MASTERARBEIT/MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Differential guidance of attention shapes infants' neural perception processes in situations of joint attention"

verfasst von/ submitted by

Marcela Victoria Catalán Lorca, B.Sc.

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

Master of Science (MSc)

Wien, 2021/ Vienna 2021

Studienkennzahl lt. Studienblatt/ degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt/ degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG200
Betreut von /Supervisor:	Univ. Prof. Dr. Stefanie Höhl

Table of Contents

1 Introduction and theoretical background

Within the first year of life, children's perception and attention undergo many crucial developments which play an important role in infant’s early learning and their exploration of the environment (Cooper & Aslin, 1990; Feldman, 2007). These developmental processes are largely shaped by the social environment in which an infant grows up (Csibra & Gergely, 2009). For this reason, the question about the interplay and reciprocal interaction of physiological maturation processes and the social context is one of the essential questions of infants’ development. The first social interaction experiences made by an infant usually occur with the caregivers; especially the mother plays an important role in early social interaction (Kuwabara et al., 2011; Lavelli & Fogel, 2013).

The mother shows high sensitivity to early behavioral signs and gestures made by the infants and usually adapts her own behavior to it which gets expressed in changes in speech or vocalization, an increase in gesturing or stimulation via touch (Lavelli & Fogel, 2005). These early social experiences have a major impact on the way infants explore the world. For the exploration of the environment, infants require the ability to direct attention and selectively attend distinct elements. Behaviorally, this is reflected by an increase in their motor activity and responsiveness to social cues like the mother’s voice (Schachner & Hannon, 2011). At the age of one year, infants are able to turn and orientate their body towards elements of interest and follow displayed social cues with their gaze (Hendry et al., 2019). An example could be a situation where a mother points out a distinct element within a book while at the same time asking the infant to look at the depicted element. The infant responds to the invitation and follows the mother's finger pointing with their attention as they gaze and look at the element shown.

However, what exactly happens in this moment within the infants’ brain? Is there a neural response observable when infants’ attention is guided towards a specific object as well as when they attend to it? And if so, how do social interactions influence attention and the neural processing of information? Answering these questions could provide important insights for understanding the development of infants’ attention processes and the impact of social interactions on attention.
Therefore, the current study aims to investigate the neural basis of infants’ attentional changes during social interaction.

1.1 Early social interaction and infants’ visual attention

Within the first weeks of life, infants already show first signs for mutual attentiveness with their mothers. Although infants’ abilities are still limited, they are already able to distinguish different low-level characteristics as for instance changes in face-like stimuli and show a preference for human faces (Hendry et al., 2019; Turati & Simion, 2002). However, mutual attentiveness is mostly depending on mothers' responsiveness to infants' attentional episodes by following infants' attention and responding with a range of maternal behaviors such as gaze turn and infant-directed speech (IDS) (Cooper & Aslin, 1990). This changes during the second month of life, when infants begin to engage in mutual interaction more actively by coordinating their behavior with their attention processes. Mothers respond to these behavioral changes by showing an increase in mirroring infants behavioral and emotional signals for instance by responding to infant smiling with distinct vocalization (Feldman, 2007; Lavelli & Fogel, 2005; Lavelli & Fogel, 2013). Mutual responsiveness and positive feedback within dyadic face-to-face interaction has a major impact on the stabilization and consolidation of early social interactions and is associated with a general increase in mother-infant behavioral matching and synchronicity (Lavelli & Fogel, 2013).

The development of infants' attention processes, social interactions and learning are strongly associated with one another. Adults tend to facilitate children's learning by communicative cues and gestures. These gestures attract infants' attention selectively towards specific entities in their environment and are commonly called ostensive cues (Csibra & Gergely, 2009; Senju & Csibra, 2008). Some of the most evident ostensive cues are direct eye gaze, gaze direction and infant-directed speech as well as referencing gestures, such as finger pointing (Butterworth & Itakura, 2000; Farroni et al., 2002; Senju & Csibra, 2008; Wass et al., 2020). For instance, mothers tend to change their voice to infant-directed speech as soon as they refer to the infant (Schachner & Hannon, 2011; Senju & Csibra, 2008). Infants are highly sensitive to ostensive cues and are more likely to gaze and allocate their attention to a referred object when preceded by an ostensive cue as for instance direct eye gaze (Senju & Csibra, 2008). This is supported by several previous studies that demonstrate an increase in infants' visual orientation and object preferencing by showing an increase in gaze following (Okumura et al., 2020; Szufnarowska et al., 2014; Wass et al., 2020). Many studies explain the sensitivity to ostensive cues on the basis that human interaction and communication are specifically adapted to transfer generic knowledge between individuals and serve as indicators for possible trustworthy interaction partners (Csibra, 2010; Csibra & Gergely, 2009; Csibra & Gergely, 2011; Schachner & Hannon, 2011). Thus, adults would use ostensive cues to indicate relevant information in the environment, and infants would interpret these cues as signs for communicative and trustworthy intent, which provides generalizable knowledge (Csibra & Gergely, 2009).

However, infants abilities to attend and follow ostensive cues undergo noticeable changes within the first months of life, which is strongly associated with their evolving ability in their attention processes (Pauen et al., 2015). During the first 3-4-months, infants largely show a preference for dyadic interactions and seem to respond much less precisely to object-directed ostensive cues. This emerges later in development: at the age of 6-months, infants show higher rates of object-directed gaze following (Feldman, 2007; Pauen et al., 2015). However, around the age of 3 months, infants show a growing interest in other objects and people which supports the general development of infants' attentional processes and leads to an increase in their environmental exploration (Feldman, 2007). Therefore, there is an increase in infants' voluntary body orientation, eye gaze fixations, disengagement and shifts in attention towards and away from objects and people (Amso & Johnson, 2006). Consequently, the underlying processes such as spatial orientation, selective attention and sustained attention are crucial for infants' perception, information processing and interaction with the world (Hendry et al., 2019).

Spatial orientation describes the ability to identify the position of objects and their direction in space, whereas selective attention processes enable the infant to selectively attend certain objects or locate them in the environment. They are influenced by object salience and infants' intentions (Amso & Johnson, 2006). Sustained attention is defined as the ability to maintain the focus of attention on relevant stimuli leading to an extended period of engagement which enhances information processing (Reynolds, 2015). This could already be shown at the age of 2 to 6 months, where infants are able to selectively attend and distinguish between a series of photographs and sustain attention at least for at least a minute (Reynolds, 2015).

Studies indicate that information encoding, and processing are influenced by the context they are presented in. As shown in the study from Yoon et al (2008) presence and absence of ostensive cues have an impact on the information encoded. Depending on the context where information is presented, infants seem to adapt their strategy of encoding; in a context involving ostensive cues, the processing of the identification-features, which refer to more generic properties of an element such as its identity, seem to be enhanced. In contrast, situations without the presence of ostensive cues seem to promote the encoding of context specific information, such as its location (Okumura et al., 2020; Yoon et al., 2008).

Ostensive cues lead to a range of behavioral changes shortly after their emergence and have an influence on attention and information processing (Wass et al., 2020; Yoon et al., 2008). Together with infants developing abilities to selectively attend and sustain their attention, ostensive cues form an essential part of social interactions. They facilitate the identification of what is being talked about and who is addressed and further direct infants' attention to relevant elements in the environment and announce or initiate communicative interaction which has an impact on the integration of information and consequently, on infants' learning.

1.2 Triadic interactions and joint attention

A key transition in infants' early development is the expansion from dyadic face-to-face interaction to more complex interactions which integrate objects as additional reference points. Infants learn to coordinate their own attention with others' and develop the ability to engage in triadic interactions (Butterworth, 2004; Carpenter et al., 1998). Triadic interactions include three foci of attention: infants' awareness of themselves, of an external, attended object and furthermore, their awareness of another person attending the same object. To engage in triadic interaction, infants have to be able to monitor the behavior of others and relate it to themselves.

Studies suggest that infants are already able to differentiate between dyadic and triadic interactions at the age of 3 months (Hoehl & Striano, 2013). However, the way they differentiate between these conditions changes with age. At the age of 3 months, infants show a preference for dyadic interaction by looking more often to the adult but already respond to triadic interaction by varying their response pattern expressed in gazing and smiling. In contrast, infants older than 3 months show a preference for triadic interaction (Hoehl & Striano, 2013; Striano & Stahl, 2005). Studies suggest that infants' sensitivity to triadic interaction develops at a very early age but their ability to functionally take part in such exchanges develops at 9 to 12 months of age (Hoehl & Striano, 2013). The transition from dyadic to triadic interaction goes along with other developmental changes as for instance eye-hand coordination. The development of more complex behavior is associated with an increase in infants response complexity towards their social partners (Hoehl & Striano, 2013).

The awareness of others attending the same object and thus, engaging in a triadic interaction forms the basis for joint attention (JA) which is an important aspect of triadic interactions. Joint attention describes the moment of shared attention between the interactional partners, them knowing that the other attends the same object of interest (Striano, Reid, et al., 2006). Conceptually, joint attention is subdivided into two types: the processes of responding to joint attention (RJA) and initiating joint attention (IJA) (Mundy & Newell, 2007). RJA, which develops earlier, is described as infants' ability to attend an external object by following another person's gestures, whereas IJA is described as a more active process where the infant voluntarily engages with the interactional partner and drives the attention of the other towards an external source (Carpenter et al., 1998; Moore & Corkum, 1998; Mundy & Newell, 2007).

The ability to engage in joint attention has a major impact on communication, perception and learning and introduces the basis for further complex interactions by enabling the infant to respond to an external stimulus, integrating information received by others and thus, attuning their own behavior to it (Butterworth, 2004; Carpenter et al., 1998; Hoehl & Striano, 2013; Tomasello & Farrar, 1986).

Many studies have investigated triadic interactions and joint attention through behavioral looking-time experiments and stimulus preferencing (Cleveland et al., 2007; Reynolds, 2015). Therefore, prolonged looking times are associated with an increase in information processing and stimulus encoding (Petersen & Posner, 2012; Reynolds, 2015). In that sense, results suggest that joint attention supports infants' visual processing and sustained attention by showing prolonged looking times which are interpreted as prolonged object exploration, information processing and object encoding (Hoehl & Striano, 2013; Striano, Reid, et al., 2006; Woodward, 2003).

6-month-olds show a significant prolongation of looking times towards novel toys following joint attention. This effect is not yet significant in 5-month-old infants (Cleveland et al., 2007). In contrast, 9-month-old infants show prolonged looking times during joint attention, but also when they are looking by themselves at a novel object. Although, looking times were significantly longer in the context of joint attention (Cleveland et al., 2007; Hoehl & Striano, 2013). Although behavioral approaches provided important insights into infants' attention processing, the inclusion of neural measuring methods such as electroencephalography (EEG) make it possible to further investigate attentional processes and infants' discriminative abilities (Christodoulou et al., 2018; Hoehl & Wahl, 2012). Therefore, Striano, Reid, et al. (2006) reported significant changes in infants neural processing by their attendance to salient distinct elements within a visual scene. The increase in infants' neural responses remains constant even when elements are presented in an immediate and delayed element recognition task (Hoehl et al., 2014). Further studies suggest that the way infants use and engage in joint attention develops during infancy as their own attentional skills continuously improve and their capacity to attend and concentrate on their own enhances (Striano, Chen, et al., 2006; Woodward, 2003; Yu & Smith, 2016).

An important aspect which has not been discussed so far is that social interactions always occur within a certain cultural context and therefore convey cultural-specific characteristics in behavior and the way the environment is explored. Thus, it is important to consider cultural differences within the development of infants' visual attention processes as they are highly influenced by the social contexts infants grow up in and furthermore to discuss how these cultural differences influence the way infants approach their surroundings (Hoehl & Striano, 2013; Striano, Chen, et al., 2006; Yu & Smith, 2016).

1.3 Cultural differences in the development of visual attention

The cultural and social context in which infants grow up also determines which content is learned and considered relevant. A lot of research has been done on cultural differences in perception and attention in adults, but much less is known about its development and manifestation in infants (Lee et al., 2017; Senzaki et al., 2016). Cross-cultural studies have shown that attention allocation differs between cultures. Adults from Eastern cultures are more likely to allocate attention on contextual information whereas adults from Western cultures are more object-focused (Imada et al., 2013; Kuwabara et al., 2011). Nisbett et al. (2001) argue that cultural differences in context-sensitivity, a phenomenon which is defined as the relative focus on the focal object versus the context of a visual scene, are the results of different cross-cultural approaches to perceiving visual scenes.

Thereby, people from Western cultures are supposed to pay more attention to focal objects of a visual scene, while people from Eastern cultures have a more holistic perception. Many studies support this claim, indicating that people from Eastern cultures show a higher sensitivity to the background and the relationship between objects within a visual scene, whereas people from Western cultures are much less incorporating and considering background information (Imada et al., 2013; Kuwabara et al., 2011; Masuda & Nisbett, 2001a; Senzaki et al., 2016). Due to the significance of attention in infants' learning, context-sensitivity plays an important aspect in the understanding of infants' cognitive development and socialization within a cultural context. A major concern of current developmental research refers to the question how context-sensitivity emerges in infancy and furthermore how it develops during early and middle childhood (Imada et al., 2013; Köster et al., 2017; Lee et al., 2017; Senzaki et al., 2016).

An important factor in the development of context-sensitivity is the influence of parental behavior. Since attention is strongly influenced by parental behavior, it is assumed that children's learning is shaped by their social interactions with caregivers and the repeated experience of culture-specific attention guidance. Especially parent-child interactions play a crucial role in the development of cross-cultural differences in context-sensitivity (Imada et al., 2013; Senzaki et al., 2016). As members of a cultural and social context, parents shape infants learning by conveying cultural characteristics and tendencies such as object preferencing. For instance, US-American mothers tend to guide infants' attention towards objects by referring at their color and shape. In contrast, Japanese mothers refer much more to the context and the relationship between objects by directing infants' attention towards multiple aspects in a visual scene (Imada et al., 2013).

As parental behavior is influenced by the cultural context, it can be assumed that the way children learn in interactions with their parents is also culturally shaped (Masuda & Nisbett, 2001b; Senzaki et al., 2014). Therefore, ostensive cues and joint attention can be seen as important mediators for infants' culture-specific learning. Through repetitive experiences of attention guidance through ostensive cues and joint attention within social interaction, perceptual preferences are transmitted to the infant at an early age, which are then internalized by the child in the course of early and middle childhood. At the age of 10 years, perceptual preferences in context-sensitivity seem to be consistently observed in children's behavior and verbal expressions (Senzaki et al., 2016).

However, the way how perceptional preferences in context-sensitivity are learned seems to be primarily influenced by the social interaction with the mother and their verbal expression and infant directed speech (Senzaki et al., 2016). Kuwabara et al. (2011) indicate that the role of parental behavior and especially the verbal account in the development of context-sensitivity becomes evident when comparing infants' behavioral performance in tasks with and without parental support. Especially, the maternal verbal descriptions which aim to guide infants' attention through visual scenes seem to have particular importance in the transmission of cultural differences in context-sensitivity. Köster and Kärtner (2018) showed that verbal preference can be observed in children's behavior earlier than it is expressed neuronally. They indicate that differences in context-sensitivity are expressed in the children's verbal description already at the age of 5 years.

