Improving Tools for the Analysis of Brain Based Measures of Infant Information Processing

STETS, MANUELA

Improving Tools for the Analysis of Brain Based Measures of Infant Information Processing

Manuela Stets

Thesis submitted for the degree of Doctor of Philosophy

Durham University, Department of Psychology

2012

Abstract

While current methodological approaches to data-acquisition and -analysis used in ERP-studies have shown to be appropriate for adult populations, their applicability with infant populations is debatable. Many researchers agree that they are, in fact, unsuitable for infants. However, due to a lack of alternatives, traditional ways of data-collection and -analysis originally designed for adults are still used with developmental and clinical populations. The four studies reported in the current thesis propose novel approaches to methodological issues involved in infant ERP-studies. Limiting the number of artefact-free trials included in the average, Study 1 shows that fewer trials from earlier stages of a test-session can convey meaningful information and provide new insights into infant cognition. Additionally, it is shown that large amounts of usable data are often discarded in a traditional data-analysis. Study 2 aimed to provide a general overview on attrition in infant ERP-studies – which often seems to be accepted as a given among developmental psychologists – and to find underlying causes for attrition which may be based on study-design features that are common to all studies. The majority of investigated features did not impact attrition. Therefore, further factors outside the scope of the current consideration are responsible for an increase in attrition. Studies 3 and 4 built on the findings of the first two. Presenting infants with eight experimental conditions, Study 3 illustrates a novel approach to study-design which both yielded more data from individual participants and decreased attrition. Finally, using the data-analysis outlined in Study 1 on the data collected for Study 3, Study 4 both substantiates the claims made in Study 1 and provides further, previously unanticipated insights into infant cognition. These findings illustrate that more infant-suitable strategies for study-design and data-analysis are possible and that they can enhance our knowledge about infant cognitive processes.

Table of Contents

Chapter 1

General Introduction …… 1

Chapter 2 – Study 1

Infant ERP Amplitudes Change Over the Course of an Experimental Session: Implications for Cognitive Processes and Methodology ……. 33

Abstract …..……. 34

Introduction ……... 35

Analysis 1 – Nc and the number of trials contributed to the average ERP per condition

Methods …… 40

Results ……..……. 43

Discussion ……..…… 45

Analysis 2 – Analysis of the first 3artifact-free trials with additional participants

Methods ……..…… 50

Results ……..…. 51

Discussion ……..… 52

Analysis 3 – Does the length of a test session alter the amplitude of the Nc?

Methods ……. 54

Results ……... 57

Discussion ……. 59

General Discussion and Conclusions ……. 60

Chapter 3 – Study 2

A Meta-Analysis Investigating Factors Underlying Attrition Rates in Infant ERP Studies ……. 71

Abstract ……..… 72

Introduction ……... 73

Methods ……..… 79

Results ……...… 88

Discussion ……..... 96

Recommendations for ERP Paradigm Construction ……....102

Appendix 1 – Papers collected for the meta-analysis ……. 111

Appendix 2 – List of factors included in the present meta-analysis ….. 120

Chapter 4 – Study 3

Infants Need More Variety – Increased Data Acquisition with Reduced Participant Attrition in Infant ERP Studies ……... 125

Abstract ……... 126

Introduction ……. 127

Materials …….. 130

Results ……. 143

Discussion ……... 153

Implications ……. 163

Conclusions ……. 166

Chapter 5 – Study 4

The Fine Print of Infant ERPs: The Effect of Additional Trials on Components …… 172

Abstract ……... 173

Introduction ……. 175

Methods ……..... 182

Results …….. 185

Discussion …… 211

Conclusion ……... 217

Chapter 6

General Discussion and Conclusions ……... 223

Declaration

I hereby confirm that no part of the material presented in the following thesis has previously been submitted by me for a degree at this or in any other university. If material has been generated through joint work, my independent contribution has been clearly indicated. In all other cases, material from the work of others has been acknowledged and quotations and paraphrases are suitably indicated.

Acknowledgements

First of all, I would like to thank my supervisory team, Dr. Vincent Reid and Dr. Mike Burt, for providing me with the opportunity of finally completing a PhD in psychology. Vincent's boundless reserves of understanding and patience made me believe that I can really do this. Mike's pragmatic idea on what a PhD is, helped to put the undertaking into perspective. I would also like to acknowledge my family, who always supported me mentally from far away, and all the friends and acquaintances near and far who helped in their own little ways to make this enjoyable, interesting, and successful three years. Among others, the people in room number 70 (Anna, Cath, Dave, Emma, Kat, and Pete) deserve credit for being the good people that they are – brownies, cakes, and cookies will never be the same again without you! Finally, I would like to give a little bit of credit to myself for having grown with respect to both personality and knowledge.

Chapter 1

General Introduction

Until as late as the 20th century, infants were seen as rather passive beings who are born with no prior knowledge of the physical and social world around them (Striano, & Tomasello, 2002). One of the first researchers to acknowledge that this is not the case and that, on the contrary, infants are very receptive and active learners, was Jean Piaget. Based on the observations that he made of his own children's development, he proposed that, through exploring items in their environment, infants refine their already existing thinking-structures as well as their perceptions of the objects and events happening around them (Piaget, 1952, 1954).

A major issue when studying populations of (pre-verbal) infants is to choose the most suitable experimental paradigm. For instance, instructing the infants in a formal and purely verbal manner will prove extremely challenging or entirely impossible (Camaioni, Perucchini, Bellagamba, & Colonnesi, 2004; Hoehl, & Wahl, 2012). Moreover, direct questions requiring oral responses or physical responses such as pressing a button are not suitable for very young populations. Therefore, developmental psychologists working in this field of research usually have to resort to alternative measures of behaviour and the cognitive processes underlying them. For example, physical responses such as pointing gestures are well-suited for use in studies with slightly older infants. Liszkowski, Carpenter, Striano, and Tomasello (2006) and Liszkowski, Carpenter, and Tomasello (2007) investigated the pointing gestures made by participants as young as 12 months. However, this test-method is also only applicable to a certain degree and will likely not yield meaningful results very much earlier than before an infant's first birthday, as typically developing infants start to use deictic gestures only from around 10 months onwards (Camaioni, 1997).

Developmental researchers have designed other behavioural research-tools in order to be able to study the period of the first 12 months of typical development. One

frequently used behavioural paradigm is to measure the infants' looking-times (e.g., Cleveland, & Striano, 2007; Symons, Hains, & Muir, 1998). For such studies, coders record the duration of a participant's gaze towards a particular stimulus compared to another one. This methodology has been used successfully with infants as young as one month of age (Striano, Henning, & Stahl, 2005) and even in newborns (Bushnell, Sai, & Mullin, 1989; Pascalis, de Schonen, Morton, Deruelle, & Fabre-Grenet, 1995). Alternatively, when studying older infants and toddlers, instead of recording the duration of a participant's gaze towards a stimulus, a reaching action could be coded (e.g., the 14-, 18-, and 24-month-olds presented in Behne, Carpenter, & Tomasello, 2005). Another behavioural study-design that is often used with populations aged six months and older is the head-turning paradigm. This methodology was designed to be used with visual stimuli but has mainly been used with auditory stimuli instead (Kuhl, 1979; see Werker, Polka, & Pegg, 1997, for a review). In this case, infants are familiarised with certain sounds and are then required to turn their head in the direction of a loudspeaker when a change in a stream of auditory stimuli is detected. Finally, another example for a methodology mainly used with very young infants is high-amplitude sucking (e.g., Floccia, Christophe, & Bertoncini, 1997; Sansavini, Bertoncini, & Giovanelli, 1997). With such a paradigm, even newborns' learning and stimulus discrimination behaviour can be investigated. The infant is repeatedly presented with a stimulus until habituation to the stimulus is manifest as indicated by a decline in sucking rates in response to the stimulus. Then, while one group of participants will be further presented with the habituated stimulus (i.e., the control group), a second group of infant will be presented with a novel stimulus. If there is a significant difference between the sucking rates of the experimental and the control group, this is seen as an indication that the change in stimuli has been detected by the

infants (Floccia, Christophe, & Bertoncini, 1997). Again, such studies are usually conducted with auditory stimuli.

The main difference between paradigms such as these and studies based on the assessment of neurophysiology is that inferences about cognitive development must be drawn from obvious, visible behaviours and responses. Therefore, the estimates for the onset-times of certain behaviours that require the understanding of social processes might differ greatly depending on the paradigms used by the researchers. For instance, Tomasello (1995, 1999) claims that, before an infant is nine months old, he or she will broadly not be able to understand someone else's actions as being intentional (e.g., eye-gaze direction). However, recent research using non-invasive, but technically advanced methodologies such as eye-tracking, Magnetoencephalography (MEG), Near-Infrared Spectroscopy (NIRS), or Event-Related Potentials (ERPs) can help to unveil and specify the accuracy of such theoretical assumptions that were made based on overt behaviour (see also Reid, & Geangu, 2008).

