage, M. L., Howe, M. L., & Squires, S. E. (2004). Individual differences in 3.5-month-olds' visual attention: what do they predict at 1 year? *Infant Behavior and Development, 27,* 19-30. doi: 10.1016/j.infbeh.2003.05.004.

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*, 680-695. doi: 10.1111/j.1467-8624.2006.00897.x.

DeLoache, J. S. (1976). Rate of habituation and visual memory in infants. *Child Development*, *47*, 145-154.

Elsner, B., Pfeifer, C., Parker, C., & Hauf, P. (2013). Infants' perception of actions and situational constraints: An eye-tracking study. *Journal of Experimental Child Psychology*, *116*, 428-442. doi: 10.1016/j.jecp.2012.11.014

Esseily, R., & Fagard, J. (2013). Ostensive Cues Orient 10-Month-Olds' Attention toward the Task But Delay Learning. *Psychology, 4,* 20-25. doi: 10.4236/psych.2013.47A003.

Gardner, R. M., Philip, P., & Radacy, S. (1978). Pupillary Changes during Recall in Children. *Journal of Experimental Child Psychology, 25,* 168-172.

Geangu, E., Hauf, P., Bhardwaj, R., & Bentz, W. (2011). Infant Pupil Diameter Changes in Response to Others' Positive and Negative Emotions. *PLoS ONE, 6*: e27132. doi:10.1371/journal.pone.0027132

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review, 24.* 33-73. http://dx.doi.org/10.1016/j.dr.2003.09.007.

Hepach, R., Vaish, A., & Tomasello, M. (2015). Novel paradigms to measure variability of behavior in early childhood: posture, gaze, and pupil dilation. *Front. Psychol., 6*, 1-12. doi: 10.3389/fpsyg.2015.00858.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple Problem-Solving. *Science*, *143*, 1190-1192.

Houston-Price, C. & Nakai, S. (2004). Distinguishing Novelty and Familiarity Effects in Infant Preference Procedures. *Infant and Child Development*, *13*, 341–348. doi: 10.1002/icd.364

Jackson, I., & Sirois, S. (2009). Infant cognition: going full factorial with pupil dilation. *Developmental Science, 12*, 670-679. doi: 10.1111/j.1467-7687.2008.00805.x.

Just, M. A., & Carpenter, P. A. (1993). The Intensity Dimension of Thought: Pupillometric Indices of Sentence Processing. *Canadian Journal of Experimental Psychology, 47*, 310-339.

Kingo, O. S., & Krøjgaard, P. (2013). Eighteen-month-old infants generalize to analog props across a two weeks retention interval in an elicited imitation paradigm. *Child Development Research, 2013*, 1-11. http://dx.doi.org/10.1155/2013/786862

Kingo, O.S., & Krøjgaard, P. (2015). Eighteen-month-olds' memory for short movies of simple stories. *Scandinavian Journal of Psychology*, *56*, 151-156. doi: 10.1111/sjop.12201

Kirkorian, H. L. (2012, June). Toddlers' processing of familiar and novel video characters: An eye movement study. Poster presented at the 18th biennial international conference on infant studies Minneapolis, MN, USA.

Kirkorian, H. L., Lavigne, H. J., Hanson, K. G., Troseth, G. L., Demers, L. B., & Anderson, D. R. (2015). Video Deficit in Toddlers' Object Retrieval: What Eye Movements Reveal About Online Cognition. *Infancy,* 1-28. doi: 10.1111/infa.12102

Kolling, T., Óturai, G., & Knopf, M. (2014). Is selective attention the basis for selective imitation in infants? An eye-tracking study of deferred imitation with 12-month-olds. *Journal of Experimental Child Psychology*, *124,* 18-35. http://dx.doi.org/10.1016/j.jecp.2014.01.016

Laeng, B., Sirois, S., & Gredeback, G. (2012) Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*, 18-27. doi: 10.1177/1745691611427305.

McDonough, L., Mandler, J.M., McKee, R.D., & Squire, L.R. (1995). The deferred imitation task as a nonverbal measure of declarative memory. *Proceedings of the National Academy of Sciences*, *U S A*, *92*, 7580-7584.

Meltzoff, A. N. (1988). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development, 59*, 217-225.

Meltzoff, A. N. (1995). What Infant Memory Tells Us about Infantile Amnesia: Long-Term Recall and Deferred Imitation. *Journal of Experimental Child Psychology*, *59,* 497-515.