However, children's neuronal attention processes do not align with their own verbal description until the age of 7 which suggests that context-sensitivity might be first learned via the verbal route until it gets internalized (Köster et al., 2017). However, the way neural attention processes develop and align with respect to behavioral measures of context sensitivity is still subject of an ongoing discussion in the literature.

Therefore, the current study aims to provide new insights into the neural basis of attention processing during social interactions. Although the cultural context in which social interactions occur always has an impact on how information is transferred and on which characteristics and perceptual preferences are conveyed, questions about the neural basis of infants' attention processes during object-background distinction shaped by social interaction and cultural context remain.

For this reason, the current study, even though not cross-culturally, will look more deeply into the neural foundation of infants' object-background sensitivity by assessing the distribution of attention and therefore, investigate how infants' visual perception can be shaped by social interaction in early infancy. In order to capture and track the infants' attentional processes within social interaction from a neural perspective, it is necessary to address the neural dynamics of ongoing brain activity and its relation to attention.

1.4 Neural dynamics of attention and steady-state visually evoked potentials

Although attention is perceived as a continuous phenomenon, many studies have shown that the neural basis of attention fluctuates over time (Busch & VanRullen, 2010; Fiebelkorn & Kastner, 2019; VanRullen, 2018). This is due to the neural architecture of brain activity which oscillates and thus causes fluctuations in information processing, showing peaks of higher and lower excitability (Thut et al., 2011). Local neuronal oscillations are dynamic, self-sustaining and vary in their frequency across the brain. However, they are able to synchronize with each other. Thus, synchronization occurs between several local cell unit oscillations, which can be detected by an enhancement within different frequency bands consistent with cognitive and perceptual functions (Uhlhaas & Singer, 2010).

The frequency and amplitude at which neuronal units oscillate changes in the course of life and thereby their ability to synchronize with other neural units changes as well. Particularly during childhood, the amplitude of oscillations seems to decrease over a wide range of frequency bands, especially in the delta and theta frequencies. Adults in comparison show much higher oscillatory frequencies than infants. This explains why adults and infants' functional frequency bands differ considerably from one another.

Neural oscillatory structures underlying visual perception processes could be measured by examining visual attention. The occurring fluctuations in attention can be assessed through the analysis of visual stimulus processing. Depending on the neural oscillatory phase a visual stimulus is presented, stimulus perception is increased or decreased (Fries, 2015). In other words, whether a stimulus is perceived depends not only on the properties it is endowed with (brightness and contrast), but also on the timing of its presentation within the oscillatory phase (Fiebelkorn & Kastner, 2019).

However, neural oscillations are not only endogenously self-sustaining, but they can also be perturbed by an external rhythmical stimulus (Thut et al., 2011). In the presence of an external periodical event, neural oscillations can become synchronized with the external rhythmic stimulus and begin to cycle in the same period as the external source (phase alignment). This phenomenon of synchronization is called neural entrainment (Calderone et al., 2014; Thut et al., 2011). Therefore, neural entrainment describes the in-phase synchronization of internal neuronal oscillations by an external oscillatory stimulus.

Due to the external synchronization ability of neuronal oscillations, many studies use the concept of neural entrainment for the investigation of neural processes and neural changes as for instance visual attention processing, selective attention, or sustained attention (Norcia et al., 2015; Vialatte et al., 2010). In order to generate visual entrainment, the occipital cortex is stimulated with flashing lights or flickering stimuli to elicit visually evoked potentials (VEPs). VEPs are used to align attentional processes by means of short-term periodic events to subsequentially prioritize certain stimulus processing as they are presented on phase peaks (Vialatte et al., 2010; Wass et al., 2020).

Although VEPs are often recorded as the neural responses to discrete visual stimuli, by presenting sinusoidal modulated visual stimulation over a longer time span, prolonged episodes of VEPs are produced in the brain. The neural response then received remains constant with an equally periodic time course (Norcia et al., 2015). These periodic neural responses are commonly called steady state visually evoked potentials (SSVEPs) (Christodoulou et al., 2017).

The advantage of SSVEPs is that although neural activity is stimulated and entrained by a periodic sinusoidal frequency over time, spectral properties remain constant. This makes it possible to isolate stimulus related responses such as changes in attention by analyzing the SSVEP frequency spectrum which shows frequency peaks that are directly related to the visually attended external stimulus frequency (Norcia et al., 2015).

In terms of directing attention, this method enables the tracking of attention within a visual scene by flickering an object in a distinct frequency. Robertson et al. (2012) could show that infants' SSVEP amplitudes increased during a visual perception task, indicating an increase in neural processing at the location attention was directed to. 12-weeks-olds showed higher SSEVP amplitudes when attention was drawn towards a toy (rotating the toy) which flickered with a frequency of 8 Hz in comparison to the control situation where the toy was present, but attention was not explicitly drawn to it. Due to the periodic and constant stimulation, SSVEP responses are relatively robust against artefacts and have a good signal-to-noise ratio (SNR) but their great advantage is that they are very suitable to reveal modulations in the spectral properties of the neuronal response triggered by the frequency stimulation (Nozaradan et al., 2018). This makes the method particularly applicable for the investigation of attention modulation and attention processing in infancy.

For the generation of SSVEPs in the visual cortex, some frequencies are more suitable than others. This is due to the fact that cognitive processes are strongly related to the intrinsic oscillations und therefore it is easier to stimulate these processes by using matching frequencies in the same frequency band (Norcia et al., 2015). Intrinsic oscillations associated with cognitive processes change over the course of life and differ between adults and children. Infants' brain oscillations show a slightly different structure and the frequency decomposition of these oscillations in terms of their functionality result in lower frequency bands compared to adults'. Thus, for infants' visual entrainment, frequencies in the range between 1-10 Hz seem to be particularly suitable (Christodoulou et al., 2017; Robertson et al., 2012).

A prominent approach in the context of SSVEPs is frequency tagging. This approach was primarily developed to investigate sensory processes in the visual modality and is based on the observation of elicited SSVEP responses which are modulated by the presentation of specific driving frequencies within a visual scene (Nozaradan et al., 2018). As the SSVEP response is influenced by a distinct periodic frequency, different objects within a visual scene can be tagged with a specific frequency which then elicits distinct SSVEP responses when attended.

Therefore, frequency tagging via SSVEPs is particularly suitable for the investigation of selective attention (Christodoulou et al., 2018; Toffanin et al., 2009). It offers the possibility to examine the processing of several stimuli within a visual scene simultaneously. Consequently, it also has the advantage of reducing overall duration of the experiments, which is useful for infant studies. Based on where the attention is directed, neural responses are elicited in the brain with the same rate as the one of the particular driving frequency (Christodoulou et al., 2018).

Due to high trial repetition and by averaging over time, individual trial differences decrease, and event related neurological effects become visible. SSVEPs offer a good signal-to noise ratio (SNR), showing less susceptibility to muscle contraction, eye blinks, movements or other artefacts which frequently occur in infant EEG research (Vialatte et al., 2010). Furthermore, it is non-invasive and easy to apply. Thus, frequency tagging is particularly suitable for collecting data from infants and children. Köster et al. (2017) successfully applied frequency tagging with children in a cross-cultural study about context-sensitivity that investigated object-background processing within a visual scene. For this, they tagged object and background with different driving frequencies, evoking two different neural responses for both object and background, and observed infants' attention distribution within the scene (Christodoulou et al., 2018; Köster et al., 2017).

Frequency tagging with SSVEPs is therefore considered a promising tool to assess neural changes in infants in order to address remaining questions in the current state of research regarding attentional processes in social interactions.

2 Research Questions and Hypotheses

Summarizing previous research, it is assumed that infants' attention processes are strongly shaped by social interaction and the repetitive experiences of attention guidance. The study aims to underpin existing behavioral results with neural measures focusing on the impact of joint attention and ostensive cues on infants' attention within the age of 11-13 month and thus serves to extend the knowledge about the influence of social interactions and guidance of attention on infants' neural processes. It furthermore aims to investigate how infants adapt their visual processing to distinct perception styles established via ostensive cues and examine if infants are able to generalize conveyed perception styles to novel visual scenes, gaining new insights into how social interaction affects infants' visual perception.

To address these questions, we applied the method of frequency tagging with SSVEPs as a measure for attention distribution and for the verification whether infants' attention can detectably be directed to certain elements of a visual scene (i.e., objects and context). Frequency tagging is particularly convenient because it allows the tagging of several elements within a visual scene separately and furthermore, tracks the attention distribution by capturing the corresponding neural response. In general, SSVEP measures are very suitable for infant studies because they are non-invasive, easily applicable with young children and relatively robust against artefacts which is always remains challenging with children's data (Köster et al., 2017; Robertson et al., 2012). Results would make a significant contribution to knowledge about infants' use of parental and supportive behavior and the effects of social interaction on infants' neural basis of attention and learning. Furthermore, little research to date applied frequency tagging to investigate infants' neural basis of attention in a live social context. Thus, results would extend the knowledge and applicability of using the method of SSVEPs in infant EEG studies.

The current study will investigate four main hypotheses which assess whether infant's attention can be successfully directed to distinct elements within a visual scene by applying two different preference styles (object preference or background preference) and furthermore, whether infants are able to generalize the learned preference style to novel elements.

This will be executed by applying a pre-post design with an interaction phase in between. Within the interaction phase, a situation of joint attention will be generated, and infant's attention will be guided via ostensive cues to either an object or the background within a visual scene.

The following four hypotheses will be examined:

I. **The distribution of attention between object and background does not differ between groups in the pre-phase.**

H1. There is no significant difference in the pre-phase between SSVEP O-scores between the object-group (O-group) and the background-group (B-group).

II. **Attention to specific entities can be increased through joint attention (JA) with ostensive cues.**

H2.1. There is a significant difference in the SSVEP O-scores between groups in the interaction phase. SSVEP O-scores are expected to be higher for O-group than for B-group.

H2.2. An SSVEP O-score change is observed from the pre-phase to the interaction-phase for both groups, according to manipulation: O-score increases for the O-group and decreases for the B-group.

III. **Attention to specific entities is increased after interaction-phase (joint attention-phase) with ostensive cues depending on the condition (object/background).**

H3. There is a significant difference in SSVEP O-scores between the pre-phase and the post-phase in both groups, according to manipulation: O-score increases for the O-group and decreases for the B-group from pre-to post-phase.

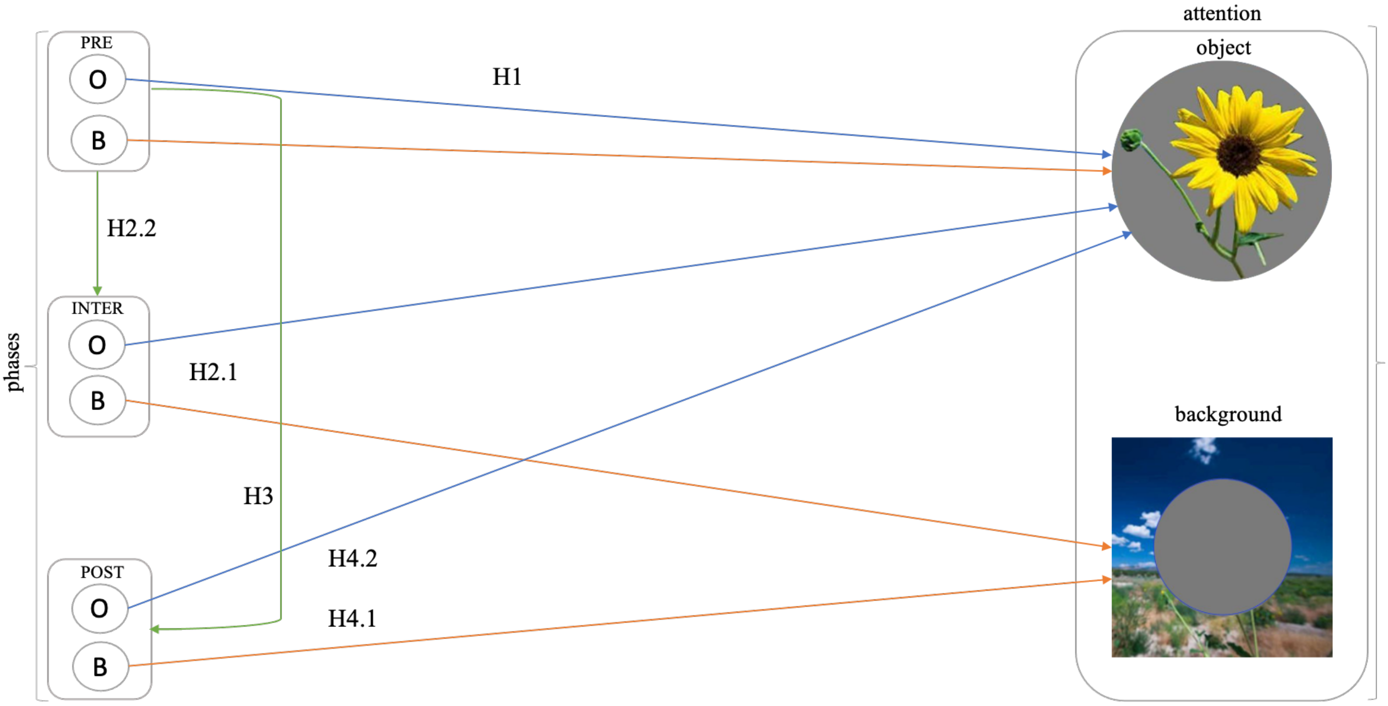
IV. **Increased attention to specific entities (object or background) is transferred to new visual scenes depending on group (object/background).**

H4.1. There is a significant difference in SSVEP O-scores between O-group and B-group in the post-phase, according to manipulation: increased O-scores for the O-group and decreased O-scores for the B-group.

H4.2. SSVEP O-scores for new and old images do not differ within groups (O-group/ B-group) in the post-phase.

Figure 1

Hypotheses plan



Note. Figure 1 illustrates the above hypotheses in a graph. The three phases are shown on the left. O and B stand for the O-scores for object or background. On the right side the visual scene is displayed, but separated between object and background to clarify where the attention is directed to in each hypothesis. The different colors of the lines refer to the respective group condition (O-group/B-group) or phase (Pre/Int/Post). Blue refers to the O-group. Orange refers to the B-group and green refers to the entire phase (e.g., pre-phase).

3 Methods

The study was conducted at the Department of Developmental and Educational Psychology in the research unit *Vienna Children Studies* (Wiener Kinderstudien) at the University of Vienna. Supervision was provided by Univ.-Prof. Dr. Stefanie Höhl and Anna Bánki, MSc.
In the following, the methods and the procedure of the study are presented and described.

3.1 Recruitment

The recruiting was carried out by the Research Unit Developmental Psychology at the University of Vienna. It was organized via the institute's own voluntary database. Parents signed up to the database after recruited at the Wiener Allgemeine Krankenhaus (hospital) and in various parent groups (e.g., baby yoga, parent-child music groups). They were contacted by phone or email and informed about the study, the general procedure and conditions of the study. They were informed that the study was associated with no risk for the infant and that the assessment could be stopped at any time. Infants with prematurity or neurological disorders such as epilepsy were excluded from the recruitment process. For participation and recording of personal data, parents were instructed to provide written informed consent prior to the study. After participation, parents received a travel compensation of 6 euros, the infants received a certificate and a toy.