One of the methodologies mentioned above which is currently becoming increasingly popular due to the apparent ease with which it can be applied to infant participants, is eye-tracking. This technique relies on recordings of infra-red light which is reflected by the eyes' cornea. When an individual looks at the same location continuously for a certain amount of time, the eye-tracker will register this behaviour as a fixation to that location. Short and rapid eye-movement between fixations are known as saccades (Salvucci, & Goldberg, 2000). Consequently, as the reflections on the cornea will change depending on the eyes' movements, the eye-tracker will calculate the gaze-shifts based on the changes in the recorded reflections (Gredebäck, Johnson, & von Hofsten, 2010). In their eye-tracking study involving groups of six- and 12-month-old infants, Gredebäck and Melinder (2010) showed that the younger

group of participants responded differently to the stimuli compared to the older group. Watching short clips of feeding actions that could either progress in a conventional manner or in an unconventional way, the 12-month-olds made anticipatory saccades towards an actress' mouth before the food, fed to her by another actress, arrived there. However, these infants only made such anticipatory saccades in the condition presenting the conventional feeding action. In the experimental condition in which the feeder placed the banana on the back of the beneficiary's hand instead of putting it into her mouth (i.e., the unconventional feeding action), the 12-month-olds followed the feeder's arm-movement with their gaze instead of making an anticipatory saccade (Gredebäck, & Melinder, 2010). The group of six-month-olds did not show such saccades in either of the two conditions and only showed the tracking of the arm-movement with their gaze. This discrepancy was interpreted as being dependent on the infants' amount of experience with being fed themselves. However, when presenting groups of six- and 10-month-olds with an actress who was feeding herself, both age-groups showed anticipatory saccades towards the mouth of the actress (Kochukhova, & Gredebäck, 2010). The difference between the performances of the six- and the 10-month-old infants lay in the speed with which the saccades were made. In this case, the 10-month-olds' saccade-latency was comparable to that of an adult control group (Kochukhova, & Gredebäck, 2010). Again, the performance-differences were seen as an indication for the amount of experience the different groups of participants had with the action of being fed. The most important finding was however that, based on the anticipatory saccades found, already at six months of age, infants can draw conclusions on another person's intentions. However, there are methodological restrictions related to eye-tracking studies that limit its use for answering certain theoretical questions. For instance, due to the technique's nature, it can only be used to

test the visual modality. Moreover, in order to make statements about infant cognition, inferences have to be made based on obvious behaviour and direct cortical measures are not available which, furthermore, makes it impossible to make statements about the cortical source for a certain response. This means that the theoretical statements that will be presented in a research article are interpretations based on inferences and can be challenged by other researchers interpreting the data in a different way.

Another non-invasive neurocognitive methodology that is becoming increasingly popular among developmental psychologists is Near-Infrared Spectroscopy (NIRS). Compared to fMRI-paradigms for which especially young participants often have to be sedated in order to avoid unwanted movements (see Altman, & Bernal, 2001), NIRS can be used when infants are awake and alert. Moreover, the infants are allowed to move to a certain degree and, additionally, no noise is emitted from the device (Lloyd, Blasi, & Elwell, 2010). For this methodology, light is sent out from a sensor in a head-mounted NIRS-set through the participant's skin, skull, and brain-tissue. Passing through the infant's head, the light will be deflected and/or absorbed, among others, by the oxygen in the blood in the brain. A second sensor then detects and records the remaining wavelength of the light after passing through the tissues (for a review see Lloyd-Fox, Blasi, & Elwell, 2010). For instance, in their study with groups of four- and seven-month-olds, Grossmann, Oberecker, Koch, and Friederici (2010) could show developmental differences between these two age-groups using NIRS-technology. Presenting their participants with either vocal sounds such as non-/words or non-vocal sounds such as birds and cars, Grossmann et al. (2010) found differences in the blood-oxygen levels between these two developmental stages. Whereas the responses recorded from the group of seven-month-old infants showed an increase in the level of oxygen in the blood at

three recording sites for vocal stimuli, such effects could not be found in the group of four-month-olds. The authors interpreted this as an indication that already at the age of seven months, compared to non-vocal stimuli, a special area of the brain is dedicated to the processing of vocal stimuli whereas this is not yet the case in the younger age group (Grossmann et al., 2010). Surprisingly, their younger participant-group showed increased levels of oxygenated blood on one recording site for non-vocal stimuli – an effect that did not appear in the group of seven-month-old infants. These results form the younger age-group were interpreted as illustrating (a) the infants' discrimination of the different classes of stimuli and (b) the immaturity of the cortical processes leading to this discrimination (Grossmann et al., 2010). However, as seen for eye-tracking studies, NIRS-studies also have some methodological issues. Despite the spatial resolution being relatively good as the data that are collected originate from very closely defined cortical areas, the temporal resolution of the technique is less optimal (see Figure 2 in Lloyd-Fox, Blasi, & Elwell, 2010, p. 271). Moreover, considering the setup of the sensors illustrated in section (d) of Lloyd-Fox, Blasi, and Elwell's (2010) Figure 1, large areas of the skull and, ultimately, the brain remain uncovered and meaningful data may remain undetected. Further, at this stage in time there are relatively few NIRS studies with infants. This makes the interpretation of data difficult unless other sources of information are also relied upon, such as EEG-data or fMRI.

Finally, for several decades, researchers are using Event-related potentials (ERPs) to study infant cognitive development. ERPs are the epoched and averaged extracts from an individual's electro-encephalograms time-locked (mainly) to the onset of the presentation of a specific stimulus (DeBoer, Scott, & Nelson, 2007; Hoehl, & Wahl, 2012; Reid, & Geangu, 2008). Parallel to eye-tracking and NIRS, ERP-methodology is a non-invasive technique as it allows for an easy application with

infant populations (Männel, 2008; Nelson, & Monk, 2001). Electrodes continuously record the electrical activity that is produced by a human brain in response to a stimulus-presentation. They can be placed on a participant's skull either individually, by using an elastic cap, or a net of sensors. After passing through an amplifier which magnifies the signal detected by the electrodes on the participant's skull, the recorded data can be processed using special software that can either come with the recording-system (e.g., NeuroScan or NetStation) or is freely accessible online for download (EEGLab). Similar to behavioural responses that frequently occur under a specific experimental conditions, if an ERP-component is repeatedly occurring in response to a specific kind of stimulus, then it can be assumed that this component is genuinely involved in the cognitive processing of such stimuli (Otten, & Rugg, 2005).

Another reason for the popularity of the ERP-methodology among developmental psychologists is that it can be used to study a wide range of cognitive processes in response to a multitude of different stimuli (see Hoehl, & Wahl, 2012 for a brief review). The nature of the stimuli can be purely auditory (e.g., Andrews-Espy, Molfese, Molfese, & Modglin, 2004; He, Hotson, & Trainor, 2007, 2009a, 2009b) or purely visual (e.g., Braddick, Birtles, Wattam-Bell, & Atkinson, 2005; Hoehl, Wiese, & Striano, 2008) or a combination of the two (e.g., Friedrich, & Friederici, 2008, 2010; Hirotani, Stets, Striano, & Friederici, 2009). Additionally, particularly when the stimuli are of a visual nature, researchers have the choice to create either non-animated or animated stimuli (e.g., Hirai, & Hiraki, 2005; Kaufman, Csibra, & Johnson, 2005). Bearing this multitude of possibilities for investigating different areas of infant cognitive development combined with the relative ease of application in mind, it is easy to understand the popularity of the methodology. Compared to this, eye-tracking studies are also very easy to administer and can answer a wide range of research-

questions. However, as mentioned earlier, this methodology is restricted to processes involving visual stimuli. Areas of research touching matters such as word-learning or syntax for instance, could only be investigated with groups of individuals who have already learned to read (e.g., Sekerina, Stromswold, & Hestvik, 2004; Staub, 2010).