Nordqvist, E., Rudner, M., Johansson, M., Lindgren, M., & Heimann, M. (2015). The relationship between deferred imitation, associative memory, and communication in 14-months-old children. Behavioral and electrophysiological indices. *Frontiers in Psychology*, *6*,. http://doi.org/10.3389/fpsyg.2015.00260.

Òturai, G., Kolling, T., & Knopf, M. (2013). Relations between 18-month-olds' gaze pattern and target action performance: A deferred imitation study with eye tracking. *Infant Behavior & Development, 36,* 736-748. doi: 10.1016/j.infbeh.2013.08.005.

Pathman, T., & Bauer, P. J. (2013). Beyond initial encoding: Measures of the post-encoding status of memory traces predict long-term recall in infancy. *Journal of Experimental Child Psychology*, *114*, 321-328. doi: 10.1016/j.jecp.2012.10.004.

Posner, M. L. (1980). Orienting of Attention. *The Quarterly Journal of Experimental Psychology, 32,* 3-25.

Richards, J. E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: an ERP study. *Developmental Science*, *6*, 312-328. doi: 10.1111/1467-7687.00287.

Sirois, S., & Jackson, I.R (2012). Pupil dilation and object permanence in infants. *Infancy, 17,* 61-78.

Snyder, K. A., Blank, M. P., & Marsolek, C. J. (2008). What form of memory underlies novelty preferences? *Psychonomic Bulletin & Review*, *15*, 315–321. doi: 10.3758/PBR.15.2.315.

Taylor, G., & Herbert, J. S. (2013a). Eye tracking infants: Investigating the role of attention during learning on recognition memory. *Scandinavian Journal of Psychology, 54,* 14-19. doi: 10.1111/sjop.12002.

Taylor, G., & Herbert, J. S. (2013b). Infant and Adult Visual Attention During an Imitation Demonstration. *Developmental Psychobiology, 56*, 770-782. doi: 10.1002/dev.21147.

Topál, J., Gergely, G., Miklósi, Á., Erdőhegyi, Á., & Csibra, G. (2008). Infants' Perseverative Search Errors Are Induced by Pragmatic Misinterpretation. *Science*, *321,* 1831-1834. doi*:* 10.1126/science.1161437.

Vivanti, G., & Dissanayake, C. (2014). Propensity to Imitate in Autism Is Not Modulated by the Model's Gaze Direction: An Eye-Tracking Study. *Autism Research, 7,* 392-399. doi: 10.1002/aur.1376.

Vivanti, G., Nadig, A., Ozonoff, S., & Rogers, S. J. (2008). What do children with autism attend to during imitation tasks? *Journal of Experimental Child Psychology, 101*, 186–205. http://dx.doi.org/10.1016/j.jecp.2008.04.008.

Woodward, A. L., Sommerville, J. A., & Guajardo, J. J. (2001). How infants make sense of intentional action. In B. Malle, L. Moses, & D. Baldwin (Eds.) *Intentions and intentionality: Foundations of Social Cognition* (pp.149-169). Cambridge, MA: MIT Press.

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Table 1 *Pearson correlation matrix of visual behavior and imitation scores*

Visual Behavior	Mean Looking in sec. *(SD)*	Encoding Test[1] Actions	Encoding Test[1] Sequences	Delayed Test[1] Actions	Delayed Test[1] Sequences
Overall looking time during both demonstrations[2]	37.81*(9.70)*	**r = -.26***	*r* = -.21	*r* = -.15	*r* = -.07
Overall looking time during 1st demonstration[2]	19.55*(5.72)*	*r* = -.22	*r* = -.16	*r* = -.14	*r* = -.05
Looking time at the three actions for all props during 1st demonstration	3.27*(.95)*	*r* = -.24	*r* = -.19	*r* = -.12	*r* = -.05
Looking time at action 3 all props during 1st demonstration	.98*(.44)*	**r = -.25***	*r* = -.22	*r* = -.06	*r* = -.01
Looking time at JJ action 3 during 1st demonstration[1]	.49*(.53)*	*r* = .11	*r* = .09	*r* = .24	*r* = .21
Looking time at SH action 3 during 1st demonstration[1]	.92*(.53)*	*r* = -.23	*r* = -.17	*r* = .02	*r* = .01
Looking time at SP action 3 during 1st demonstration[1]	1.52*(.93)*	*r* = -.12	*r* = -.09	*r* = -.07	*r* = -.00
Looking time at the experimenter during both demonstrations[3]	11.41*(4.32)*	*r* = -.15	*r* = -.16	*r* = -.17	*r* = -.11
Habituation	-	***r* = .25***	***r* = .27***	*r* = .10	*r* = .10
Pupil dilation JJ at action 3 during 1st demonstration[1]	-	*r* = .10	*r* = .11	*r* = -.04	*r* = -.02
Pupil dilation SH at action 3 during 1st demonstration[1]	-	*r* = .01	*r* = .06	*r* = .24	***r* = .24***
Pupil dilation SP at action 3 during 1st demonstration[1]	-	*r* = .24	***r* = .32***	*r* = .20	***r* = .26***