3.2 Participants

The total sample size consisted of n = 38 infants. A priori G-Power analysis carried out with the software G*Power suggested a total sample size of 44 participants (Faul et al., 2007). Due to Covid-19 regulations, data collection was suspended thus the intended sample size could not be achieved. The sample size used for further EEG-analysis consisted of a subsample of n = 10 typically developing, full-term born infants. During the preprocessing of the data two participants had to be excluded because of incomplete data assessments (for inclusion, infants had to complete at least 2 trials per condition). Thus, the final sample size consisted of n = 8 infants.

3.3 Research Design

In this between-subjects study, a pre-post design with a training phase in between was applied to examine the effect of differential attention guidance on infants' attention distribution within a visual scene. During pre- and post-phases infants sat on their mother's lap and watched flickering images depicting natural object (e.g., animals, everyday objects) on a computer screen. The training phase consisted of a short joint-attention interaction task which was accompanied by attention directive ostensive cues. During training, the infants watched the same images while an experimenter sat next to the infant and pointed out either the object or the background of the visual scenes (dependent on which group the infant was assigned to). For the infants assigned to the ‘object’ group, the experimenter always pointed to the object, whereas for the ‘background’ group, to the background. The number of infants for each group was balanced (n=4 in each group).

3.4 Demographic data

A total of n = 38 subjects was integrated in the descriptive data analysis, which includes all children tested in the study and those who were tested during piloting which preceded the final study procedure. Infants were all in the age group of 11 months and 0 days to 12 months and 30 days (*M* = 12 months, 4 days; *SD* = 0.62). The proportion between boys and girl was relatively balanced (male = 20, 52.6%, female = 18, 47.4%).

Due to the above-mentioned cultural differences in context sensitivity between people from Eastern and Western cultures, the migration background was also surveyed. N = 21, (55.3%) of the parents reported that at least one parent had a migration background, n = 17 (44.7%) did not report any migration background. 27.5% of the mothers were first generation immigrants, living on average 8.14 years in Austria, 10% were second or third generation immigrants. Since all infants were born in Austria and mothers were living in Austria for a substantial amount of time, the effect of cultural background on infants’ attention style was not assessed further in this work. This was supported by the fact that during training, infants interacted with an experimenter trained previously to consequently point out object or context on the visual scenes.

Across the whole sample, the educational level of the mothers was high, 86.8% (n = 33) reported a university degree. Most of the infants were cared for at home (86.8%, n = 33) and did not go to kindergarten or to a childminder. N = 8 infants with sufficient data quality (at least 2 completed trials per condition) were included in the final EEG-analysis. This subsample was balanced in terms of gender distribution (male = 4, female = 4). The average age was 12 months and 5 days (SD = 0.72). All included mothers had an academic education (university degree). Regarding the migration background, n= 6 (75%) parents reported that at least one parent had a migration background. All infants were cared for at home and did not go to kindergarten.

Table 1

Demographic information about the participants

Study Group	EEG-Sample n = 8		Full Sample N = 38	
	n	*%*	*n*	*%*
Gender				
Female	4	50	18	52.6
Male	4	50	20	47.4
Care Taking				
Kindergarten/childminder	-	-	5	13.2
At Home	8	100	33	86.8
Migration Background				
yes	6	75.0	21	55.3
no	2	25.0	17	44.7
Education Level Mother				
University degree	8	100	33	86.8
Other	-	-	5	13.2

Study Group	EEG-Sample n = 8				Full Sample N = 38			
	n	*%*	*M*	*SD*	*n*	*%*	*M*	*SD*
Age (yrs)								
Infant	8	100	0.976	0.021	36	94.7	1.011	0.009
Mother	8	100	34.38	0.822	38	100	34.34	4.97
Infant Siblings								
yes	3	37.5	1.67	1.15	16	42.1	1.19	*.54*
no	5	62.5	-	-	22	57.9	-	-

Note. Table 1 shows the demographic information about the final sample (EEG-Sample, n = 8) which describes the sample which was included into further analysis and the full sample (N = 38). The full sample included all participants which were assessed during piloting and further design adjustment.

3.5 Study Procedure and Stimuli

The study was conducted in the testing room facilities of the Vienna Children Studies of the University of Vienna. A testing session, including instruction and preparation, was scheduled for approximately 60 minutes. At the beginning, the infant was familiarized with the surroundings, while the caregiver was asked to fill out an informed consent form. The mother was then instructed about the EEG measurement, the EEG cap and the general procedure of the study.

The actual experiment was preceded by a short warm-up phase. The intent of the warm-up phase was to create a comfortable, trustworthy situation for the infant. Therefore, the experimenter showed five toys, each for 30 seconds to the infant while speaking and communicating with the infant. The experimenter played the role of an ‘expert’ who showed interesting features of the toys to the infant, to establish a similar situation as in the experiment during the later joint-attention interactional task, so the infant can start to consider her as an ‘informant’ and pay attention to her.

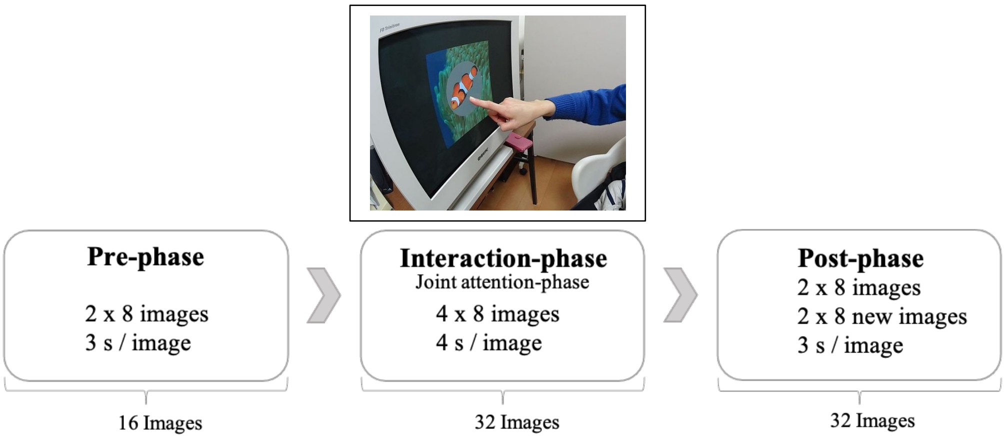
Afterwards, the experimenter showed five infant-friendly visual scenes to the infant in a picture book (“Sachen Suchen”) by pointing with the index finger for at least 3 seconds at one object and one background element on each of the scenes, to further practice for the upcoming joint-attention interaction. Finger pointing was accompanied by infant-directed speech using several predefined phrases: ‘Can you see this?’; ‘Look at this!’; ‘Look here!’; ‘Oh!’; ‘What is this?’. After the warm-up phase, the EEG cap was then positioned on the infant’s head and a constant signal transmission was established.

The experimental task with the EEG recording took approximately 10 minutes, dependent on the infant’s attention span, spontaneous breaks were introduced. In the presence of the caregiver, infants attended the visual attention task: viewing images on a computer screen. During the whole experiment, infants’ neural activity was measured with EEG. The visual attention task was presented on a 17-inch stimulus CRT (cathode-ray tube) monitor with a 85 Hz refresh rate at a viewing distance of 60 cm. The task was sub-divided into three phases. Pre-phase, interaction phase and post-phase.

Parent and infant started the experimental procedure alone, while the experimenter left the room (a second experimenter coordinating software and equipment but not interacting with the dyad stayed in the room behind a separator wall, out of sight of the infant). In the pre-phase, 8 flickering images were shown twice for 3 seconds each. The interaction-phase was characterized by the participation of the experimenter. In this phase, the same 8 images as in the pre-phase were presented four times for 4 seconds each. Thereby, infant and experimenter jointly attended the scenes while the experimenter was pointing out either the object or background, dependent on which group the infant was randomly assigned to (object/background). The allocation to group (object or background) was made prior to the assessment. Finger pointing was accompanied by the same standardized infant-directed speech comments as the ones in the warm-up phase. The five pre-defined phrases were said within a duration of a maximum one second from image onset. During the pointing, the experimenter's face was approximately 20-30 cm away from the computer screen. Direct eye contact between the infant and the experimenter was avoided. After the interaction phase, the experimenter left the room. In the post-phase, parent and infant attended the visual task alone. They had to observe the same 8 images twice as in the pre-phase, but also saw 8 new natural images shown twice (according to the procedure by Köster et al. 2017) intermingled with the old images, for 3 seconds each.

Figure 2

Structure of the study



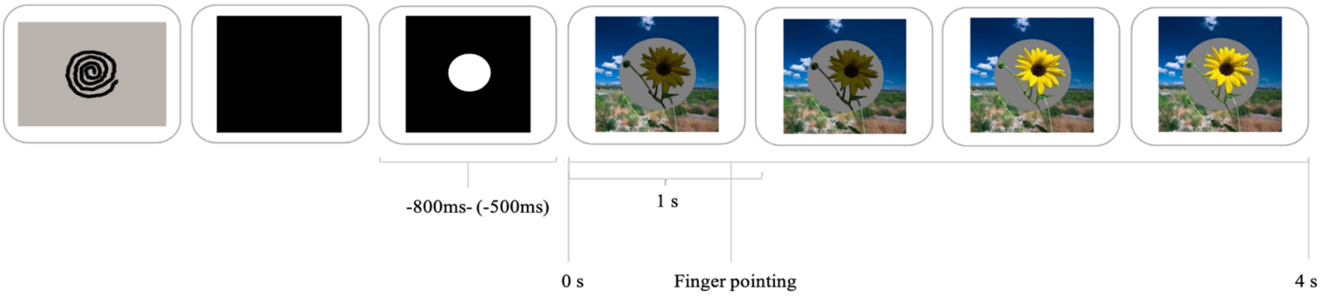
Note. Figure 2 shows the structure of the research design. Photograph adapted from https://www.kinderstudien.at/forschung/laufende-projekte/wie-soziale-interaktionen-wahrnehmungsprozesse-in-der-fruehen-kindheit-praegen-ein-vergleich-zwischen-japan-und-oesterreich/

The implementation of the new images in the post-phase was made to test generalization effects (differential attention guidance in the training phase leads to generalized object-/background-focused attention.) Stimuli consisted of 16 distinct natural, child-friendly images. The images showed visual scenes with e.g., an animal, plant or object displayed on a natural background (see Figure 3). Image order was randomized across phases as well as across participants.

Breaks were given at any time if needed, for instance if the infants became tired or distracted. Then, a child-friendly animation (circulating black spiral in front of a white background) was presented to the child accompanied with music. After the testing session, the caregiver was debriefed on the aim of the study and compensated for the travel costs. The infants received a certificate and a toy.

Figure 3

Interaction phase (joint attention): trial structure



Note. Figure 3 shows trial structure in the interaction-phase (including joint attention). Within the first second experimenter places finger on the screen indicating to the assigned condition. Spiral appears between trials and in breaks. An attention getter was presented within the time window -800 to 0ms. Whole pre-stimuli time-window was (backscreen to stimuli presentation) -1000 to 0ms long. The duration of the attention getter varied within the time window. Trial structure was the same for all phases, with the only difference that for pre- and post-phase images presentation was 3 seconds long.

3.6 Measures

In the following section, the behavioral and neural measurement instruments are presented and discussed in more detail.

Video Recording of the Gaze Behavior

Infants' gaze behavior was recorded by video camera. (Sony Action Digital HD Video Camera Recorder supplied with Live-View Remote RM-LVR2V, http://www.sony.net/SonyInfo/News/Press/201402/14-019E/) with a rate of 60 frames per second. This was conducted to track infants' looking behavior during the experiment and exclude trials when infants were not looking at the screen. Behavioral micro-coding of infants' looking behavior with the software Mangold INTERACT (Mangold, 2020) was carried out to identify such trials.

Electroencephalography (EEG) recordings

Due to the fact that behavioral measurements have their limitations in accessing neurocognitive developmental processes, several studies include non-invasive neural measurements, as for instance electroencephalography (EEG) to explore infants' brain activity during visual attention paradigms (Christodoulou et al., 2018; Toffanin et al., 2009; Vialatte et al., 2010).

Apparatus

EEG was recorded continuously with 24 passive Ag/AgCl ring electrodes (EasyCap GmbH, Herrsching, Germany) from 24 scalp locations determined according to the international 10-20 system. Data were recorded using a wireless SMARTING amplifier system (mBrainTrain LLC, Serbia) at a sampling rate of 500 Hz. Electrode impedances were controlled at the beginning of the experiment and kept under 20 kΩ. Electrode signals were referenced to the vertex (Cz).

Frequency Tagging for Object and Background via SSVEPs

For our study, we strongly relied on the paper from Köster et al. (2017) and used frequency tagging as a measure for the distribution of attention or more specifically on object/background processing.

Therefore, SSVEP frequency tagging was applied to the object and the background within a visual scene. Depending on the element attended, stronger SSVEPs should be elicited in the relevant driving frequency (Köster et al., 2017). Thus, tagging certain features within a visual scene with a distinct frequency enables the disentangling of the neural responses for distinct elements and thus, support the confirmation of the exact location attention is directed to (Köster et al., 2017).

Due to evidence that infants have lower spectral boundaries in EEG, low stimulus frequencies were used to effectively elicit SSVEP responses (Christodoulou et al., 2018).

Therefore, we relied on the frequencies 5.67 Hz and 8.5 Hz, used in the paper from Köster et al (in prep.). For our study, over all conditions (pre-phase, interaction-phase and post-phase), object and background were flickered either in 5.67 Hz or 8.5 Hz. This resulted in two frequency combinations (see Figure 4): if object was flickered in frequency 8.5 Hz, background was flickered in 5.67 Hz (frequency combination 1). If object was flickered in 5.67 Hz, background was flickered in 8.5 Hz (frequency combination 2).

Counterbalancing

To avoid the possibility that one of the two frequencies would elicit higher responses for one element specifically, frequency combinations were counterbalanced between elements. This means that in each phase for 50% of the images, object was shown in frequency 8.5 Hz and for the other 50% of the images, in frequency of 5.67 Hz. For the pre-phase (2x8 images) 4 of the 8 images were shown with object in frequency combination 1 and the other 4 with the object in frequency combination 2.

Each image was repeated twice. Within-subject, a distinct image was always tagged with the same frequency combination in all phases (pre/interaction/post), sub-conditions (new/old images) and repetition. For instance, for participant 7, the image showing a flower was always flickered with frequency combination 1. Between-subjects, frequency combination for each image was randomized across image IDs. Thus, for participant 8, the flower-image, might have been flickered in frequency combination 2.