There are a few physical aspects of ERP-components that are frequently included in analyses of the waveforms when researchers try to relate event-related potentials to the processing of visual and/or auditory stimuli. Firstly, the peak- or mean-amplitude/s of the ERP-wave/s can be compared for two or more experimental conditions within a certain time-window after a stimulus-onset (e.g., Hoehl, Reid, Mooney, & Striano, 2008). For instance, Hoehl, Palumbo, Heinisch, and Striano (2008) investigated the Negative component (Nc) in groups of seven-month-old infants. The Nc is a negative peak occurring around 400 ms after the onset of a stimulus and it has repeatedly been related to attention-allocation (see also Hoehl, Reid, Mooney, & Striano, 2008; Richards, 2003). In their first experiment, Hoehl, Palumbo, Heinisch, and Striano (2008) presented their participants with images depicting adult faces displaying either a fearful or a neutral facial expression looking towards a toy which was located at eye-level next to the face. In order to investigate the processing differences between the experimental conditions and following standard analysing-procedures frequently used in infant ERP-studies, Hoehl et al. (2008) first manually inspected the participants' EEG-data for artefacts and discarded contaminated trials. Then, they averaged the artefact-free trials per condition for each of the participants and excluded data from infants, who did not provide minimally 10 artefact-free trials for both conditions, from further analyses. Finally, the authors established the most negative amplitudes for the conditions in the time-window between 400 and 600 ms after the onset of the stimuli, and compared the averages of

minimal amplitudes provided by their participants for the two experimental conditions in a paired-samples *t*-test (Hoehl, Palumbo, Heinisch, & Striano, 2008). This analysis revealed significant differences in the Nc-amplitudes between the two experimental conditions with the Nc-peak amplitude being more negative for faces displaying a fearful expressing compared to the faces showing a neutral expression. The authors interpreted these results as an indication that, at seven months of age, infants are already sensitive to the emotion conveyed by a facial expression and can use such information to decide what type of stimulus to attend more closely (Hoehl, Palumbo, Heinisch, & Striano, 2008). In a follow-up study which was presented in the same research article and in which the actors also displayed either a fearful or a neutral facial expression but did not look at the toy depicted next to the head, the authors did not replicate this effect. Therefore, Hoehl et al. (2008) concluded that the adults' fearful facial expression in combination with the gaze being directed at a novel toy serves as an indicator for the infants whether the unfamiliar object could potentially be dangerous and pose a threat.

Another aspect of ERP-components which are frequently investigated are the latencies of the amplitude-peaks which are seen as indicators for the effort needed to process a stimulus (Rugg, & Coles, 1995). For example, Purhonen, Kilpeläinen-Lees, Valkonen-Korhonen, Karhu, and Lehtonen (2005) found significant condition-dependent differences in the latencies of four of the five ERP-components they were investigating. Presenting their four-month-old infants with words spoken either by their mothers or by an unfamiliar female, the authors interpreted the shorter processing-times in the condition with the voice of the mother as an indication for heightened attention to these stimuli. Putting this into an evolutionary perspective, Purhonen et al. (2005) further argued that increased attention to one's mother's voice

was of importance when considering the individual's survival. In their group of eight-month-old infants, Scott, and Nelson (2006) found significant differences between conditions on the latency of the N290 but not on its amplitude when presenting their participants either with unchanged faces they had been familiarised with or with faces they had never seen before. The data of their group of eight-month-olds showed longer latencies for the peak of the N290 for the familiarised faces compared to the novel ones. However, no differences had been found in the latencies between the familiar and the unfamiliar stimuli when compared to those presenting faces the infants had been familiarised with but which were slightly changed in a featural or in a configurational aspect (Scott, & Nelson, 2006).

Finally, the location of the channel/s or channel-groups at which a component(s) appeared is also often analysed and reported (e.g., Mills, Plunkett, Prat, & Schafer, 2005). Significant effects of location can occasionally appear in conjunction with an effect of condition (e.g., Scott, and Nelson (2006). Differences were also seen in processing-speed between the hemispheres (i.e., the right hemisphere showing a later N290-peak compared to the left hemisphere). In the data from their group of eight-month-olds, the N290 also peaked later in the channel-group over the right hemisphere for the faces the infants had been familiarised with compared to the novel ones (Scott, & Nelson, 2006). In a study comparing ERP-responses collected from a group of eight-month-olds and from a group of adults in response to standard or deviant syllables, Pang, Edmonds, Desjardins, Khan, Trainor, and Taylor (1998) found differences in the location of the mismatch negativity depending on the participant's age. Whereas their group of adults showed mismatch responses over wide areas of both hemispheres and on central channels, the same responses could only be found in two channels over the left hemisphere in the infant population (Pang et al., 1998).

Studies comparing cognitive responses to the same stimuli from different developmental groups can provide valuable insights into potential changes in the cognitive processes underlying human behaviour. Moreover, given that EEG-methodology can be used with both very young participants such as newborns as well as with toddlers, older children, and adults, this enhances the comparability of results between age groups when contrasted with other methods, such as looking time measures. Therefore, studies such as Pang et al. (1998) or Brannon, Libertus, Meck, and Woldorff (2008), in which responses from infant populations and from adult groups are presented side by side, need to make the basic assumption that these different developmental groups have the same physical properties and that the respective ERPs will have been generated and recorded under the same conditions. However, this assumption cannot be met as there are differences between the physiological development of brains at different developmental stages; e.g., between infants and adults (e.g., Casey, Tottenham, Liston, & Durston, 2005; Reid, & Geangu, 2008). Whereas adult brains are highly myelinated, neuronal connectivity is less dense in infants' brains (e.g., DeBoer, Scott, & Nelson, 2007; Matsuzawa, Matsui, Konishi, Noguchi, Gur, Bilker, & Miyawaki, 2001). Moreover, while adults have a closed and thick skull, the fontanel in an infant's skull is not entirely closed until around 18 to 24 months of age (Duc, & Largo, 1986). Therefore, as Reid and Geangu (2008) stated, it is safer to assume that cognitive responses towards the same stimuli collected from adults and from infants will probably not be alike but may be correlated in a manner that is most likely unknown.

Methodological literature providing advice on conducting EEG-/ERP-studies in general (e.g., Luck, 2005; Picton et al., 2000) and on studies with infant populations (e.g., (DeBoer, Scott, & Nelson, 2007; Hoehl, & Wahl, 2012) are plentiful. Picton et

al. (2000) and Hoehl and Wahl (2012) acknowledge that testing strategies need to be adjusted to the needs of infant populations and that paradigms used with adults cannot be simply applied with infants. In their guidelines, Picton et al. (2000, p. 129) specifically state that "[…] tasks need to be of shorter duration for clinical and developmental studies than for ERP studies in normal young adult […]". Hoehl and Wahl (2012) advise researchers to take care not to present the infants with too many different experimental conditions. Moreover, the authors state that the attrition rate experienced in a study using three or four experimental conditions is likely to increase to a large extent as well. Additionally, research articles reporting on studies featuring relatively high attrition rates commonly contain short explanations stating that this might be due to high task demands or very strict criteria for data-inclusion (e.g., de Haan, Pascalis, & Johnson, 2002; Parise, Reid, Stets, & Striano, 2008). In general, attrition rates between 50 and 75% are considered as being relatively normal or typical and as to be expected when testing infants (e.g., DeBoer, Scott, & Nelson, 2007). As a comparison, Lloyd-Fox et al. (2010) state that NIRS-studies with infant populations typically feature attrition rates of about 40% which is slightly lower than what is usually reported and expected for infant ERP-studies. In contrast, the attrition rates reported for infant eye-tracking studies tend to be much lower (around 10%) when compared to the other two methodologies (e.g., Gredebäck, & Melinder, 2010; Hespos, Gredebäck, von Hofsten, & Spelke, 2009).

The most frequently mentioned reason for (high) attrition rates and restrictions on paradigm-design is the infants' limited attention-span (see DeBoer, Scott, & Nelson, 2007; Hoehl, & Wahl, 2012; Picton et al., 2000; Reid, & Geangu, 2008). This concept seems to be generally accepted among developmental researchers and has, likely, never been challenged. It is commonly assumed that, if the load on an infant's

cognitive system is too high – due to too many experimental conditions for instance –, the infant will enter a state usually termed "fussy". This state will become apparent to a researcher when the participant starts to move around agitatedly, yawns, vocalises, or shows any other possible signs of distress. However, a concrete definition of "fussiness" which is followed by all developmental researchers, does not exist. This issue has been mentioned repeatedly in methodological literature and suggestions have been made, but, so far, consensus has not been found (see also Hoehl, & Wahl, 2012). Consequently, the decision, if an experimental session should be interrupted or needs to be discontinued (maybe even in its earliest stage), is very subjective. It lies solely at the discretion of the researcher in the respective test-session. Making this decision requires extensive training and much experience on the researcher's side as well as a feeling for the infants (see DeBoer, Scott, & Nelson, 2007). There is always a chance that an infant will reduce negative affect to the point where the test-session can be continued – even if only for a few more trials. However, a less-experienced researcher who has not been trained sufficiently and, therefore, cannot draw on a wide repertoire of intervention-strategies when an infant shows the first signs of distress, is more likely to discontinue a test-session at an early stage and not to try again after a short break.