Note. * p<.05 (two-tailed). JJ = Jumping Jack, SH = Shaker, SP= Spinner

[1] = Please note that every time the visual behavior is from a specific prop the correlation is carried out in relation to the specific imitation scores belonging to that prop and not the overall mean. For the looking time, *n* varied between 66 and 68, for the pupil dilation *n* varied between 59 and 67.

[2] = Please note that the correlations are based on proportional looking to control for length differences.

[3] = Please note that these correlations are based on the proportional looking to the experimenter's face.

Notice: This is the author’s version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Figure 1

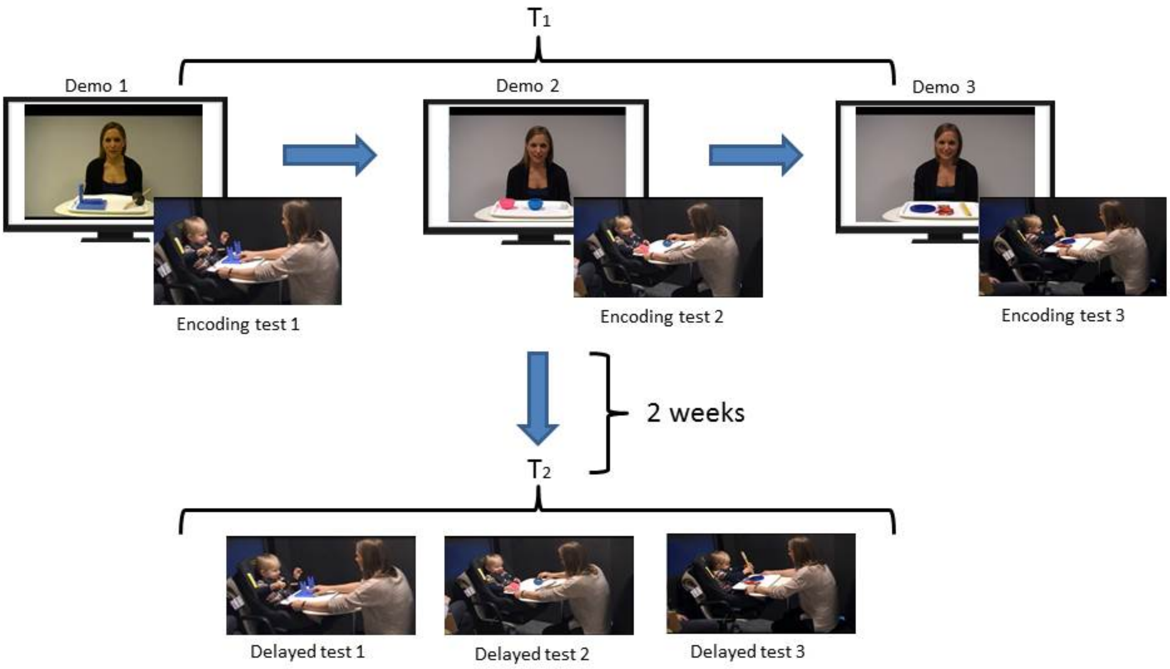


Figure 1: A schematic illustration of the procedure. Note that the order of the props was counterbalanced. For each infant the given order was kept the same at the encoding test and the delayed test. The baseline condition only followed the procedure as depicted for T_2.