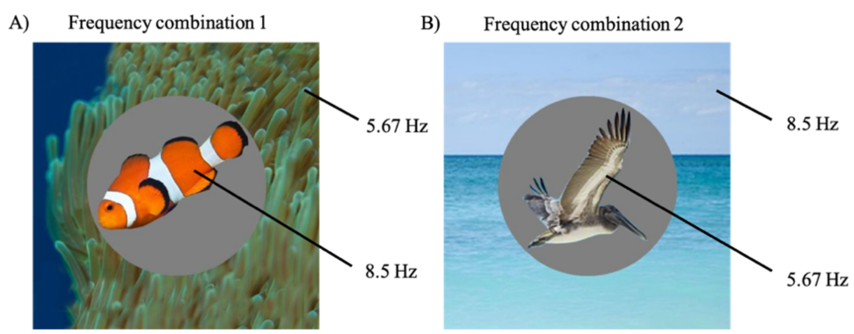
Triggers

To identify and monitor the phase (pre/interaction/post), and the displayed frequency combination for each image, and individual image ID, images were marked with triggers within the continous EEG data. Therefore, pre-phase trials were marked with the trigger-numbers 100 and 200, interaction-phase trials were marked with the numbers 300 and 400 and the post-phase trials were marked with the triggers 500 and 600, whereas post-phase images where again subdivided into old and new images and marked by image ID.

If trigger-numbers were odd (100, 300, 500) the object was shown in frequency combination 1 (8.5 Hz), and if trigger-numbers were even (200, 400, 600), object was shown in frequency combination 2 (5.67 Hz).

Figure 4

Image frequency combination



Note. Figure 4 shows frequency combinations. A) frequency 1: object is displayed in frequency 8.5 Hz. B) frequency combination 2: object is shown in frequency 5.67 Hz. Regarding the Triggers, A) could be Triggers 100, 300, 500 B) Triggers 200, 400, 600.

4 Data Analysis

In the following, we present the data preparation and analysis. This includes the analysis of the videos, the analysis of the EEG data, as well as the analysis of the statistical data.

4.1 Video data analysis

Initially, all recorded videos from all participants were manually coded and analyzed based on the infant's looking behavior. The aim of the coding was to identify those trials in which the children had successfully looked at the screen during stimulus presentation. The excluded trials were further removed from the analysis of the EEG data. For each trial, the onset and offset of the trial were coded based on infants looking behavior within the whole trial.

The Coding procedure for pre- and post-phases was identical. Onset of the trial was defined as image flickering started. Offset was at the end of the trial at image offset or when the infant looked away. In the interaction phase, onset was defined as the moment finger pointing started (within the 1 s after image onset). The trial offset was at the end of the trial or when the infant looked away (see coding manual in Appendix A).

Based on the agreement between two independent coders, the trials identified as when the infant looked away prior to image offset or started looking 500 ms after finger pointing in the interaction phase were excluded from further EEG analysis (interrater reliability $K = 0.89$).

4.2 EEG data analysis

Before the actual data analysis, raw data had to be prepared to become more suitable and interpretable for further analysis. This step is called preprocessing. For EEG data, preprocessing usually refers to the rejection of noisy trials and movement in the data to get closer to the actual neural signal. The inclusion of additional noise during data collection is practically inevitable and the signal derived from the skull does not necessarily represent the actual signal from the brain.

During EEG recording, spatial information gets lost and different sources can contaminate the signal for instance eye blinks and body movements. Thus, the idea of preprocessing is therefore to remove as much artefacts as possible from the recorded data to get as close as possible to the actual neural signal. Preprocessing is an important and time-consuming step before data analysis. It is important to be accurate and meticulous to ensure that the actual signal remains, but that as much noise as possible is removed. The procedure is quite individual and depends strongly on the data and conditions of data collection. For the preprocessing of our study, we strongly follow suggestions and procedures of Mike Cohen (Cohen, 2014; Köster et al., 2017)

Preprocessing

Preprocessing and later data analysis was conducted with the software MATLAB (The Mathworks Inc., version R2019b) and the Toolbox EEGLAB (Delorme & Makeig, 2004). Channel FCz was used as reference channel. EEG signal was band-pass filtered from 1 Hz to 70 Hz. Band-pass filtering is used to keep a certain frequency range between a lower and upper bound. Thus, further analysis focuses on the kept frequency range of interest.

Epochizing

Epochizing facilitates data analysis by extracting certain segments of the data which are task related. Therefore, stimulus onset and offset are used as boundaries to define epochs. For our study, for the pre- and post-phases, epochs of -1s to 4s were extracted. For the interaction phase, epochs had a length of -1s to 5s. A linear baseline-subtraction (-500ms to 0ms) was applied on the data to eliminate differences in the baseline across trials or for instance any drifts in the data which are not removed by the high-pass filter.

Trial Rejection and Channel Rejection

Before the trial rejection, trials which have already been marked by video coding as noisy trials or as unusable trials due to infants looking behavior, were excluded from further data analysis.

Trial rejection was executed manually by the visual examination of two independent inspectors. Due to trial structure similarities, trials from pre- and post-phased were exanimated together. Trials from the interaction-phase was examined separately. For the visual inspection of the trials, scale was set to 250mV. Each trial was inspected, and noisy trials were removed. Therefore, especially trials with noise in the channels of interest (occipital channels; O1, O2, POz) were inspected cautiously. Within a trial, a special emphasis was placed at artefacts between the -500ms before stimulus onset (0s) and 500ms after stimulus onset. Noise in this segment of the trial would have an effect on the baseline and thus impact data quality in the whole trial.

Channel rejection was executed after trial rejection. Therefore, notes about impedances made during data collection were checked and channels with high impedances or bad connection were excluded. The rest of the channels were manually checked by a scale of 500 mV and if necessary, eliminated (maximum of 3 channels for each participant, in addition to mastoid channels). The aim was to keep as much channels as possible, especially the occipital channels and frontal channels (for later independent component analysis) and rather preferred to exclude noisy trials. In the case of more than 5 noisy trials for a channel, channel removal was performed. Based on the agreement between two independent inspectors, final trial and channel rejection was executed.

Independent component analysis (ICA)

Independent component analysis belongs to the blind source separation methods which were developed to face the cocktail phenomenon within a collection of data. The idea is the identification of individual distinct sources within additional sources which is considered as noise.

Thus, the ICA is used to identify noise which does not correspond to brain activity but rather evoked by eye blinks, eye movement or muscle artefacts. By weighting the EEG-signal on the basis of data from each electrode, ICA creates a set of components. Thus, each component has a single time course and is weighted over all electrodes, showing the weight distribution for each component. The aim of the ICA procedure is to remove and identify irrelevant artefacts and noise. Due to several electrodes used during data collection, it is possible to statistically separate the neural signal in the EEG into several independent components.

The actual removal of the components was conducted manually. Due to this reason, component removal is not a standardized procedure, two independent inspectors did the component rejection. Furthermore, scripts from Mike Cohen were used as a reference (see (Cohen, 2014)
For every participant, ICA components were analyzed. Therefore, at first general component weight distribution was checked. Components which had very isolated weight distribution e.g., one electrode were particularly examined in more detail (see Figure 5).

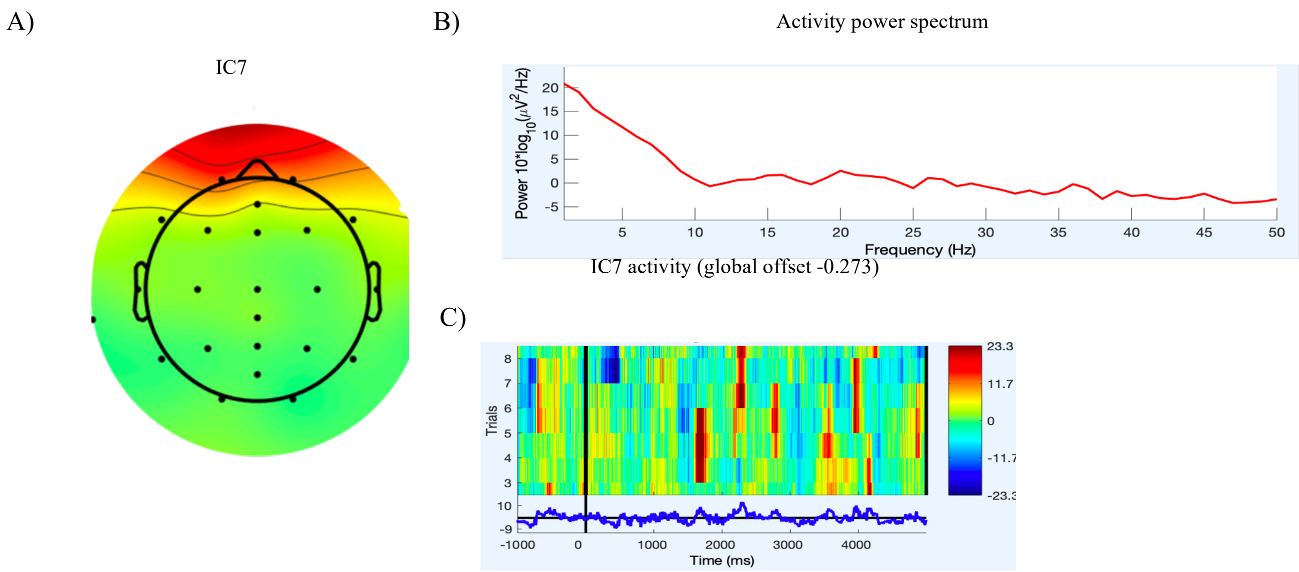
Re-referencing

After the ICA procedure, data were re-referenced offline (Hoehl & Wahl, 2012). The idea of re-referencing is to relate the recorded signal to another reference source. As reference source, any recorded channel can be selected, or an average of all channels. In this study, a so-called common average referencing approach was used. Therefore, the averaged electrical activity measured across all scalp channels was used as the new reference source.

During average re-referencing, the calculated average of all scalp channels is subtracted from the signal of each channel and the overall amplitudes of the EEG signal will be affected and reduced. After re-referencing and before further analysis, data which were already split into conditions (pre-phase, interaction-phase and post-phase) were sub-divided into sub-conditions based on the frequency-combinations (object shown in frequency 1 or frequency 2) and saved separately.

Figure 5

ICA Component analysis



Note. Figure 5 A) shows a very anterior distribution of weights, probably eye blink component. B) shows power spectrum values. If they drop to fast and increases later on it could be an indication for muscle activity and noise. C) ERP-Image: each line corresponds to trial and color corresponds to the amplitude of the single trial component time course. Red sparks indicate high amplitude dynamics e.g., eye blinks. The blue line beneath indicates the time course which does not really seem very task related.

Morlet wavelet analysis

To obtain the evoked potentials (ERPs) over time, trial data were averaged across trials for each participant depending on condition (pre-phase, interaction-phase and post-phase) and sub-condition (frequency combination). Therefore, triggers were used to differentiate between conditions and subconditions. Trigger numbers 100 and 200 indicated the pre-phase, 300 and 400 the interaction-phase, 500 and 600 the post-phase (old images and new images were distinguished based on image IDs).To localize changes in the frequency characteristics over time Morlet wavelet analysis was used with 7 cycles per wavelet (Tallon-Baudry et al., 1997).

Frequencies increased linearly in steps of 0.4 Hz and number of frequencies was 27. The maximum and minimum of the wavelet frequency was 15 Hz and 2 Hz. For the analysis, the signal of three occipital electrodes (O1, O2, POz) electrodes were included with time windows of -1000 too 3500ms for the pre/post-phases and -1000 to 4500ms for the interaction-phase. For SSVEP analysis, amplitude values were calculated and plotted to visually check spectral characteristics and noise features of the signal. The inspection of spectral characteristics of the amplitude values showed an increase after image onset. Amplitude values were baseline corrected following the wavelet-analysis.

Baseline correction

Baseline correction was implemented after a visual inspection of the raw time frequency plots in order to obtain stimulus-related activity during the trial and remove potential noise prior to stimulus onset. As baseline, a time-window between -500 to -250 milliseconds prior to stimulus onset was chosen. This was correspondent with the presentation duration of the attention getter (white circle on black screen) which was randomly set to -800/500 to 0 milliseconds. In addition, resulting decrease in temporal precision and temporal smoothing from time frequency, baseline time-window was not defined until time point 0 prior stimulus onset (see Mike Cohen).

4.3 Statistical Analysis

The statistical analysis was carried out with the Software's *IBM SPSS* and *R* and *Microsoft Excel.* Evoked amplitude values were calculated to SSVEP O-scores, weighted, and used for further statistical analysis. For all statistical calculations, the required assumptions were examined in advance. Normal distribution was tested by Shapiro-Wilk test and analysis of Q-Q-plots. Variance homogeneity was tested by Leven's test.

Calculation for Object-SSVEPs scores

To obtain the object and background SSVEPS values for each participant and sub-condition (pre-phase, interaction-phase and post-phase and within these, the two frequency combinations), evoked amplitude values (in microvolt) were calculated and then subtracted from one other (object-evoked amplitude value – background-evoked amplitude value = object-related amplitude-difference).

In terms of comparability, all SSVEPs were calculated in the direction of the object (SSVEP O-score = object - background). Thus, object-related SSVEP O-scores trended more positively than background-related SSVEP O-scores. In this way, object-related SSVEP (SSVEP O-) scores were generated, and then averaged between sub-conditions within each phase (2-2 sub-conditions in pre- and interaction phases, 4 sub-conditions in post-phase). For instance, for the pre-phase: frequency combination 1 (Tigger 100): the amplitude value for the background-frequency was subtracted from the amplitude value for the object-frequency. The main SSVEP O-scores for each sub-condition were obtained in this way.

Weighting of SSVEP O-scores

Due to the small sample size and the fact that each participant viewed different number of trials, the weighting of sub-conditions based on O-SSVEPs was implemented for each condition (pre-, interaction, post-phase) based on the number of trials each participant executed successfully per sub-condition. For the weighting, sub-condition-based O-scores were multiplied by the number of trials completed in that particular sub-condition and then averaged over the whole number of trials the participant completed for the whole condition.

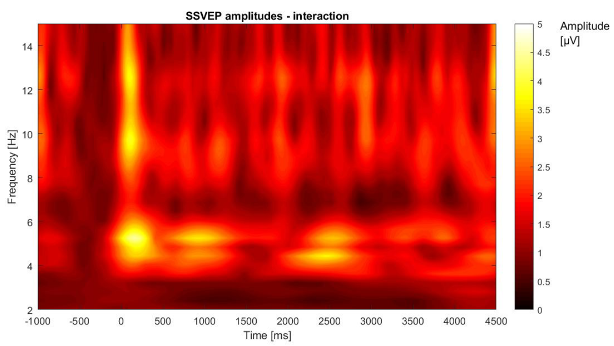
For instance, for the pre-phase the trial number for frequency combination 1 was multiplied with its difference-score (difference_score100), the same was done for frequency combination 2 and then averaged over the whole trial number of the pre-phase. Finally, the weighted SSVEP O-scores were then used for the statistical testing of the hypotheses.

5 Results

The examination of the EEG analysis revealed that regarding the two selected frequencies (5.67 Hz, 8.5Hz), a clear activation at each frequency can be detected: more pronouncedly for the frequency 5.67 Hz and with a slightly reduced activation for frequency of 8.5 Hz. In addition, the first harmonic frequency of 5.67 Hz can also be seen at approximately 11 Hz. This is particularly noticeable in the interaction phase (Figure 6).