Apart from these methodological difficulties related to the preparation of an infant ERP-study and to conducting of the test-sessions, there are also several issues with respect to data-analysis. Both Luck (2005) and Picton et al. (2000) give advice on how many artefact-free trials are necessary per experimental condition from each participant in order to obtain clean and meaningful averages (see also Hoehl, & Wahl, 2012). In contrast to Luck (2005), Picton and colleagues (2000) do not state specific numbers for the lowest possible threshold of usable trials per condition. Rather, it is

merely stated that the number of presented trials should ideally be twice or three-times as high as compared t that typically used in ERP-studies with adults. This claim is supported by the argument that the number of trials that will need to be rejected for various reasons will be much higher in such young populations (Picton et al., 2000). Luck's (2005) article, which is mainly addressed at researchers conducting studies with adult participants, provides the reader with very specific instructions on this issue. He explains that – depending on the typical size of a specific ERP-component – the number of artefact-free trials needs to relate to the remaining noise in the signal in such a way that the following equation equals 0.4 in the worst case and 10.0 in the best case (Luck, 2005, p. 30):

$$(1/\sqrt{N}) \times R.$$

In the above equation, *N* stands for the number of artefact-free trials which will be included in an individual's average per condition. The variable *R* represents the noise that is left in a single trial after filtering, re-referencing, and automatically rejecting contaminated trials from the data. (Here, the notion of noise relates to randomly occurring frequencies that will be different and differently distributed over a time-window from trial to trial within one experimental condition. They are in contrast to the more stable frequencies representing the ERP-component/s (e.g., Repovš, 2010).) These random frequencies will be Furthermore, Luck (2005) explains the equation using the example of an ERP-component with a typical amplitude-size of 20 μV and an assumed amplitude-size of the noise on one trial of 50 μV. Based on these assumed values, the signal-to-noise ratio for this one trial would be

$$20 / 50 = 0.4.$$

Adding data from more artefact-free trials with the same or a very similar amplitude for the component as well as different levels of noise in the individual trials, the signal-to-noise ratio will slowly improve in favour of the ERP-average. However, as Luck (2005) states, already for adult populations the number of artefact-free trials required to achieve a perfect signal-to-noise ration of 10.0 (i.e., that the ERP is absolutely pure and no noise is left in the signal) is extremely high. For the above example of an ERP-component with a typical amplitude of 20 μV, 625 usable trials would be needed in each experimental condition (Luck, 2005). Furthermore, in accordance with Picton et al. (2000), he stated that, when conducting studies with non-adult or clinical populations, twice or three-times as many artefact-free trials will be needed to obtain a clean average. As Luck (2005) admits himself, ensuring that such high numbers of usable trials will be collected per condition from an adult participant already seems very challenging. Based on the above-mentioned obstacles such as a shorter infant attention-span as compared to adults, the probability that an infant will provide that many analysable trials per condition will need to be estimated as being close to zero.

One infant ERP-study reported to have been able to analyse 1186 on average out of maximally 1600 trials for an experimental condition (see He, Hotson, & Trainor, 2007). However, this particular study used very short auditory stimuli (stimulus-duration = 600 ms) presented in very short intervals (duration of interstimulus-intervals = 200 ms). In comparison, typical infant ERP-studies using visual stimuli such as images of faces portraying fearful or neutral emotional expressions (e.g., Hoehl, Wiese, & Striano, 2008) or successions of images showing

three different stages of a self-feeding action with either an expected or an unexpected outcome (see Reid, Hoehl, Grigutsch, Groendahl, Parise, & Striano, 2009) typically feature stimulus-durations of 1000 ms and often equally long or slightly longer inter-stimulus intervals. Moreover, He et al. (2007) used an odd-ball paradigm with the stimuli belonging to the standard category being presented with a frequency of 80% and those belonging to the odd-ball category being presented with a frequency of 20%. At this stage, it remains to be seen whether basic differences between paradigms have any influence on methodological issues such as the applicability of a certain kind of paradigm with a specific group of participants or attrition.

As was described earlier with respect to the question when a test-session should be interrupted or terminated, a similar problem arises when data are prepared for statistical analysis. Two basic steps of data-preparation are filtering and re-referencing. When an EEG is filtered, frequencies above and/or below a certain pre-determined threshold will be deleted from the EEG-stream (see Reid, & Geangu, 2008). During data-acquisition EEGs are typically referenced to the vertex electrode, which is located at the crossing of imagined lines running (a) front to back from the nasion to the inion and (b) left to right from ear to ear according to the international 10-20-system developed by Jasper (1958). As a second step during data-preparation, the EEG is then re-referenced to either the average of all electrodes or to one or two specific electrodes which are typically far away from the scalp-area and channels where the cortical responses are expected to appear for a certain kind of stimulus (e.g., Reid, & Geangu, 2008). Finally, when adult EEG-data are prepared for statistical analysis, contaminated trials are typically removed from further analysis by way of general algorithms. These are preinstalled in the respective software-packages used by

the various labs and will automatically reject any trials that fit specific predetermined criteria (e.g., an eye-blink or other muscle artefacts in the data).

It seems safe to apply such practices with data collected from typically developing adult participants as they can be clearly instructed and should, therefore, be expected to blink only at certain times during the stimulus-presentation (i.e., in the inter-stimulus interval) and to keep any physical movements to an absolute minimum. Infants, however, will not benefit from such instructions (see DeBoer, Scott, & Nelson, 2007). Therefore, infant EEGs are typically inspected visually and contaminated trials are rejected manually. However, as it stands, the guidelines as to which trials need to be rejected and which could be kept despite possible small imperfection on not to be analysed channels are relatively lenient. They seem to be closer to rules of thumb than to strict regulations. Again, practices may differ from lab to lab and, in fact, are likely to differ between individual researchers within a lab. This means that the decision for including or excluding trials (and, consequently, participants) is entirely up to the individual who is manually rejecting the data and preparing them for analysis. As is the case with the proceedings in the test-sessions, these are highly subjective decisions to make and require much experience and skill. In behavioural studies, it is common practice to have a second researcher code about 20% of the data again (e.g., Geangu, Benga, Stahl, & Striano, 2010; Liszkowski, Carpenter, Striano, & Tomasello, 2006). This is performed as an attempt to ensure that the scores recorded by the first scorer are reliable. Considering the above-mentioned inter-individual differences in researchers' strategies to data-inspection and trial-rejection, it seems necessary to investigate the currently used strategies, to find the largest discrepancies and to install guidelines which would ideally be used by all researchers working with infant EEG-data. However, the test-sessions and data-analyses for numerous (infant) research

studies – both behavioural and ERP – are in fact performed by trainees who likely do not have extensive training and much experience (see also Hoehl, & Wahl, 2012). Again, given the inter-individual differences between rejecting-strategies of experienced coders, it seems to indicate this practice might merely increase the likelihood for errors to be introduced into the data.

In the current thesis, I aim to highlight and discuss methodological issues in infant ERP-studies such as the ones outlined previously. Such challenges for developmental researchers using EEG-methodology are well-known in the research community. However, despite various attempts to improve the situation – especially with respect to data-collection – the challenges have remained mainly unresolved as of yet. This is clearly indicated by the lack of propositions for new strategies in the related methodological literature. As has been stated above, the existing sources mainly provide advice on the acquisition of adult EEG-data in an experimental session and only briefly refer to more challenging populations. Alternatively, methodological literature targeted at developmental psychologists working with infants, typically describe the current state-of-the-art strategies, especially those used in the respective labs but do not propose new approaches to the existing problems (e.g., DeBoer, Scott, & Nelson, 2007).

In the first study (see Chapter 2), the behaviour of a prominent ERP-component, the Negative component (Nc), is examined over the course of a test session using a novel approach to statistically analysing infant ERP-data. The data used in this study had been previously analysed and reported in Hoehl, Reid, Mooney, and Striano (2008) and originated from a group of 17 out of 64 four-month-olds. These infants had been presented with two experimental conditions of which one depicted a situation in which an adult was looking toward a toy while, in the second condition,

the person averted their gaze from the toy. The participant's data had been analysed according to conventional strategies using 10 artefact-free trials as a minimal criterion for including an infant's data into the analyses. Then, the negative peak-amplitudes in the time-window between 400 and 600 ms after stimulus-onset were compared for the two conditions (Hoehl et al., 2008). Additionally, the authors had compared the mean-amplitudes between 700 and 1000 ms after stimulus-onset. However, in my re-analysis, for reasons of simplicity, I decided to test this novel approach to data-analysis on the data for the Nc-amplitude only. In order to establish if the rationale for only including infants contributing minimally 10 artefact-free trials per condition into a study's statistical analysis is justified, I prepared data-sets including the first three, five, seven, eight, nine, and 10 usable trials per condition and conducted the same statistical tests as those presented in Hoehl, Reid, Mooney, and Striano (2008). In case my statistical analyses including fewer trials yield the same results as those presented in the original publication, then the rationale for having a minimum criterion for infant ERP-studied would lose much of its justification. Furthermore, it could then be argued that current assumptions about the acquisition of infant ERP-data need to be re-evaluated. Finally, this would allow for shifts in methodology which can potentially help to solve the issues related to collecting a sufficient amount of data from infant participants mentioned earlier. This would be especially helpful with respect to the duration of the test-sessions and the infants' limited attention-span. If fewer trials are necessary to achieve the same results as a conventional data-analysis including minimally 10 artefact-free trials, then test-sessions could be shorter and, consequently, more suitable for infant populations.