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Figure 2

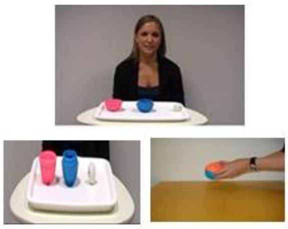
Prop	Objects	Description	Target actions
Shaker (SH)		The Shaker consisted of two half ball shells (one in pale red and one in light blue; diameter = 9 cm), and a white ball (4 cm in diameter).	First you place the ball in one of the half ball shells (action 1). Then you cover up the ball by mounting the remaining half ball shell to the other half ball shell (action 2). Finally you shake the ball with both hands (action 3).
Jumping Jack (JJ)		The Jumping Jack consisted of a blue, circular shaped base (diameter = 14 cm) with a center-placed hole; a yellow 27 cm long wooden stick with a strip of Velcro attached on the top part; and a jumping jack yellow and brown tiger (max h X w = 16 X 16 cm) with velcro mounted on the back side.	The stick is mounted vertically in the center-placed hole of the base (action 1). Then the jumping jack tiger is attached to the stick by use of the velcro parts on both objects (action 2). Finally, the jumping jack tiger is made jumping by pulling the string on the jumping jack (action 3).
Spinner (SP)		The Spinner consisted of a blue square shaped 15 x 15 cm base with two 9 cm tall blue poles (of which one was firm, whereas the other was hinged); and a 20 cm long wooden stick with a black center located, horizontally mounted 'wheel' (diameter = 8 cm).	First, you raise the hinged pole to an upright position (action 1). Then you place the stick on the poles as a cross bar (action 2). Finally, you turn the spinner (action 3).

Figure 2: Still-pictures from the videos and of the three props including descriptions of the target actions.

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Figure 3

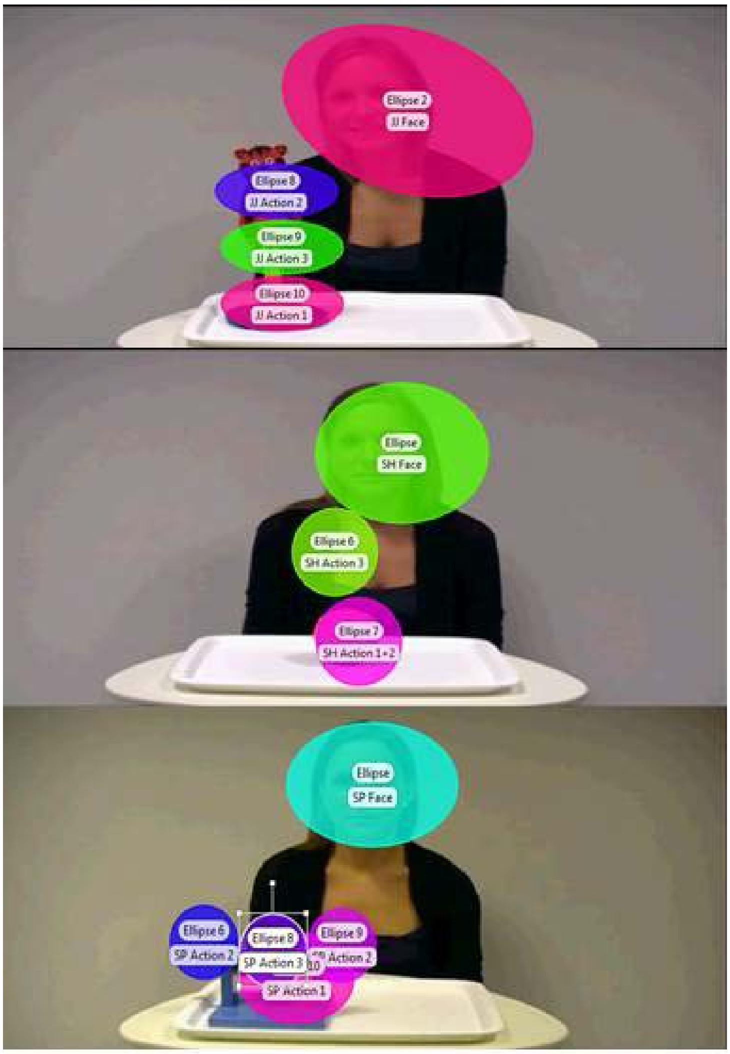


Figure 3: A graphic illustration of the AoIs for the different props. Please note that the whole visual field (indicating everything that is visible on the photo) also functioned as an AoI. From the top: Jumping Jack, Shaker, and Spinner.