Figure 6

Grand Means SSVEP Amplitudes in the Interaction phase

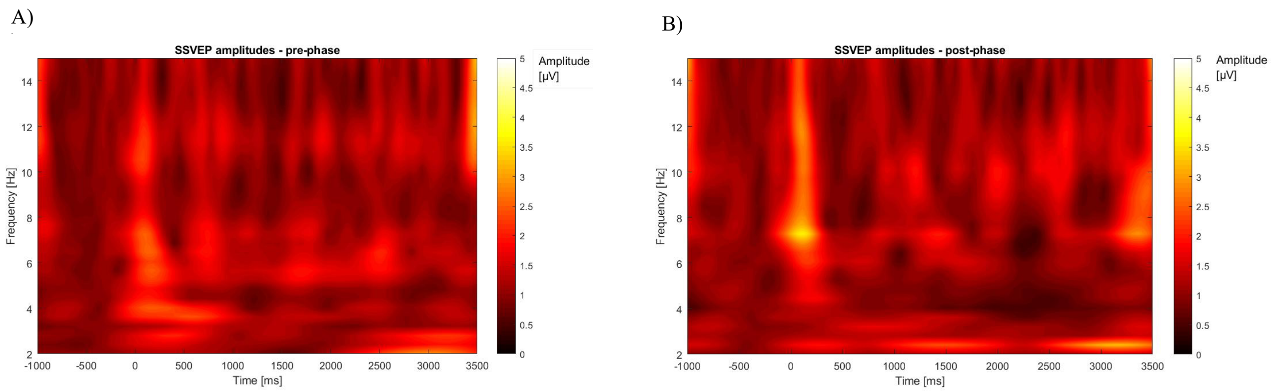


Note. Figure 6 shows the baseline corrected grand-mean TF-plots for the interaction phase (n = 8).

The EEG analysis of the pre- and post-phase shows a slight activation in both frequencies of interest (5.67 Hz and 8.5 Hz). Although, activation seems to be a bit stronger in the frequency 5.67 Hz.

Figure 7

Grand Means SSVEP Amplitudes for the Pre- and Post-phase



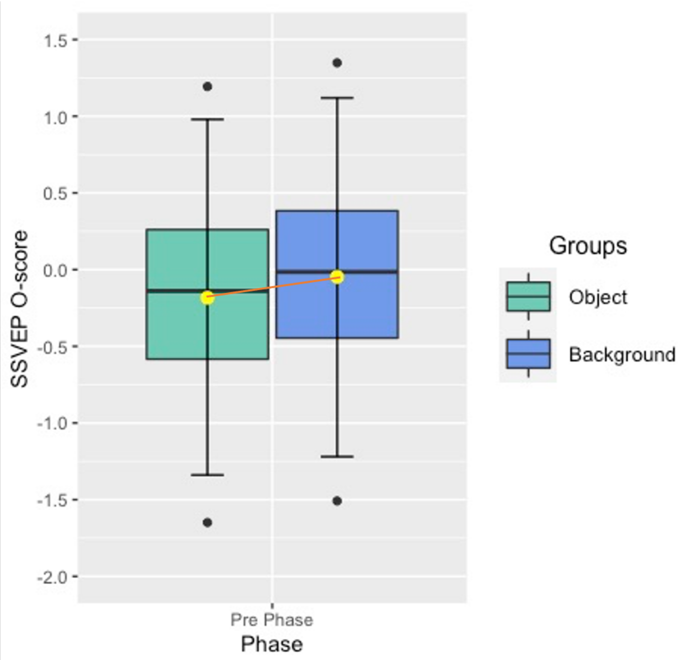
Note. Figure 7 shows the baseline corrected grand-mean TF-plots for A) the pre-phase and B) the post-phase (n = 8).

Hypothesis 1

For hypothesis 1, it was assumed that the distribution of attention between groups (object-group/background-group) does not differ in the pre-phase. H1. examined whether there was a difference in the SSVEP O-scores between object-group and background-group in the pre-phase. Therefore, a univariate ANOVA was calculated. Normal distribution was given, tested by Shapiro-Wilk test, $p = 0.31$. Variance homogeneity was present, tested by Levene's test: $F(1,6) = 0.00$, $p = 0.99$. SSVEP O-Scores for the object-group were lower ($M = -0.18$, $SD = 1.16$) than for the background-group ($M = -0.05$, $SD = 1.17$). However, there was no statistically significant difference in the SSVEP O-scores between groups: $F(1,6) = 0.027$, $p = 0.87$. Hypothesis 1., which assumed that there was no difference between groups, could be accepted.

Figure 8

SSVEP O-score distribution for both groups in the Pre-phase (H1)



Note. Figure 8 shows the SSVEP O-score distribution for both groups (Object/Background) in the Pre-phase. Means are displayed as yellow circles. Median as black lines. Errorbars = 1+-SD

Hypotheses 2

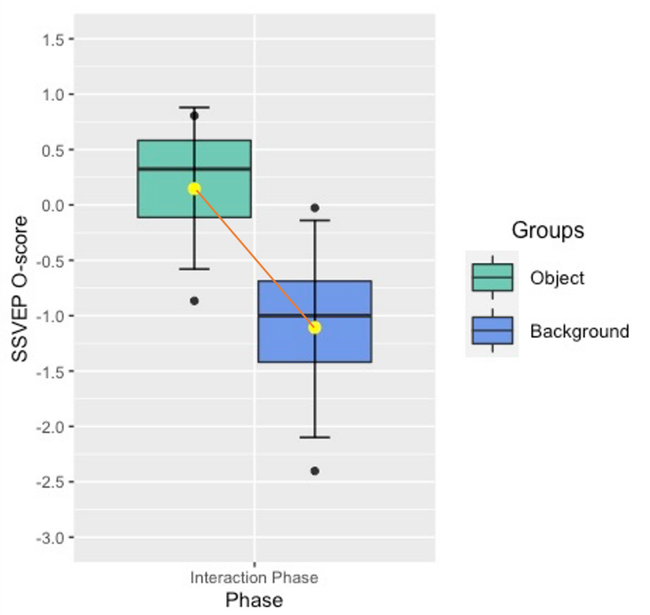
For hypothesis 2, it was assumed that attention to specific entities is increased through joint attention accompanied with ostensive cues in both groups, depending on group-specific manipulation. H2.1 tested whether there was a significant difference in the SSVEP O-scores between groups in the interaction phase, according to manipulation: higher SSVEP O-scores for the object-group than for the background-group.

H2.2 tested if SSVEP O-scores for both groups changed significantly from the pre-phase and interaction phase, according to manipulation: higher SSVEP O-scores for the object-group and lower O-scores for the background-group from pre- to interaction phase.

For H2.1, a univariate ANOVA was calculated. Normal distribution was given for both groups in the interaction phase (Shapiro-Wilk test: object-group, $p = 0.54$; background-group, $p = 0.77$). Variance homogeneity was present, tested by Levene's test: $F(1,6) = 0.13$, $p = 0.73$. SSVEP O-scores were higher for the object-Group ($M = 0.15$, $SD = 0.73$) than for the background-group ($M = -1.12$, $SD = 0.98$). However, there was no statistically significant difference in the SSVEP O-scores between groups, $F(1,6) = 4.23$, $p = 0.09$. Thus, H2.1. which assumed that there was a significant difference between the groups, could not be accepted.

Figure 9

SSVEP O-score distribution for both groups in the Interaction phase (H2.1)



Note. Figure 9 shows the SSVEP O-score distribution for both groups (Object/Background) and their mean distribution in the Interaction phase. Means are displayed as yellow circles. Median as black lines. Errorbars = 1+-SD

For H2.2, a mixed ANOVA was calculated. Normal distribution was given in both conditions (pre-phase/interaction phase), tested with Shapiro-Wilk test ($p > 0.05$).

Table 2

Assumption Checks of hypotheses 2

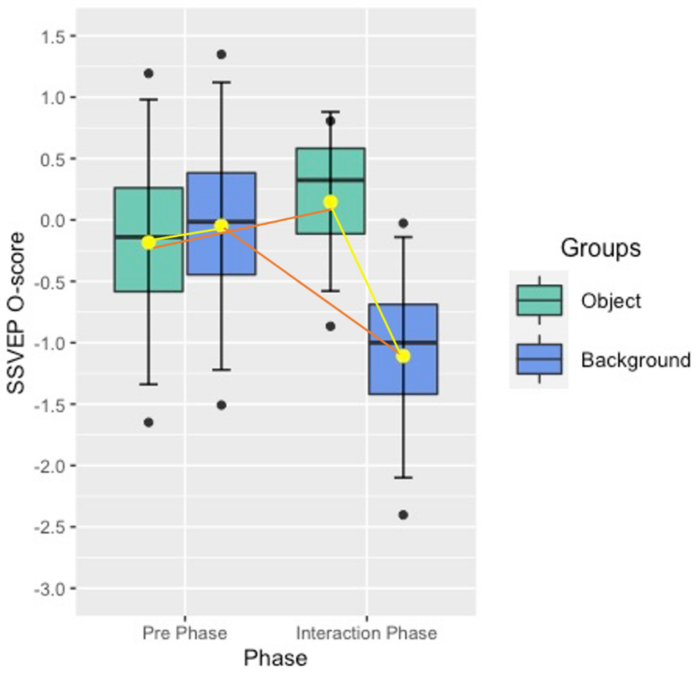
	F	df1	df2	*p*
O-score - pre phase	2.429e -5	1	6	0.996
O-score - interaction	0.129	1	6	0.732

Note. Test for Equality of Variances (Levene's)

SSVEP O-scores for the object-group were higher in the interaction phase (*M* = 0.15, *SD* = 0.73) than in the pre-phase (M = -0.18, *SD* = 1.16). SSVEP O-scores for the background-group were lower in the interaction phase (*M* = -1.11, *SD* = 0.98) than in the pre-phase (*M* = -0.05, *SD* = 1.17). However, no significant interaction could be found between conditions (pre-phase/interaction-phase) and groups (object-group/background-group): $F(1,6) = 2.36$, $p = 0.18$, partial $\eta^2 = 0.28$. There was no statistically significant main effect for phase (pre-phase/interaction phase): $F(1,6) = 0.65$, $p = 0.45$, partial $\eta^2 = 0.09$, nor for group: $F(1,6) = 0.97$, $p = 0.36$, partial $\eta^2 = 0.14$.

Figure 10

SSVEP O-score distribution for the Pre- and Interaction phase (H2.2)



Note. Figure 10 shows the SSVEP O-score distributions for the Pre-phase and Interaction phase. The yellow dots display the means. Yellow lines indicate the mean values between groups, the orange lines indicate the mean differences between phases within a group. The black lines represent the median.

Hypothesis 3

For hypothesis 3, it was assumed that attention to specific entities is increased following the interaction phase (joint attention phase with ostensive cues), depending on group (object-group/background-group). Therefore, H3 tested if SSVEP O-scores within group changed significantly between pre-phase and post-phase, according to manipulation: SSVEP O-score increases for the object-group and decrease for background-group. Therefore, a mixed ANOVA was calculated. Normal distribution for both groups was given in both phases (pre-phase/post-phase), tested with Shapiro-Wilk test ($p > 0.05$).

Table 3

Assumption Checks of hypothesis 3

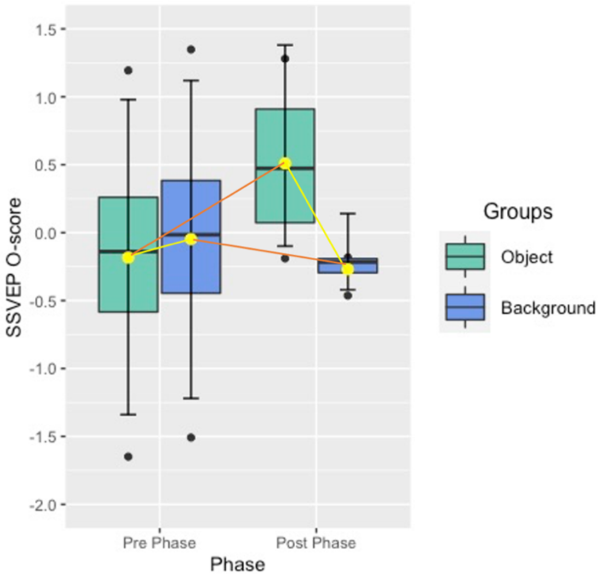
	F	df1	df2	*p*
O-score – pre phase	2.429e -5	1	6	0.996
O-score - post	3.084	1	6	0.130

Note. Test for Equality of Variances (Levene's)

SSVEP O-scores for the object-group were higher in the post-phase ($M = 0.64$, $SD = 0.74$) than in the pre-phase (M = -0.18, $SD = 1.16$). SSVEP O-scores for the background-group were lower in the post-phase ($M = -0.14$, $SD = 0.28$) than in the pre-phase ($M = -0.05$, $SD = 1.17$).

Figure 11

SSVEP O-score distribution for Pre- and Post-phase (H3)



Note. Figure 11 shows the SSVEP O-score distributions for the Pre-phase and Post-phase. The yellow dots display the means. Yellow lines indicate the mean values between groups, the orange lines indicate the mean differences between phases within a group. The black lines represent the median.

However, no significant interaction could be found between condition (pre-phase/post-phase) and groups (object-group/background-group): $F(1,6) = 0.76$, $p = 0.42$, partial $\eta^2 = 0.11$. There was no statistically significant main effect for phase (pre-phase/interaction phase): $F(1,6) = 0.48$, $p = 0.51$, partial $\eta^2 = 0.08$, nor for group: $F(1,6) = 0.71$, $p = 0.43$, partial $\eta^2 = 0.11$. Thus, hypothesis 3 could not be accepted (see Figure 10).

Hypotheses 4

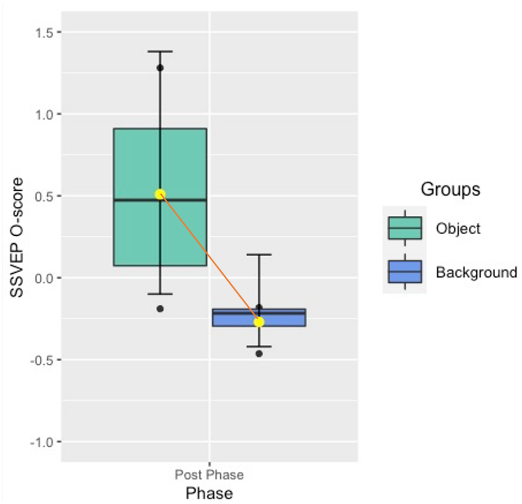
For hypothesis 4, it was assumed that attention to specific entities is transferred to new visual scenes depending on group (object-group/background-group).

H4.1 tested whether there was a significant difference in the SSVEP O-scores between groups for all post-phase images, according to manipulation: SSVEP O-scores increase for the object-group and decrease for background-group. H4.2 tested if SSVEP O-scores within group (object-group/background-group) differ between old and new images.