The second study presented in this thesis (see Chapter 3) is aiming to provide a more general overview over the different strategies and study-designs which have been

used in infant ERP-studies over the last 34 years and, particularly, their effect on the attrition rates of the respective studies. Therefore, this study will be addressing more general issues related to study-design and its potential impact on the progress of an individual test-session. Parallel to the meta-analysis on behavioural paradigms presented in Slaughter and Suddendorf (2007), information on 34 study-features such as which EEG-system had been used to record the data, the duration of the inter-stimulus interval, and the number of experimental conditions were collected for 149 published research articles. These publications could be split up further into 314 experimental groups which means that some articles either presented data from experimental groups belonging to the same age-group but experiencing slightly different conditions or, alternatively, that the groups represented different age-groups but were presented with the same stimuli. As a first step, a meta-analysis of 181 of the originally collected 314 experimental groups was conducted, providing a general overview over the 34 study-features under investigation. The remaining 133 groups had to be excluded from further analyses due to a lack of information on the attrition rates experienced in these groups. In a second step, a meta-regression was run in which each of the factors was removed one at a time to determine their impact on the studies' attrition rates. These results could give an insight into which strategies and study-designs might be easier for infants to cope with compared to others and which cause infants to attrition more easily than others. This knowledge could help developmental psychologists – as well as the wider research community – to design infant ERP-studies such that more data could be obtained from less distressed groups of infants. Additionally, lower attrition rates could be achieved.

As a consequence of the findings from the meta-analysis described in Chapter 3, the empirical study presented in Chapter 4 aims to further explore the impact of a

study's design on the data to be obtained. This is referring both to the data's quality and quantity. As has been claimed repeatedly (e.g., DeBoer, Scott, & Nelson, 2007; Reid, & Geangu, 2008), infants' attention-span is very limited. In addition to researchers' eagerness to achieve an average which is as free of noise as possible (i.e., obtaining sufficient numbers of artefact-free trials per condition to have a cleaner average; see Luck, 2005), this limitation of attention typically causes researchers to present very few experimental conditions. Typically, infant ERP-studies involve two conditions which are represented by stimuli that are only slightly different from each other. Chapter 4 presents a study in which a group of 18 12-month-old infants were presented with eight experimental conditions that originated from three previously published research articles. The original three studies were theoretically unrelated with respect to the cognitive processes that were investigated. Two of the eight experimental conditions, a gap-/no gap-task, originated from Csibra, Tucker, Volein, and Johnson (2000) who had presented a group of one-year-olds with this task. Following Halit, de Haan, and Johnson (2003), four of the eight conditions used in the empirical study (see Chapter 4) presented the 12-month-olds with human or monkey faces shown in an upright or inverted orientation. Finally, the last two of the eight conditions originated from Hoehl, Reid, Mooney, and Striano (2008) and depicted the social situation described earlier; i.e., an adult who is either looking toward or away from a toy. Importantly, the different experimental conditions were not presented in a block-design that would have caused them to appear as still being presented as separate studies. As this study aimed to explore the effects of study-design on both the infants' behaviour and the data, the eight conditions were presented intermixed with each other as if all presented stimuli belonged to one study. As minimally 10 artefact-

free trials could be obtained for all of the conditions, the data were analysed in a conventional manner including 10 or more trials into the analyses.

Finally, Study 4 (see Chapter 5), a re-analysis of the data collected in the third study, is mainly based on the findings of the Study 1 (see Chapter 2). Moreover, I aim to combine and further the insights gained in studies 1 and 3 in order to lend support to these previous findings. Therefore, in an attempt to replicate the findings from Study 1 and to investigate if they will hold for ERP-components other than the Negative component, the data collected in the empirical study (Chapter 4) will be analysed in the fashion outlined in Chapter 2. Datasets including increasing numbers of trials will be created to form the respective individual averages for the analyses of the separate ERP-components. In contrast to Study 1 in which only the averages of the first three, five, seven, eight, nine, and 10 artefact-free trials have been analysed, Study 4 will analyse datasets including the very first usable trial, the average of the first two analysable trials up to the first 10 artefact-free trials. The results of the separate analyses for the different components will be compared to each other to form a profile of the ERP-components' behaviours depending on the number of included trials. Furthermore, as the influence of including varying numbers of artefact-free trials has only been investigated in the components' amplitudes as of yet, additional analyses along the same lines will be conducted on the latencies of the respective ERP-components.

As this introduction makes clear, the studies presented in the current thesis aim (a) to propose new ways for both acquiring and analysing infant EEG-/ERP-data (see Chapters 2 and 4) and (b) to further the knowledge of which pre-existing strategies may be more successful than others to help facilitate the progress of a study (see Chapter 3). Therefore, a novel strategy is designed and tested to make more efficient

use of the limited number of artefact-free trials that can typically be obtained from an infant participant. Additionally, based on a meta-analysis providing insights into the potential processes underlying typical infant EEG test-sessions, a novel approach to study-design is proposed and tested. The overall aim of this thesis is therefore to create studies which can help to bring about substantial changes in the traditions according to which developmental researchers approach data-acquisition and -analysis in general. Furthermore, the aim of some analyses is to help to standardise the methodologies that are in use at the moment. In so doing, this thesis strives to potentially improve the quality (and quantity) and generalisability of infant ERP-results that are currently published.

References

Altman, N.R., & Bernal, B. (2001). Brain activation in sedated children: Auditory and visual functional MR imaging. *Radiology*, *221*, 56-63.

Andrews-Espy, K., Molfese, D.L., Molfese, V.J., & Modglin, A. (2004). Development of auditory event-related potentials in young children and relations to word-level reading abilities at age 8 years. *Annals of Dyslexia*, *54*(1), 9-38.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, *8*(6), 492-499.

Braddick, O., Birtles, D., Wattam-Bell, J., & Atkinson, J. (2005). Motion- and orientation-specific cortical responses in infancy. *Vision Research*, *45*, 3169-3179.

Brannon, E.M., Libertus, M.E., Meck, W.H., & Woldorff, M.G. (2008). Electrophysiological measures of time processing in infant and adult brains: Weber's law holds. *Journal of Cognitive Neuroscience*, *20*(2), 193-203.

Bushnell, I.W.R., Sai, F., & Mullin, J.T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, *7*(1), 3-15.

Camaioni, L. (1997). The emergence of intentional communication in ontogeny, phylogeny, and pathology. *European Psychologist*, *2*(3), 216-225.

Camaioni, L., Peruccini, P., Bellagamba, F., & Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy*, *5*(3), 291-308.

Casey, B.J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development?. *TRENDS in Cognitive Sciences*, *9*(3), 104-110.

Csibra, G., Tucker, L.A., Volein, A., & Johnson, M.H. (2000). Cortical development and saccade planning: the ontogeny of the spike potential. *NeuroReport*, *11*(5), 1069-1073.

Cleveland, A., & Striano, T. (2007). The effects of joint attention on object processing in 4- and 9-month-old infants. *Infant Behavior & Development*, *30*, 499-504.

DeBoer, T., Scott, L.S., & Nelson, C.A. (2007). Methods for acquiring and analysing infant event-related potentials. In M. de Haan (ed.), *Infant EEG and Event-Related Potentials* (pp. 5-38), Hove, England: Psychology Press.

de Haan, M., Pascalis, O., & Johnson, M.H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199-209.

Duc, G., & Largo, R.H. (1986). Anterior Fontanel: Size and Closure in Term and Preterm Infants. *Pediatrics*, *78*(5), 904-908.

Floccia, C., Christophe, A., & Bertoncini, J. (1997). High-amplitude sucking and newborns: The quest for underlying mechanisms. *Journal of Experimental Child Psychology*, *64*, 175-198.

Friedrich, M., & Friederici, A.D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *NeuroReport*, *19*(18), 1757-1761.