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Figure 4

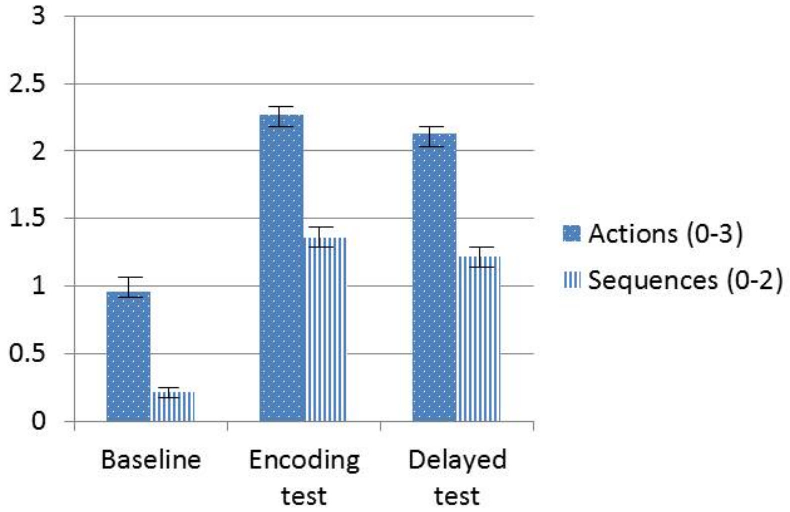


Figure 4: A graphical representation of the actions and sequences produced at Baseline (by a separate group of infants), at the Encoding Test, and at the Delayed Test for all props combined. Error bars display ± SE.

Notice: This is the author's version of a work that was accepted for publication in Infancy. A definitive version was subsequently published in Infancy, 21, 728-750. DOI: 10.1111/infa.12141

Figure 5

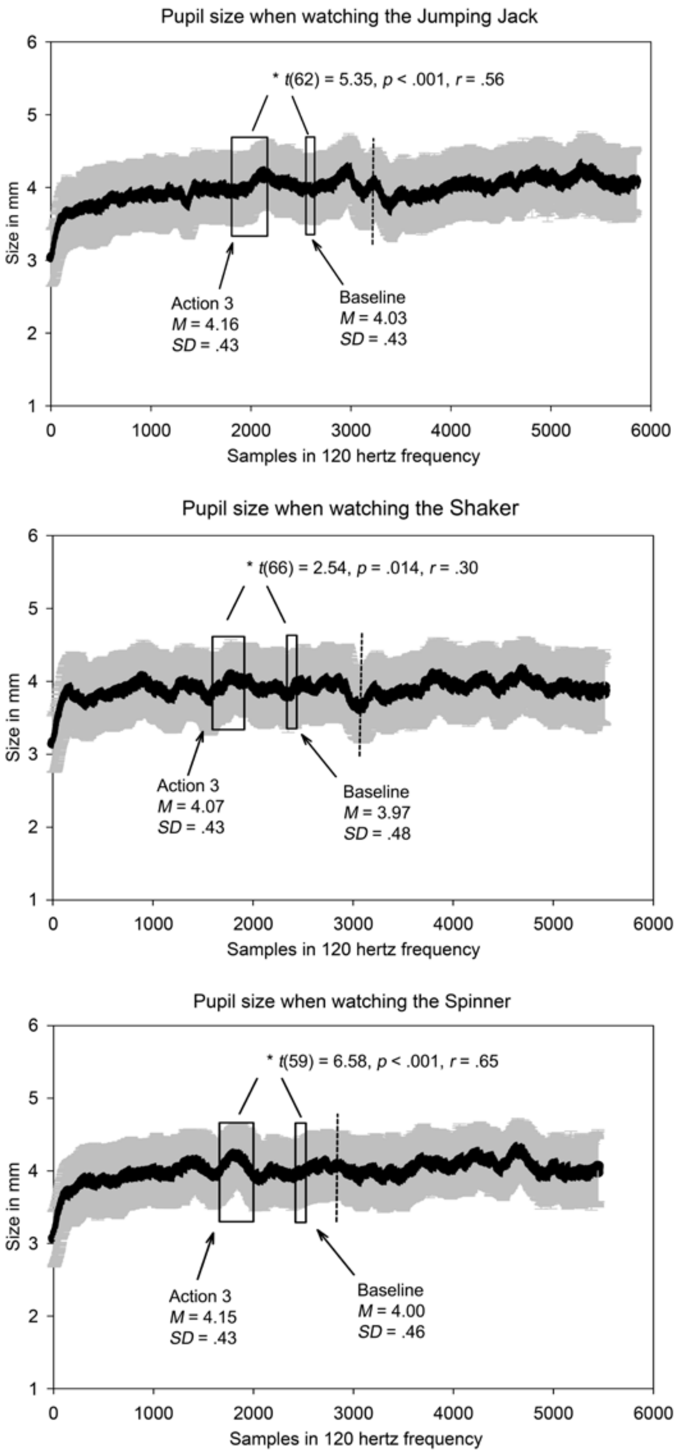


Figure 5: A graphical representation displaying pupil size over time when watching the three different props assessed by samples in 120 hertz frequency. The dotted vertical lines represent the change from first to second iteration and the grey area indicates SD.