For H4.1 a univariate ANOVA was calculated. Normal distribution was given for both groups in the post-phase (Shapiro-Wilk test: object-group, $p = 0.73$; background-group, $p = 0.41$). Variance homogeneity was present, tested by Levene's test: $F(1,6) = 3.08$, $p = 0.13$. SSVEP O-Scores were higher for the object-group ($M = 0.64$, $SD = 0.74$) than for the background-group ($M = -0.14$, $SD = 0.27$). However, there was no statistically significant difference in the SSVEP O-scores between groups, $F(1,6) = 3.87$, $p = 0.09$. Thus, H4.1. which assumed that there was a significant difference between the groups, could not be accepted.

Figure 12

SSVEP O-score distribution for both groups in the Post-phase (H4.1)



Note. Figure 12 shows the SSVEP O-score distribution for both groups (Object/Background) and their mean distribution in the Post-phase. Means are displayed as yellow circles. Median as black lines. Errorbars = 1+-SD

For H4.2 a mixed ANOVA was calculated. Normal distribution for both groups was given in both conditions (post-phase old images/post-phase new images), tested by Shapiro-Wilk test ($p > 0.05$).

Table 4

Assumption Checks of hypotheses 4

	F	df1	df2	*p*
O-score - post_old	0.708	1	6	0.432
O-score - post_new	6.888	1	6	0.039

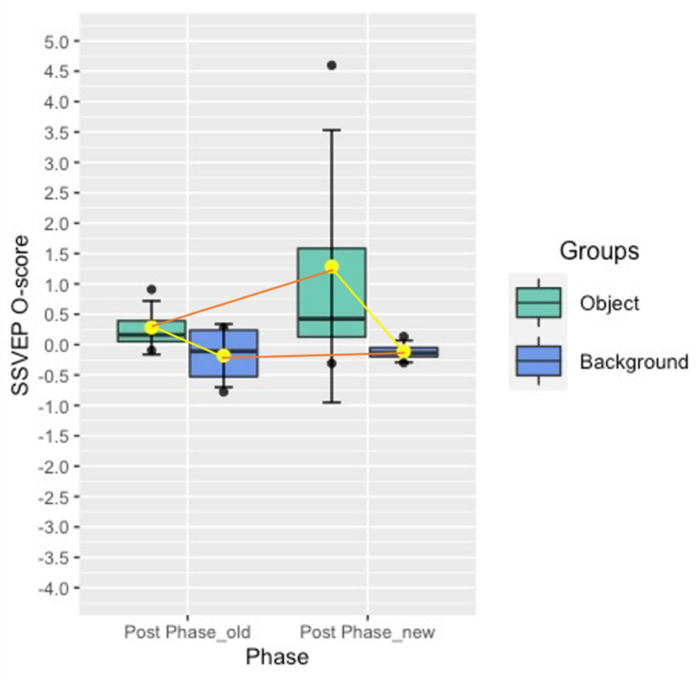
Note. Test for Equality of Variances (Levene's)

SSVEP O-scores for the object-group were higher for the new images ($M = 1.29$, $SD = 2.24$) than for the old images ($M = 0.28$, $SD = 0.44$). SSVEP O-scores for the background-group were higher for the new images ($M = -0.11$, $SD = 0.18$) than for the old images ($M = -0.18$, $SD = 0.52$).

No significant interaction could be found between images (old/new) in the post-phase and groups (object-group/background-group): $F(1,6) = 0.54$, $p = 0.49$, partial $\eta^2 = 0.08$. There was no statistically significant main effect for image sub-condition (old/new): $F(1,6) = 0.71$, $p = 0.43$, partial $\eta^2 = 0.12$, nor for group: $F(1,6) = 3.02$, $p = 0.13$, partial $\eta^2 = 0.34$.

Figure 13

SSVEP O-score distribution for the Post-phase, old and new images (H4.2)



Note. Figure 13 shows the SSVEP O-score distributions for the Post-phase for the old images and Post-phase for the new images. The yellow dots display the means. Yellow lines indicate the mean values between groups, the orange lines indicate the mean differences between phases within a group. The black lines represent the median.

6 Discussion

The current study aimed to investigate the neural dynamics of infants' visual perception altered by preferential attention guidance towards specific elements within a visual scene and furthermore examined whether these socially learned preferences are transferred to novel elements within a visual scene. Therefore, two perceptual preference styles were applied, each of which directed infants' attention either to the object or to the background of a visual scene. Infants' neural processing was tracked via frequency tagging with SSVEPs by applying two distinct frequencies (5.67 Hz and 8.5 Hz). Results were expected to provide further insight into the neural processes of infants' attention in the context of social interactions and additionally, expand the knowledge on the applicability of SSVEPs and frequency tagging in infancy research.

Considering the background of the current research, it has been assumed that the frequency tagging of object and background with both frequencies would elicit two separated neural responses in corresponding frequency bands. This was confirmed by the analysis of time frequency plots which show separated activation for both frequencies, implying that infant's visual processing was successfully entrained via SSVEPs. Across all phases, the most prominent activation in both frequencies is shown in the interaction phase (Figure X). Taking into account current literature, a possible explanation for this could be the impact of the social interaction on infants' visual processing. Because only during interaction phase, infants' visual processing is directly affected by the social interaction which involves the experience of joint attention together with the mediation of a distinct preferential perception style via ostensive cues. Infants might have looked more particularly at the directed element by following the ostensive cues and therefore, visual processing of the elements was stronger influenced by the presented flicker frequency. Therefore, as infants' attention is most strongly influenced and affected during the interaction phase, it is reasonable to expect higher SSVEP amplitudes in both frequencies of interest than in the other phases. However, frequency 5.67 Hz appears to elicit a slightly stronger activation than frequency 8.5 Hz. A reason for the stronger activation for 5.67 Hz could be that lower frequencies actually work better for infants in this age and elicit stronger activation. However, and even though the suitability of those frequencies has to be object of future research, time frequency plots suggest that frequency tagging of distinct elements within a visual scene seems to be suitable to produce an activation in the neural data.

With regard to the hypotheses, hypothesis (H1) examined the pre-phase in which infants were looking on their own at the displayed visual scenes. In line with the current literature which suggests that infants intrinsic, culturally biased preferencing evolves much later in infancy, it was assumed that although infants were already assigned into groups (object-group/ background-group), no significant differences between infant's visual processing in attention distribution would be found between both groups. This hypothesis (H1) could be accepted. Accordingly, results suggest that infants at the age of 11-13 months do not yet independently show any perceptual preference for object or background of a visual scene since they had not received any social cues (Köster & Kärtner, 2018; Senzaki et al., 2016). Another interesting aspect is shown considering the SSVEP O-means between groups (object-group/ background-group). It is shown that SSVEP O-means for the object-group are slightly positive, whereas SSVEP O-means for the background group are slightly negative, indicating a marginal difference between both groups, however remaining non-significant. Generally, if infants visual processing would be biased by a perceptual preference style, in case of object-preferencing the SSVEP O-means would show significantly positive values whereas background-preferencing would lead to negative SSVEP O-means. Therefore, slightly positive values for the object-group and slightly negative SSVEP O-means indicate that, although results remain non-significant, SSVEP O-means align with the expected trend direction.

With respect to the hypotheses 2, two separate hypotheses were examined. Whereas H2.1 considered infants visual processing differences between groups (object-group, background-group) for the interaction phase, H2.2 investigated phase differences between pre-phase and interaction phase. In light of the current literature, it was assumed that due to the influence of the social interaction, including joint attention and preferential guidance via ostensive cues, differences between groups would enlarge, implying successful directing infants' attention to either object or background. In addition, and assuming a successful preferential guidance, phase comparison between pre-phase and interaction phase would show a prominent difference for phase and group. However, no significant differences could be found neither between the groups (H2.1) (object-group / background-group) nor between the pre- and interaction-phase (H2.2). This means that the way infants directed their attention to either object or background did not differ significantly between both groups and the impact of the social interaction was not as strong as expected by showing no significant differences between pre-phase and interaction phase. However, considering SSVEP O-means between the groups within the interaction phase, SSVEP O-means for the object-group become more positive, while they become more negative for the background-group (Figure 9). SSVEP O-means comparison between pre-phase and interaction phase shows an equal trend (Figure 10). Therefore, although the results did not reach significance, changes in SSVEP O-means support the underlying hypothesis and encourage further investigation.

The third hypothesis (H3) tested whether infants' visual processing changed significantly between pre-phase and post-phase. It was assumed that infants visual processing in attention allocation would have changed in the post-phase accordingly to the conveyed preferential perception style (object-preference/ background-preference) experienced during interaction phase. However, no significant changes of infants' visual processing could be found for groups between pre-phase and post-phase. Thus, SSVEP O-scores did not significantly differ between groups (object-group/ background-group) in the pre-phase to the post-phase. This implies that the stimulation on infants' attention experienced during social interaction is not strong enough to significantly influence infants' visual processing of the attentional distribution beyond the interaction phase.

Although results remained non-significant, by analyzing the SSVEP O-means of the groups between pre- and post-phase, SSVEP O-means change in the expected directions (Figure 11). As to say, SSVEP O-means for the object-groups, which were expected to become more positive if attention allocation was successfully guided towards the object, changed accordingly. The same trend was observable for the background-group, as results demonstrate a more negative SSVEP O-mean in the post-phase than in the pre-phase. However, particularly the object-group shows a larger change of the SSVEP O-means between the phases. This stronger change between pre- and post-phase for the object-group could be explained with the presentation of the new images in the post-phase. In addition to an overall novelty effect, an object-preference may have occurred with the new images. This could be due to the fact, that within the visual scene, the objects were always presented in the center of the image which for the new images was more eye-catching than the background. Another reason is related to the structure of the image: the object was always displayed on a grey circle which separated object from the background. Accordingly, the gray circle might also have caused the background to be less well perceived.

The fourth hypotheses (H4) took a closer look on the post-phase. Therefore, two separate hypotheses were examined. Hypothesis 4.1 examined infants visual processing in attention allocation between groups (object-group, background-group) in the post-phase and hypotheses 4.2. investigated depending on the group, whether there was a significant difference in the visual processing of attention distribution between the old and new images displayed in the post-phase.

However, hypothesis 4.1 remained non-significant, indicating that group differences were not strong enough to reach significance. It was assumed that if infants would apply the socially mediated preferential perception style, differences between groups (object-group/ background-group) within the post-phase would have been found. However, no significant perceptual preference in attention allocation could be found for either object group or background group. Considering the SSVEP O-means between object-group and background-group, however, SSVEP O-means for the object-group show positive values while the SSVEP O-means for the background are slightly negative which is in alignment with the expected trend (Figure 12).

With regard to hypothesis 4.2. no significant results were found, indicating that depending on the group, infants' visual processing of new and old images did not significantly differ between each other. This implies that infants visual processing did not significantly changed with the presentation of the new images. However, although results remained non-significant, considering SSVEP O-means between groups for old and new images it is shown that SSVEP O-means change and become slightly more positive as compared to the old images. Most prominently this is shown in SSVEP O-means of the object-group but also for the background-group. Although the SSVEP O-mean changes are small, it is noteworthy to mention that SSVEP O-mean enlarged in a positive direction (Figure 13). This enlargement raises questions about the perception of the new images and if their perception is generally more object focused.

Furthermore, changes in the SSVEP O-means for the new images provide questions about a possible novelty effect which seems worthy to prove in further studies. A novelty effect would imply that the new images are noticed by the infants and perceived as new and different from the already known images. For this reason, more attention would be drawn to the new images and SSVEP O-means would enlarge. Assuming that a novelty effect is present, the SSVEP O-scores from the pre-phase which only included old images, were tentatively compared with the new images from the post-phase (Appendix B). Therefore, by exclusively comparing the new images from the post-phase with the old images from the pre-phase, differences remained non-significant. However, only the group mean SSVEP O-score for the object-group enlarge for the new pictures. Therefore, an interplay between novelty effect and preference style should be examined in further studies.

The current findings suggest that the mediated preferential perception style which was conveyed during social interaction in a situation of joint attention via ostensive cues was not successfully applied and transferred to the post-phase. Already during interaction phase, where infants' visual processing was directly influenced by the social interaction no significant differences between groups in their attention allocation were found. Therefore, several reasons have to be mentioned as possible explanations for the current results. It should be considered whether, during interaction phase, the social stimulation via joint attention and pointing was not strong enough to elicit a constant transference of the preferential perception style within interaction phase but also to the post-phase.

Consequently, it should be discussed whether it is possible to enhance the social stimulation by an additional ostensive cue as for instance direct eye gaze. Considering current literature, direct eye gaze is a strong ostensive cue which enhances infants' gaze following and information encoding, especially in combination with finger pointing (Senju & Csibra, 2008). Therefore, direct eye gaze as a preceding cue to pointing could work as a sort of attention getter and furthermore strengthen infant's relationship with the experimenter. This could improve infant' willingness to follow the mediated preference style conveyed by the experimenter. However, it is possible that direct eye gaze is perhaps too strong as a stimulus and thus leading to a distraction of infants' attention away from the visual scene.

Further aspect which could have had an impeding influence on the results, is the low number of trials and the low sample size. Infants' mobility and fussiness could have had a significantly stronger influence on the results and thus may have masked actually existing effects. Infants' fussiness and concentration loss could have been influenced by the length of the experiment, as well as by the strong flickering caused by the visual stimulus. Especially for infants used to little screen time, the strong stimulation could have been perceived as very exhausting and thus led to a loss of concentration. Nevertheless, a larger sample size would shed light on potential differences between the groups within and between phases. Therefore, further research is needed to confirm these assumptions.

Limitations and Future Research

A main limitation of the study was its small sample size. The previously conducted power analysis resulted in an ideal sample of 44 participants. However, the aimed sample size could not be achieved due to the global Covid-19 pandemic, as safety could not be guaranteed for the participants. Another difficulty was the low number of trials within the individual phases, which, taken together with the small sample, had an additional impeding effect on the analysis and interpretation of the results.

A high number of trials had to be excluded since the EEG data of many of the collected trials were very noisy and had artifacts (e.g., movement, eye-blinks), which is often found in the data collection of infants, since they move much more frequently, are less controllable and have a shorter attention span. However, apart from artifacts in the neural signal, infants' looking behavior was also assessed via video recording and coded with a behavioral coding manual (Appendix A). This was made to only include and correctly identify trials where the infants were actually looking at the screen. However, regarding the video coding, there were some difficulties in identifying usable trials because in some cases, the determination of whether the infant was looking at the computer screen was ambiguous. Therefore, it may be necessary to discuss alternative or additional verification techniques. As a quality check of the data and to obtain further insights into existing data, the signal-to-noise ratio (SNR) was calculated post hoc.

The SNR calculation showed low values in all phases overall, but especially in the post-phase. This suggests that EEG signal was strongly influenced by present noise which again stresses the importance of a cautious interpretation of the results.

The especially low SSVEP amplitude and SNR values in the post-phase could additionally indicate that infants' attention dropped in the post-phase. Despite trying to maintain ecological validity, the overall experiment including preparation and warm-up phase, might still have taken too long for the infants in order to concentrate on the stimulus until the end (post-phase). Furthermore, experimental task repetition could have become too repetitive and thus, infants got bored observing the same thing, using the same pointing, and variating the same 5 sentences. Another explanation could be the flickering of the stimulus, i.e. the strong flickering could have made it slower and more exhausting for infants to pay attention for a longer period of time.