Friedrich, M., & Friederici, A.D. (2010). Maturing brain mechanisms and developing behavioural language skills. *Brain and Language*, *114*(2), 66-71.

Geangu, E., Benga, O., Stahl, D., & Striano, T. (2010). Contagious crying beyond the first days of life. *Infant Behavior and Development*, *33*(3), 279-288.

Gredebäck, G., & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition*, *114*, 197-206.

Gredebäck, G., & Melinder, A. (2011). Teleological reasoning in 4-month-old infants: Pupil dilations and contextual constraints. *PLoS One*, *6*(10), 1-5.

Grossmann, T., Oberecker, R., Koch, S.P., & Friederici, A.D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, *65*, 852-858.

Halit, H., de Haan, M., & Johnson, M.H. (2003). Cortical specialisation for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*, *19*, 1180-1193.

He, C., Hotson, L., & Trainor, L.J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*, *19*(5), 878-892.

He, C., Hotson, L., & Trainor, L.J. (2009a). Maturation of cortical mismatch responses to occasional pitch change in early infancy: Effects of presentation rate and magnitude of change. *Neuropsychologia*, *47*, 218-229.

He, C., Hotson, L., & Trainor, L.J. (2009b). Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *European Journal of Neuroscience*, *29*, 861-867.

Hespos, S., Gredebäck, G., von Hofsten, C., & Spelke, E.S. (2009). Occlusion is hard: Comparing predictive reaching for visible and hidden objects in infants and adults. *Cognitive Science*, *33*, 1483-1502.

Hirai, M., & Hiraki, K. (2005). An event-related potentials study of biological motion perception in human infants. *Cognitive Brain Research*, *22*, 301-304.

Hirotani, M., Stets, M., Striano, T., & Friederici, A.D. (2009). Joint attention helps infants learn new words: Event-related potential evidence. *NeuroReport*, *20*(6), 600-605.

Hoehl, S., Palumbo, L., Heinisch, C., & Striano, T. (2008). Infants' attention is biased by emotional expressions and eye gaze direction. *NeuroReport*, *19*(5), 579-582.

Hoehl, S., Reid, V.M., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science*, *11*(1), 10-16.

Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, *37*(3), 187-209.

Hoehl, S., Wiese, L., & Striano, T. (2008). Young infants' neural processing of objects is affected by eye gaze direction and emotional expression. *PLoS*, *3*(6), 1-6.

Jasper, H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371-375.

Kaufman, J., Csibra, G., & Johnson, M.H. (2005). Oscillatory activity in the infant brain reflects object maintenance. *PNAS*, *102*(42), 15271-15274.

Kochukhova, O., & Gredebäck, G. (2010). Preverbal infants anticipate that food will be brought to the mouth: An eye tracking study of manual feeding and flying spoons. *Child Development*, *81*(6), 1729-1738.

Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustic Society of America*, *66*(6), 1668-1679.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, *7*(2), 173-187.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Pointing out new news, old news, and absent referents at 12 months of age. *Developmental Science*, *10*(2), F1-F7.

Lloyd-Fox, S., Blasi, A., & Elwell, C.E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience and Behavioral Reviews*, *34*, 269-284.

Luck, S.J. (2005). Ten simple rules for designing ER experiments. In T.C. Handy (ed.), *Event-related potentials: a methods handbook* (pp. 17-32). Cambridge, MA: MIT Press.

Matsuzawa, J., Matsui, M., Konishi, T., Noguchi, K., Gur, R.C., Bilker, W., & Miyawaki, T. (2001). Age-related volumetric changes of brain gray and white matter in healthy infants and children. *Cerebral Cortex*, *11*(4), 335-342.

Männel, C. (2008). The method of event-related brain potentials in the study of cognitive processes. In A.D. Friederici, & G. Thierry (eds.), *Early language development: bridging brain and behaviour; Series: Trends in Language Acquisition Research (TiLAR), Vol. 5* (pp. 1-22), Amsterdam/Philadelphia: John Benjamins.

Mills, D.L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: effects of vocabulary size and experience. *Cognitive Development*, *20*, 19-31.

Nelson, C.A., & Monk, C.S. (2001). The use of event-related potentials in the study of cognitive development. In C.A. Nelson, & M. Luciana (eds.), *Handbook of developmental cognitive neuroscience* (pp. 125-136), Cambridge, MA: MIT Press.

Otten, L.J., & Rugg, M.D. (2005). Interpreting event-related brain potentials. In T.C. Handy (ed.), *Event-related potentials: a methods handbook* (pp. 3-16). Cambridge, MA: MIT Press.

Pang, E.W., Edmonds, G.E., Desjardins, R., Khan, S.C., Trainor, L.J., & Taylor, M.J. (1998). Mismatch negativity to speech stimuli in 8-month-old infants and adults. *International Journal of Psychophysiology*, *29*, 227-236.

Parise, E., Reid, V.M., Stets, M., & Striano, T. (2008). Direct eye-contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience*, *3*(2), 141-150.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, *18*, 79-85.

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: International University Press.

Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Norton.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, Jr., R., … Taylor, M.J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standarads and publication criteria. *Psychophysiology*, *37*, 127-152.

Purhonen, M., Kilpeläinen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2005). Four-month-old infants process own mother's voice faster than unfamiliar voices – Electrical signs of sensitization in infant brain. *Cognitive Brain Research*, *24*, 627-633.

Reid, V., & Geangu, E. (2008). Research methodology and social cognition. In T. Striano, & V.M. Reid (eds.), *Social cognition: Development, Neuroscience and Autism* (pp. 17-34). Oxford: Blackwell.

Reid, V.M., Hoehl, S., Grigutsch, M., Groendahl, A., Parise, E., & Striano, T. (2009). The neural correlates of infant and adult goal prediction: Evidence for semantic processing systems. *Developmental Psychology*, *45*(3), 620-629.

Repovš, G. (2010). Dealing with noise in EEG recording and data analysis. *Informatica Medica Slovenica*, *15*(1), 18-25.

Richards, J.E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: An ERP study. *Developmental Science*, *6*(3), 312-328.

Rugg, M.D., & Coles, M.G.H. (1995). The ERP and cognitive psychology: conceptual issues. In M.D. Rugg, & M.G.H. Coles (eds.), *Electrophysiology of mind – Event-related brain potentials and cognition. Oxford psychology series, No. 25* (pp. 27-39). New York, NY: Oxford University Press.

Salvucci, D.D, & Goldberg, J.H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (pp. 71-78), New York, NY: ACM.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, *33*(1), 3-11.

Scott, L.S., & Nelson, C.A. (2006). Featural and configural face processing in adults and infants: A behavioral and electrophysiological investigation. *Perception*, *35*, 1107-1128.

Sekerina, I., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns?. *Journal of Child Language*, *31*, 123-152.

Slaughter, V., & Suddendorf, T. (2007). Participant loss due to "fussiness" in infant visual paradigms: A review of the last 20 years. *Infant Behavior and Development*, *30*, 505-514.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*, 71-86.

Striano, T., Henning, A., & Stahl, D. (2005). Sensitivity to social contingencies between 1 and 3 months of age. *Developmental Science*, *8*(6), 509-518.

Striano, T., & Tomasello, M. (2002). Infant development: Physical and social cognition. In N.J. Smelser & P.B. Baltes (eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 7410-7414). Pergamon: Oxford.

Symons, L.A., Hains, S.M.J., & Muir, D.W. (1998). Look at me: Five-month-old infants' sensitivity to very small deviations in eye-gaze during social interactions. *Infant Behavior and Development*, *21*(3), 531-536.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (eds.), *Joint attention: Its origins and role in development* (pp. 103-130). Hillsdale, NJ: Lawrence Erlbaum.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Werker, J.F., Polka, L., & Pegg, J.E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development and Parenting*, *6*, 171-178.

Chapter 2

Infant ERP Amplitudes Change Over the Course of an Experimental Session: Implications for Cognitive Processes and Methodology

Manuela Stets and Vincent M. Reid

Brain and Development (2011), *33*(7), 558-568.

Abstract

Event-related Potentials (ERP) studies are a widely used methodology to investigate the early cognitive development in infants of all age ranges. We examined changes in amplitude contribution in a set of previously published data (see [4]) in relation to the Negative component as a function of number of trials contributing to each individual average and with time as a co-variate to that contribution. With only seven trials contributing to each individual's average the Nc for the object-directed condition was significantly more negative than the one for the averted-gaze condition, which is the opposite of the effect reported in Hoehl et al. [4]. The analysis including time as a co-variate revealed that the effect did not alter according to the amount of time taken during the experiment to collect each trial, suggesting that data of the quality needed to contribute to the ERP is itself a measure of time varying components of the experimental session. We conclude that infants initially allocate more attentional resources to object-directed gaze. This suggests that the temporal dynamics of infant ERP violates assumptions present in adult ERP, and that differences between conditions in relatively few trials inform us of cognitive capacities during infancy.

Introduction

Event-related Potentials (ERPs) are used to investigate the cognitive and neural architecture of many phenomena in the field of psychology. As it is a non-invasive way of investigating cognitive processes related to the perception and processing of a certain stimulus or a kind of stimuli, it is well-suited for research with infants and young children (e.g., [1-2]). This is particularly the case with infant populations as behavioral responses that can be measured in older infants are often not clearly present at younger ages (see [3]). For example, an infant may have the capacity to react to certain stimuli and differently process multiple streams of stimuli, but no evidence for this would be manifested at the behavioral level despite differences in processing seen at the level of neural correlates (see [4-6]).

One important characteristic of ERP-studies conducted to date with infants is that the methodologies applied and the ways of interpreting the data are derived from cognitive neuroscience with adult populations. These two populations need to be treated differently with respect to how ERP data are recorded and analyzed (see also [7]). Should these elements be standardized, there would still remain differences between infants and adults concerning several physical characteristics such as the thickness of the skull or myelination, leading to different morphologies of the resultant ERPs (see [2, 7]). Also, the cognitive processes triggered by the same kind of stimulus might be different in these two populations (e.g., [3]). The common practice of conducting a study with adult participants initially in order to create predictions for infant participants unintentionally introduces complicating factors, such as the possible introduction of biases in the way that the infant data will be analyzed.

Scientists working with ERPs hold that the noise in the signal will decrease should more artifact-free trials be added to the analysis for each individual participant's ERP average. Picton et al. [8] produced guidelines that are commonly considered when conducting, analyzing, and publishing an ERP-study. However, these guidelines mostly apply to adult participant populations. Even though it is mentioned repeatedly that clinical populations and (young) children require different treatment, the report lacks detail concerning these issues. Parallel to Picton et al. [8], Luck [9] lists ten rules for how to run a successful ERP-study. These include that the number of trials present in the final analysis should be as high as possible to ensure that noise will be averaged out of the procured signal. For instance, he states that the perfect figure for a signal-to-noise ratio of 10.0 would require the collection of 625 trials per condition from each participant in a study investigating a large component such as the P300 (see [9]). Further, it is stated that the minimal number for a badly rated signal-to-noise-ratio of 0.4 requires between 30 and 60 usable trials per condition and participant. As these numbers are suggested for studies with adult participants, Luck [9] further states that they ideally would need to be doubled or tripled for infant and other populations to still achieve the aim of attaining clean data with as little noise in the signal as possible. This, however, poses an obvious problem to all ERP-studies investigating infant populations as, normally, far fewer trials are presented to infants due to limits in infant attention span. This issue is compounded as the number of trials acceptable for analysis is further reduced via behavioral artifacts, such as movement or eye blinks, that are present in recorded data. Consequently, experimental sessions with infants are designed to be brief in order to fit with the infants' attention span while still obtaining enough artifact-free trials to be able to conduct an appropriate statistical analysis. It is therefore recommended that two conditions are optimal for ERP studies

with infants, with more conditions increasing attrition rates due to the raised probability of too few trials included in the average per condition for each participant (see [10]).

In isolation to the issue of signal-to-noise, some work has previously examined how ERP components may change over the course of an experimental session. Using the example of the Negative component in the ERPs of 6-month-old infants elicited by the frequent stimulus in an odd-ball paradigm, Nikkel and Karrer [11] suggested that the amount of experience a participant has with a specific stimulus has the capacity to influence the strength of a reported ERP component. The authors split their test session into three blocks and compared the amplitudes of the ERP components measured in the individual test blocks with each other. As had been predicted, they found a trend for the amplitudes of the Nc to decrease from one test block to the next. The authors suggested that the decrease in the amplitude of the Nc was connected to the amount of effort the infants made to attend to the stimuli. This suggests that the more often the infants saw the stimuli, the less attention they would need to pay to those stimuli in order to establish a memory trace for the stimulus (see [11]).

In another study by Snyder, Webb, and Nelson [12], the effect of familiarity with a stimulus on the Negative component was investigated in 6 month olds by comparing the amplitudes and latencies of infants who produced varying numbers of acceptable trials for analysis. They found that this factor was related to the amplitude and latency of the Nc, with differences related to the degree of encoded information derived from the stimuli. They then related this to differences between studies in terms of the number of trials infant's observed when contrasted with the number included in the final analyses, suggesting that large variability between infants will distort or diminish effects.

There is also the added complexity of the participants' age when considering infant data. In a study on the P300-component evoked by auditory stimuli, McIsaac and Polich [13] investigated an age range of between 6 and 10 months and reported the results as one averaged ERP. The authors reasoned that the age range was not problematic as changes in ERP-components evoked by auditory stimuli are relatively rare during infancy. In contrast, Kushnerenko, Ceponiene, Balan, Fellman, Huotilaine, and Näätänen [14] found that there is high variability in the ERP components between infant subjects of the same age group. As a consequence of such research, it has been suggested that longitudinal studies should be conducted to determine how many trials are optimal for a final analysis for each age group (see also [3, 15]).

In the present study, we followed a different approach to examine signal-to-noise issues in infant data. We are interested in determining the appropriate number of artifact-free trials per condition needed for infant populations, and, additionally, we are including time as a co-variate. This means that we are including the time it took an individual participant to collect a specific number of usable trials from the start of the stimulus presentation as an additional factor in our examination of the ERP data. Instead of testing the same subjects at different ages, we re-analyzed an existing set of previously published EEG-data (see [4]). When investigating infant social-cognitive processing capacities, these authors showed that infants as young as 4 months of age differentially process stimuli depicting either an adult looking toward an object positioned at eye-level to the side of her head (the "object-directed" condition) or the same female adult looking away from an object positioned at eye-level to the side of her head (the "averted" condition). The ERPs that Hoehl et al. [4] extracted from the data of 17 participants who fulfilled their inclusion criterion of contributing at least 10 artifact-free trials per condition to the average, showed a more pronounced Negative

component (Nc) in frontal areas in the time window from 400 to 600 ms after stimulus onset for the condition in which the adult was not looking at the object. Along with other studies reporting on the amplitude of the Nc, Hoehl et al. [4] interpreted the enhanced Nc for the averted condition such that the infants allocated more attention to these stimuli compared to those depicting the object-directed gaze.

As described previously, the Nc usually appears in frontal regions (see [16]) and might also spread to central areas (see [17]). It is usually associated with attention allocation, with a stimulus eliciting a more negative amplitude Nc compared to another condition with more attentional resources dedicated to its processing than a condition featuring a diminished negative amplitude (e.g., [4]). Reynolds and Richards [16] modeled the source of the Nc as the anterior cingulate, which is in line with adult research on attention systems in the human brain.

As evidenced in many published studies, often fewer than 20 trials per condition are included in the average. As Nickel and Karrer [11] reported reductions in Nc amplitude between blocks, we hypothesized that such reductions would also show in the variation in Nc amplitude over the course of a typical test session. The implication of this hypothesis is that when trials from the end of an experimental session are included in an average they may yield fewer differences in amplitude between conditions when compared to the trials from the early stages of the same infant during the same test session. Therefore, we hypothesize that focusing the statistical analysis on fewer trials from the early stages of the test session may make it possible to still achieve valuable results, which might have implications for subsequent work with infant populations. Overall, our study had the aims of monitoring the behavior of the Nc according to the number of trials included per condition and the effect of time on the amplitude of the Nc.

Analysis 1

Nc and the number of trials contributed to the average ERP per condition

Methods

Participants

For the study outlined in Hoehl, Reid, Mooney, and Striano [4], a total of 64 infants were tested but the final analysis comprised of the data of 17 of these. The data of the other 47 participants had been excluded for reasons of fussiness (n = 19), failing to reach the minimum requirements for adequate averaging of the ERP data, including artifacts (n = 14), and technical problems during the saving of collected data (n = 14). The infants were born full-term (after 37 to 41 weeks of gestation) and had an average age of 4 months +/- 9 days at the time of testing. For this analysis we examined the same data of the 17 infants, who were included in the final analysis of the original article.

Stimuli

The stimuli used in the article by Hoehl et al. [4] depicted the face of a female adult who was either looking toward a toy positioned at eye-level next to the head ("object-directed" condition), or the same face looking away from a toy positioned at eye-level ("averted" condition; see [4]).