Therefore, length of the experimental task, the repetitive character and the stimulus qualities of the experiment should be revised and discussed in future studies while taking into account infants' smaller attention span and lower concentration capacity.

Another limitation that has to be considered and discussed in future research relates to the fact that only two very distinct frequencies were used. Therefore, it would be interesting to test other frequencies and their suitability to entrain infants' visual attention processes. One last aspect which could be considered and discussed is the impact of the preceding warm-up phase on infants' behavior and willingness to engage with the experimenter during the experiment. It is possible that without the warm-up phase, infants would respond stronger to the social cues as they would not have experienced them in the warm-up phase. To address these questions, the implementation of a control group could help to find answers accordingly.

Conclusion

The current study aimed to investigate the influence of social interaction on the visual processing of infants' attention distribution. Therefore, infants' attention was guided to either the object or the background within a visual scene, transferring a preferential perception style.

For the investigation of infants' attention distribution in the context of object/background preferencing, the current study is one of the few studies which uses the method of frequency tagging in early infancy to track attention processes in a live social interaction. Although, the results remained non-significant, they indicate that visual stimulation via frequency tagging is a promising tool to investigate infants' neural dynamics and visual perception processes. However, many questions have arisen, which should be considered and tested in future research with similar paradigms. For this reason, the main limitation imposed by a small sample size must be improved in order to obtain results which have enough power to make representative assumptions about the influence of social interaction on infants' neural dynamics and visual attention processing. However, the current study and future research will help to gain further insights into infants' attentional processes, thereby contributing to a better and more conclusive understanding of underlying neural processes of the development of attention and social interactions early in life.

Bibliography

Amso, D., & Johnson, S. P. (2006). Learning by Selection: Visual Search and Object Perception in Young Infants. *Dev Psychol, 42*, 1245. https://doi.org/10.1037/0012-1649.42.6.1236

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences, 107*(37), 16048-16053. https://doi.org/10.1073/pnas.1004801107

Butterworth, G., & Itakura, S. (2000). How the eyes, head, and hand serve definite reference. *The British Journal of Developmental Psychology, 18*, 25. https://doi.org/10.1348/026151000165553

Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences, 18*(6), 300-309. https://doi.org/10.1016/j.tics.2014.02.005

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development, 63*(4), i-174. https://doi.org/10.2307/1166214

Christodoulou, J., Leland, D., & Moore, D. (2017, 12/28). Overt and Covert Attention in Infants Revealed Using Steady-State Visually Evoked Potentials. *Developmental Psychology, 54*. https://doi.org/10.1037/dev0000486

Christodoulou, J., Leland, D. S., & Moore, D. S. (2018, May). Overt and covert attention in infants revealed using steady-state visually evoked potentials. *Dev Psychol, 54*(5), 803-815. https://doi.org/10.1037/dev0000486

Cleveland, A., Schug, M., & Striano, T. (2007). Joint attention and object learning in 5- and 7-month-old infants. *Infant and child development, 16*, 306. https://doi.org/10.1002/icd.508

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT press.

Cooper, R. P., & Aslin, R. N. (1990). Preference for Infant-Directed Speech in the First Month after Birth. *Child Development, 61*(5), 1584-1595. https://doi.org/10.2307/1130766

Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language, 25*(2), 141-168. https://doi.org/https://doi.org/10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148-153. https://doi.org/10.1016/j.tics.2009.01.005

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1567), 1149-1157. https://doi.org/doi:10.1098/rstb.2010.0319

Delorme, A., & Makeig, S. (2004, Mar 15). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002, Jul 9). Eye contact detection in humans from birth. *Proc Natl Acad Sci U S A, 99*(14), 9602-9605. https://doi.org/10.1073/pnas.152159999

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007, 2007/05/01). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Feldman, R. (2007, Mar-Apr). Parent-infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *J Child Psychol Psychiatry, 48*(3-4), 329-354. https://doi.org/10.1111/j.1469-7610.2006.01701.x

Fiebelkorn, I. C., & Kastner, S. (2019, 2019/02/01/). A Rhythmic Theory of Attention. *Trends in Cognitive Sciences, 23*(2), 87-101. https://doi.org/https://doi.org/10.1016/j.tics.2018.11.009

Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron, 88*(1), 220-235. https://doi.org/10.1016/j.neuron.2015.09.034

Hendry, A., Johnson, M. H., & Holmboe, K. (2019). Early Development of Visual Attention: Change, Stability, and Longitudinal Associations. *Annual review of developmental psychology, 1*, 275. https://doi.org/10.1146/annurev-devpsych-121318-085114

Hoehl, S., Michel, C., Reid, V. M., Parise, E., & Striano, T. (2014). Eye contact during live social interaction modulates infants' oscillatory brain activity. *Soc Neurosci, 9*(3), 300-308. https://doi.org/10.1080/17470919.2014.884982

Hoehl, S., & Striano, T. (2013). Further Evidence for Continuity in Infants' Joint Attention Development: Commentary on de Barbaro, Johnson, and Deák. *Human development, 56*, 253.

Hoehl, S., & Wahl, S. (2012, 2012/04/01). Recording Infant ERP Data for Cognitive Research. *Developmental Neuropsychology, 37*(3), 187-209. https://doi.org/10.1080/87565641.2011.627958

Imada, T., Carlson, S. M., & Itakura, S. (2013). East–West cultural differences in context-sensitivity are evident in early childhood. *Dev Sci, 16*, 208. https://doi.org/10.1111/desc.12016

Köster, M., Castel, J., Gruber, T., & Kärtner, J. (2017, 2017/12/01/). Visual cortical networks align with behavioral measures of context-sensitivity in early childhood. *NeuroImage, 163*, 413-418. https://doi.org/https://doi.org/10.1016/j.neuroimage.2017.08.008

Köster, M., & Kärtner, J. (2018). Context-sensitive attention is socialized via a verbal route in the parent-child interaction. *PLOS ONE, 13*(11), e0207113. https://doi.org/10.1371/journal.pone.0207113

Kuwabara, M., Son, J. Y., & Smith, L. B. (2011, 2011/10/01). Attention to Context: U.S. and Japanese Children's Emotional Judgments. *Journal of Cognition and Development, 12*(4), 502-517. https://doi.org/10.1080/15248372.2011.554927

Lavelli, M., & Fogel, A. (2005, Jan). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: the 2-month transition. *Dev Psychol, 41*(1), 265-280. https://doi.org/10.1037/0012-1649.41.1.265

Lavelli, M., & Fogel, A. (2013, Dec). Interdyad differences in early mother-infant face-to-face communication: real-time dynamics and developmental pathways. *Dev Psychol, 49*(12), 2257-2271. https://doi.org/10.1037/a0032268

Lee, H., Nand, K., Shimizu, Y., Takada, A., Kodama, M., & Masuda, T. (2017, 2017/10/01). Culture and emotion perception: comparing Canadian and Japanese children's and parents' context sensitivity. *Culture and Brain, 5*(2), 91-104. https://doi.org/10.1007/s40167-017-0052-0

Mangold. (2020). *INTERACT Benutzerhandbuch*. M. I. G. (Hrsg.). www.mangold-international.com

Masuda, T., & Nisbett, R. E. (2001a). Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. *J Pers Soc Psychol, 81*, 934. https://doi.org/10.1037//0022-3514.81.5.922

Masuda, T., & Nisbett, R. E. (2001b, Nov). Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans. *J Pers Soc Psychol, 81*(5), 922-934. https://doi.org/10.1037//0022-3514.81.5.922

Moore, C., & Corkum, V. (1998). Infant gaze following based on eye direction. *British Journal of Developmental Psychology, 16*(4), 495-503. https://doi.org/https://doi.org/10.1111/j.2044-835X.1998.tb00767.x

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

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and Systems of Thought: Holistic Versus Analytic Cognition. *Psychological review, 108*, 310. https://doi.org/10.1037/0033-295X.108.2.291

Norcia, A. M., Gregory Appelbaum, L., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *J Vis, 15*, 46. https://doi.org/10.1167/15.6.4

Nozaradan, S., Keller, P. E., Rossion, B., & Mouraux, A. (2018, 2018/03/01). EEG Frequency-Tagging and Input–Output Comparison in Rhythm Perception. *Brain Topography, 31*(2), 153-160. https://doi.org/10.1007/s10548-017-0605-8

Okumura, Y., Kanakogi, Y., Kobayashi, T., & Itakura, S. (2020, 2020/02/01/). Ostension affects infant learning more than attention. *Cognition, 195*, 104082. https://doi.org/https://doi.org/10.1016/j.cognition.2019.104082

Pauen, S., Birgit, T., Hoehl, S., & Bechtel, S. (2015). Show Me the World: Object Categorization and Socially Guided Object Learning in Infancy. *Child Development Perspectives, 9*(2), 111-116. https://doi.org/https://doi.org/10.1111/cdep.12119

Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annu Rev Neurosci, 35*, 89. https://doi.org/10.1146/annurev-neuro-062111-150525

Reynolds, G. D. (2015, 2015/05/15/). Infant visual attention and object recognition. *Behavioural Brain Research, 285*, 34-43. https://doi.org/https://doi.org/10.1016/j.bbr.2015.01.015

Robertson, S. S., Watamura, S. E., & Wilbourn, M. P. (2012). Attentional dynamics of infant visual foraging. *Proceedings of the National Academy of Sciences, 109*(28), 11460. https://doi.org/10.1073/pnas.1203482109

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Dev Psychol, 47*, 25. https://doi.org/10.1037/a0020740

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology, 18*(9), 668-671. https://doi.org/10.1016/j.cub.2008.03.059

Senzaki, S., Masuda, T., & Ishii, K. (2014). When Is Perception Top-Down and When Is It Not? Culture, Narrative, and Attention. *Cognitive Science, 38*(7), 1493-1506. https://doi.org/https://doi.org/10.1111/cogs.12118

Senzaki, S., Masuda, T., Takada, A., & Okada, H. (2016). The Communication of Culturally Dominant Modes of Attention from Parents to Children: A Comparison of Canadian and Japanese Parent-Child Conversations during a Joint Scene Description Task. *PLOS ONE, 11*(1), e0147199. https://doi.org/10.1371/journal.pone.0147199

Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006, 2006/09/01). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology, 3*(3), 289-299. https://doi.org/10.1080/17405620600879779

Striano, T., Reid, V. M., & Hoehl, S. (2006). Neural mechanisms of joint attention in infancy. *European Journal of Neuroscience, 23*(10), 2819-2823. https://doi.org/https://doi.org/10.1111/j.1460-9568.2006.04822.x

Striano, T., & Stahl, D. (2005, Jul). Sensitivity to triadic attention in early infancy. *Dev Sci, 8*(4), 333-343. https://doi.org/10.1111/j.1467-7687.2005.00421.x

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014, 2014/06/16). Is ostension any more than attention? *Scientific Reports, 4*(1), 5304. https://doi.org/10.1038/srep05304

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Permier, J. (1997, Jan 15). Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. *J Neurosci, 17*(2), 722-734. https://doi.org/10.1523/jneurosci.17-02-00722.1997

Thut, G., Schyns, P. G., & Gross, J. (2011). Entrainment of Perceptually Relevant Brain Oscillations by Non-Invasive Rhythmic Stimulation of the Human Brain. *Frontiers in psychology, 2*, 170. https://doi.org/10.3389/fpsyg.2011.00170

Toffanin, P., de Jong, R., Johnson, A., & Martens, S. (2009, 2009/06/01/). Using frequency tagging to quantify attentional deployment in a visual divided attention task. *International Journal of Psychophysiology, 72*(3), 289-298. https://doi.org/https://doi.org/10.1016/j.ijpsycho.2009.01.006

Tomasello, M., & Farrar, M. J. (1986). Joint Attention and Early Language. *Child Development, 57*(6), 1454-1463. https://doi.org/10.2307/1130423

Turati, C., & Simion, F. (2002, 2002/12/01/). Newborns' recognition of changing and unchanging aspects of schematic faces. *Journal of Experimental Child Psychology, 83*(4), 239-261. https://doi.org/https://doi.org/10.1016/S0022-0965(02)00148-0

VanRullen, R. (2018). Attention Cycles. *Neuron, 99*(4), 632-634. https://doi.org/10.1016/j.neuron.2018.08.006

Vialatte, F.-B., Maurice, M., Dauwels, J., & Cichocki, A. (2010, 2010/04/01/). Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. *Progress in Neurobiology, 90*(4), 418-438. https://doi.org/https://doi.org/10.1016/j.pneurobio.2009.11.005

Wass, S. V., Whitehorn, M., Marriott Haresign, I., Phillips, E., & Leong, V. (2020). Interpersonal Neural Entrainment during Early Social Interaction. *Trends in Cognitive Sciences, 24*(4), 329-342. https://doi.org/10.1016/j.tics.2020.01.006

Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science, 6*(3), 297-311. https://doi.org/10.1111/1467-7687.00286

Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences, 105*(36), 13690-13695. https://doi.org/10.1073/pnas.0804388105

Yu, C., & Smith, L. B. (2016, May 9). The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Curr Biol, 26*(9), 1235-1240. https://doi.org/10.1016/j.cub.2016.03.026

Figure Captions

Table Captions

Abstract (english)

From early infancy, the development of infants' attentional processes is strongly influenced by social interactions and the attentional guidance through their interactional partners. While there is many behavioral evidence on infants' attention processing in current state of research, gaps remain in addressing the neural dynamics of infants' attention processes in situations of attentional guidance, shaped by social interactions. To address these questions, the current study examines infants' neural dynamics of infants' attention distribution altered by socially conveyed preferential styles which guide infant's attention to specific elements within a visual scene. Therefore, eight infants' (11-13 month) were assessed with a flickering image task in a pre-post design with an interaction phase in-between, involving joint attention and ostensive cues which directed attention to either background or object. In order to capture infants' attentional distribution to specific elements of a visual scene, steady-state visually evoked potentials (SSVEPs) were elicited in infants and captured with EEG and depicted elements were tagged with two separate frequencies (frequency tagging). Results revealed that, infants did not show any perceptual preferences style for neither object or background when attention has not been socially influenced and guided towards depicted elements within the visual scene. However, against the expectations, infants' visual processing of the respective elements in the visual scene (object/ background) also did not significantly change due to conveyed preferential perception styles during interaction phase. Implication of these results and limitations of the present study are discussed. All things considered, the present study points to important aspects for the investigation of infants' attention distribution during social interaction and addresses the suitability of frequency tagging as a method to capture infants visual processing of attention allocation. The study contributes to a better and more conclusive understanding of the underlying neural processes of infants' attention during social interactions and provides possible future directions and questions for research on the neural dynamics of infant attentional processing in the context of social interaction.

Abstract (deutsch)

Die Entwicklung der Aufmerksamkeitsprozesse von Säuglingen wird schon früh im Leben stark von sozialen Interaktionen und der Aufmerksamkeitslenkung durch ihre Interaktionspartner beeinflusst. Während es viele verhaltensbezogene Forschungsergebnisse zur Aufmerksamkeitsverarbeitung von Kleinkind gibt, bestehen weiterhin offene Fragen zur neuronalen Dynamik der kindlichen Aufmerksamkeit in Situationen von Aufmerksamkeitslenkung, welche durch soziale Interaktion geprägt sind. Um diese Fragen zu adressieren, untersucht die aktuelle Studie die neuronalen Prozesse der kindlichen Aufmerksamkeitsverteilung, die durch sozial vermittelte, perzeptuelle Präferenzstile beeinflusst wird, welche die Aufmerksamkeit des Kleinkindes auf spezifische Elemente innerhalb einer visuellen Szene lenken. Dafür wurden acht Kleinkinder (11-13 Monate) mittels Flackerbild-Aufgabe in einem Prä-Post-Design, mit dazwischen liegender Interaktionsphase untersucht. Diese beinhaltete eine Situation der gemeinsamen Aufmerksamkeit (engl. joint attention) und ostensive Hinweise, mittels welchen die Aufmerksamkeit des Kleinkindes entweder auf den Hintergrund oder das Objekt gelenkt wurde.

Um die Aufmerksamkeitsverteilung von Kleinkindern auf bestimmte Elemente einer visuellen Szene zu erfassen, wurden Steady-State visuell evozierte Potentiale (SSVEPs) bei Kleinkindern ausgelöst und mit dem EEG erfasst, wobei die dargestellten Elemente mit zwei separaten Frequenzen markiert wurden (engl. frequency tagging).

Die Ergebnisse zeigen, dass die kindlichen Aufmerksamkeitsprozesse, wenn sie nicht sozial beeinflusst und gelenkt wurden, keine Wahrnehmungspräferenz für weder das Objekt noch für den Hintergrund in einer visuellen Szene in ihrer Aufmerksamkeitsverteilung aufweisen. Entgegen den Erwartungen änderte sich die visuelle Verarbeitung der jeweiligen Elemente allerdings auch dann nicht signifikant, wenn ein präferenzierender Wahrnehmungsstil (Objekt Präferenz/ Hintergrund Präferenz) in der Interaktionsphase vermittelt wurde. Implikationen dieser Ergebnisse und Limitationen der vorliegenden Studie werden diskutiert. Insgesamt weist die vorliegende Studie auf wichtige Aspekte für die Untersuchung der Aufmerksamkeitsverteilung von Kleinkindern hin und diskutiert die Eignung von Frequency Tagging als Maß für die Erfassung der Aufmerksamkeitsverteilung bei Kleinkindern. Die Studie trägt damit zu einem besseren Verständnis der zugrundeliegenden neuronalen Prozesse der kindlichen Aufmerksamkeit während sozialer Interaktionen bei und liefert mögliche zukünftige Richtungen und Fragestellungen für die Forschung zur neuronalen Dynamik der kindlichen Aufmerksamkeitsverarbeitung im Kontext sozialer Interaktion.

Appendix

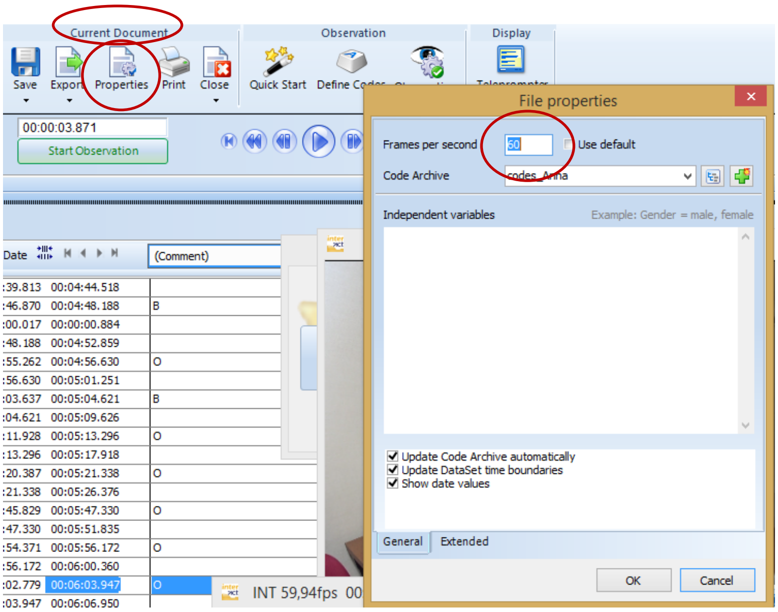
Appendix A:

Video Coding Manual

First step:

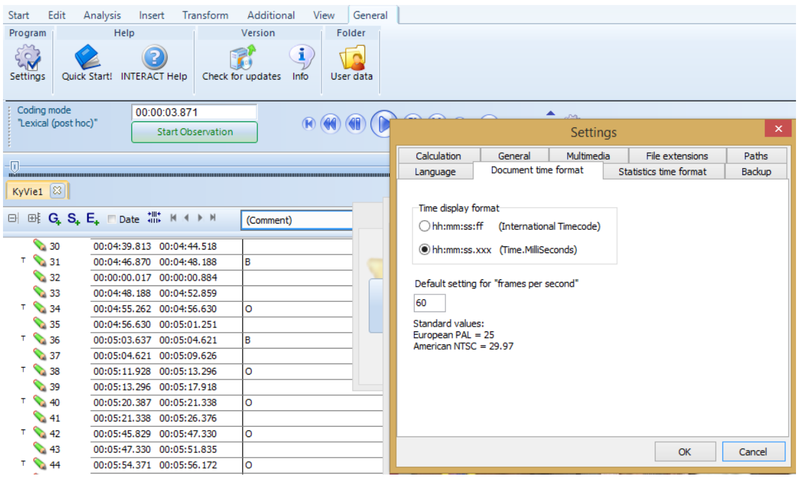
Open Interact, open video + code file, then set the fps to 60

Current Document > Properties > File Properties > Frames per second



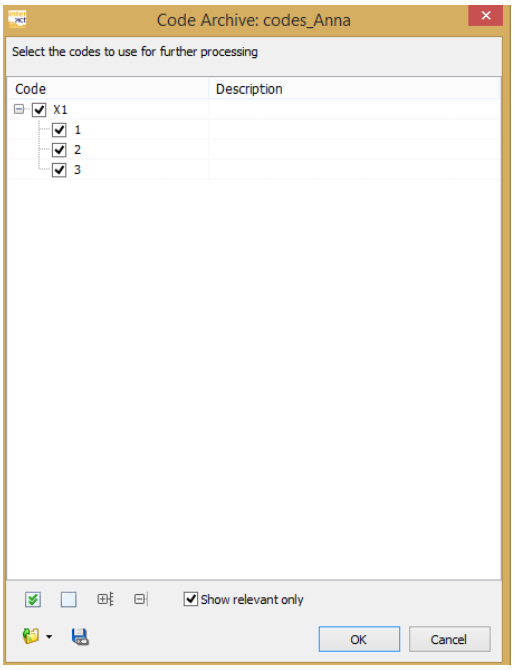
Second step:

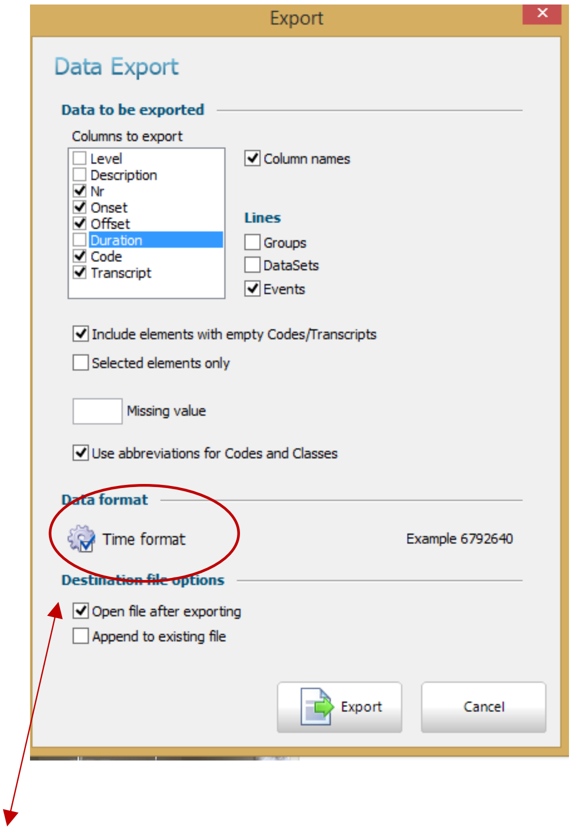
Check setting: General > Program > Settings > Document time format > Time display format: hh.mm.ss.xxx (Time Milliseconds) and also here: > Default setting for "frames per seconds": 60



Third step:

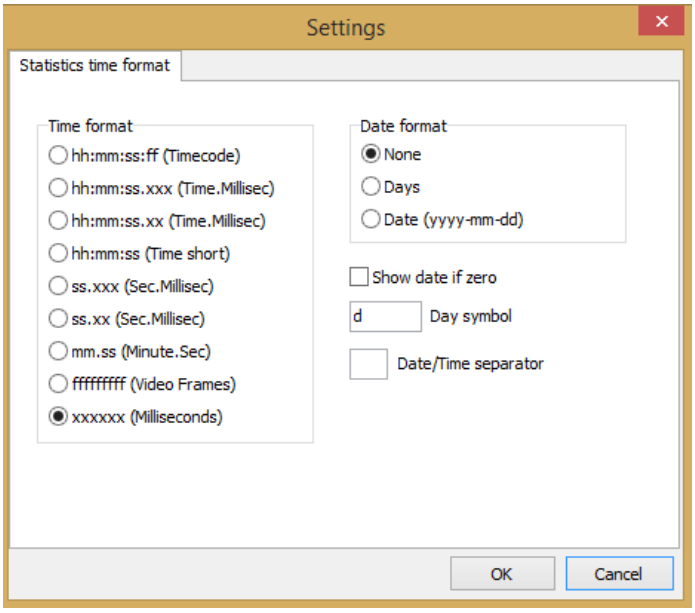
Set the Export details as follows: (Duration could be clicked if you want to check for yourself your trial lengths) Instead of Nr choose Dataline number





Fourth step:

By clicking on Time format on 2nd picture above, check if it is set in Milliseconds.



Fifth step:

Start CODING:

<u>Pre-Phase and Post-Phase:</u> Only using Code 1

For each trial = in Excel table should be 1 row with 2 timestamps in 2 cells next to each other (times for onset of pic AND offset of pic / if the kid looked away before offset then look away and not offset)

1. Press code 1 for onset of pic

2. Again press code 1 for offset of pic (3 seconds later) [If child looked away earlier than offset: press code 1 in that very moment when he or she looked away, and not at offset of pic!]

- These 2 timestamps will be in the same line

- Then comment window will appear. Leave it empty and press enter. OR: if the child looked away before offset of pic add "L" to comment window

- Offset of pic is the last frame when the picture is still visible but a bit darker. Make sure not to code the over-glow. That can make trial length too long. If you are unsure about trial lengths, export your file with durations shown in Excel to see the time passed btw onset and offset.

- If child was not looking during the whole trial: copy-paste onset of pic into the field of offset of pic so that they will be IDENTICAL → this will result in 0 duration later on for that trial → MISSED trial, add comment "M".

- If child was not looking from beginning but looked to the screen within 30 frames and looked until the end ---> add comment "G"

- If child was not looking from beginning but looked to the screen within 30 frames and then again looked away ---> add comment "L" (look away trial)

<u>Interaction Phase:</u> Using code 2 and 3 for each trial, one after the other

For each trial in the Excel table, there should be <u>2 rows</u> with 4 timestamps (onset of pic, onset pointing; onset pointing, offset pic /look away)

<u>For each trial:</u>

1. Press code 2 for onset of pic

2. Again, press code 2 for onset of pointing (moment when finger touches the screen)

- These 2 timestamps will be in the same line
- Leave comment window empty
- Pause video and click on the field of onset of pointing to jump back to that moment in the video and press code 3

3. When you pressed code 3, now you got the exact time for onset of pointing in the first cell of this current row from the second cell of the previous row

4. Again, press code 3 for offset of pic / if child looked away earlier, press code 3 in that moment instead of coding offset pic

- These 2 timestamps will be in the same line
- Type "L" in the comment window if the child looked away before offset of pic
- If the infant is not looking at onset of pointing, we are checking if it starts looking within 30 frames after pointing (then we code onset looking instead of onset pointing (code 3) and add a "G" to the comment window for code 2.
- If child was not looking from beginning but looked to the screen within 3ß frames and then again looked away → add comment "L" (look away trial)
- If child was not looking from the beginning: code offset with same time as onset pointing, so that they will be identical → to have 0 duration later on for that trial → MISSED trial and add "M" into comments
- Typ "L" in the comment window if the child looked away before offset of pic
- If the infant is not looking at onset of pointing, check if it starts looking within 30 frames after pointing (then we code onset looking instead of onset pointing (code 3) and add a "G" to the comment window for code 2.

- If child was not looking from the beginning but looked to the screen within 30 frames and then again looked away → add comment "L" (look away trial)

- If child was not looking from the beginning: code offset with same time as onset pointing, so that they will be identical → to have 0 duration later on for that trial → MISSED trial and add "M" into comments

In the exported Excel table, there should be all together 120 rows: 16 for pre-assessment (CODE 1) + 32 x 2 = 64 (CODE 2 & CODE 3) interaction, and 32 for post-assessment (CODE 1)

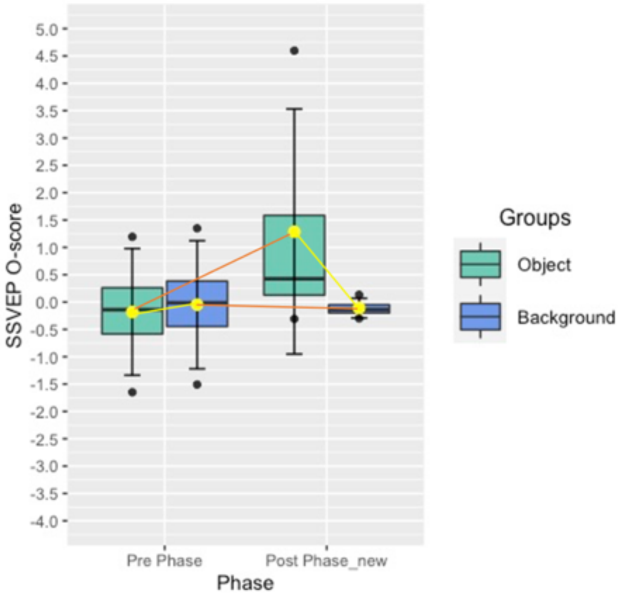
Extra - for Interaction:

Note down the trial numbers for yourself in which the experimenter leaves her finger on the screen for a shorter time than 3 sec for some reason.

Appendix B

Figure 14

SSVEP O-score distribution for the Pre-phase and Post-phase (only new images)



Note. Figure 14 shows the SSVEP O-score changes from the pre-phase for the old images and the post-phase for the new images.