Procedure and data-collection

Infants sat on their caregivers' laps and observed a randomized succession of the stimuli described above. The stimuli were presented with the constraints that the

same condition was not presented three times consecutively and that the number of presentations of each set of stimuli was balanced in every 20 trials presented. The program consisted of five blocks of 20 repeating stimuli. Before the stimuli were presented for 1000 ms, a fixation object was shown for 500 ms. After the presentation of the stimulus the screen was left black for a period of 800 to 1000 ms. Thus, if an infant would watch the entire program from beginning to end without any interruptions this would take from 230 s (3.833 min) to 250 s (4.167 min).

The EEGs were recorded according to the international 10 – 20 system using Ag-AgCl-electrodes with a sampling rate of 250 Hz. The data were referenced on-line to Cz and were re-referenced off-line to the linked mastoids (channels TP9 and TP10).

Data analysis

When investigating the Nc, Hoehl et al. [4] examined the time window between 400 and 600 ms after stimulus onset from those participants who met their inclusion criterion of having at least 10 artifact-free trials per condition. The authors conducted a repeated measures ANOVA for this time window to ascertain if there would be a significant difference between the negative peaks for the two conditions. The area of interest covered the frontal channels (F3, F4, Fz, FC3, FC4, C3, C4, and Cz).

Following Hoehl et al. [4], we compared the negative peaks in the time window from 400 to 600 ms after stimulus onset in the same area of interest for the two conditions. However, we did not include all the trials an individual infant contributed to their individual average as was reported in the paper. Rather, we compared the negative peaks for varying amounts of artifact-free trials. We reasoned that as attention modulated throughout the course of an experimental session, we would observe

differences between conditions as a function of time. Therefore, we examined the first three artifact-free trials where the infant was attending to the screen per condition for the 17 subjects reported in Hoehl et al. [4] and compared the negative peaks of the two conditions by means of a repeated measures 2 x 3-ANOVA with condition (object-directed versus averted gaze) and channel location (left frontal, F3 + FC3 + C3, versus fronto-central, Fz + Cz, versus right frontal, F4 + FC4 + C4) as the independent variables. Then, we followed the same procedure for the first five, seven, eight, nine, and 10 trials. This way we were able to examine the evolution of the morphology and amplitude of the Negative component as a function of the number of analyzable trials increased within the individual average of each participant.

Due to the manner in which the datasets have been created, they are highly dependent on each other. In order to test the possibility that differences, which may be found between conditions, would occur by chance in these multiple comparisons (i.e., the occurrence of a Type I error; [18]), a one-way ANOVA with a Bonferroni-correction was performed. For the condition with the directed gaze, only the dataset including the first three artifact-free trials was significantly different from the other datasets for this condition and for the three channel-groups; $p \leq .029$. This is likely due to the larger variation or noise which remains in the data when only three trials are forming an average as compared to averages derived from larger numbers of artifact-free trials. The size of the error bars in Figure 2 below are indicative of this variation. No significant differences were found between datasets originating from the averted gaze condition. This indicates that datasets were generally very similar to each other – apart from the first dataset for the directed gaze condition. Therefore, the likelihood that effects may be found due to outliers in certain dataset, should be relatively low.

Results

Hoehl, Reid, Mooney, and Striano [4] reported a significantly more negative peak of the Negative component for the averted gaze condition compared to the object-directed condition for 17 participants ($F(1, 16) = 4.8$, $p = .044$; see Figure 1a for a sample electrode which displays the overall effect of Nc). The peak amplitude for the object-directed gaze condition was reported to be $M = -14.3$ μV, $SE = 2.26$, whereas the peak amplitude for the averted gaze condition was $M = -19.6$ μV, $SE = 2.0$. This analysis comprised of all the artifact-free trials an individual participant could contribute to the average. The range of artifact-free trials contributed per condition and participant varied across their 17 participants between 12 and 37 for the object-directed condition and between 11 and 37 for the averted-gaze condition.

Reducing the number of included trials per condition to three, five, seven, eight, nine, or 10 reverses this picture. The repeated measures 2 x 3-ANOVA conducted on the first three usable trials per condition for the originally included 17 participants revealed an effect of condition which was approaching significance, $F(1, 16) = 3.595$, $p = .076$, with the stimuli depicting the object-directed gaze eliciting the more negative peak (object-directed peak amplitude, $M = -56.96$ μV, $SE = 10.94$; averted peak amplitude, $M = -29.99$ μV, $SE = 7.37$). The repeated measures ANOVAs for the first five artifact-free trials revealed that there was no difference between the peaks for the two conditions, $F(1, 16) = .001$, $p = .978$ (object-directed, $M = -24.25$ μV, $SE = 6.56$; averted, $M = -24.01$ μV, $SE = 6.07$). When comparing the negative peaks for the first seven analyzable trials per condition, we found a significant effect of condition with the object-directed condition eliciting a more negative peak, $F(1, 16) =$

6.103, p = .025 (object-directed, M = -18.15 μV, SE = 3.00; averted, M = -10.32 μV, SE = 2.57; see Figure 1b for an illustration of the grand average of the area of interest).

From eight trials onwards, the polarity of the Nc started to alter towards the direction reported by Hoehl et al. [4]. Thus, the repeated measures 2 x 3-ANOVA for the dataset including the first eight artifact-free trials revealed that there was no significant difference between conditions (F(1, 16) = 2.841, p = .111 with object-directed, M = -16.79 μV, SE = 2.79, and averted, M = -11.75 μV, SE = 2.83). When the first nine trials per condition were examined, there was no difference between conditions (F(1, 16) = .077, p = .785 with object-directed, M = -18.63 μV, SE = 3.21, and averted, M = -17.68 μV, SE = 3.28). Finally, when we included the first 10 trials into the analysis, there was no significant effect of condition, however, the conditions had now altered positions in terms of eliciting the most negative mean response (F(1, 16) = .062, p = .806 with object-directed, M = -17.73 μV, SE = 2.66, and averted, M = -18.39 μV, SE = 3.6). This indicates that the final outcome presented in Hoehl et al. [4] was due to amplitude contributions made in the latter half of the test session. An illustration of the development of the Negative component according to how many trials were included from the beginning of the experimental session is given in Figure 2.

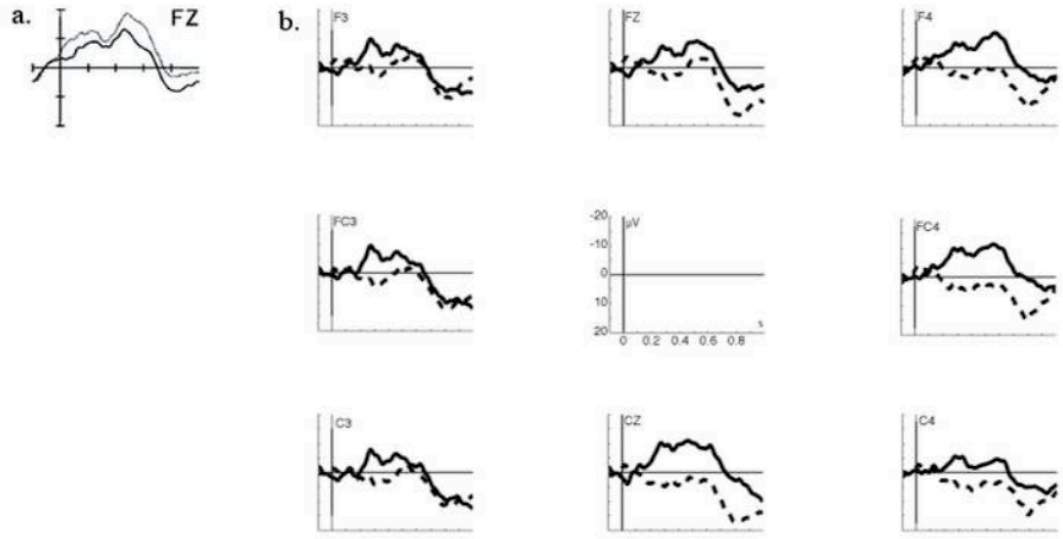

Figure 1. Part (a) depicts Fz as a sample electrode for the grand average presented in Hoehl et al. [4]. Part (b) illustrates the grand average of the area of interest (channels F3, F4, Fz, FC3, FC4, C3, C4, and Cz) for 17 subjects with the first seven artifact-free trials per condition included in the statistical analysis (solid line for the object-directed condition, dotted line for the averted gaze condition).

Discussion

Our results show that the Nc changes over the time-course of the experimental session. ERPs extracted from the first few artifact-free trials contribute differences in amplitudes between conditions, which are the opposite of those contributed in the latter half of the experimental session. As can be seen in Figure 2, there are many changes in the development of the difference between the object-directed and the averted gaze conditions. However, already with the first three artifact-free trials the difference between the conditions indicates a strong trend towards significantly diffe-

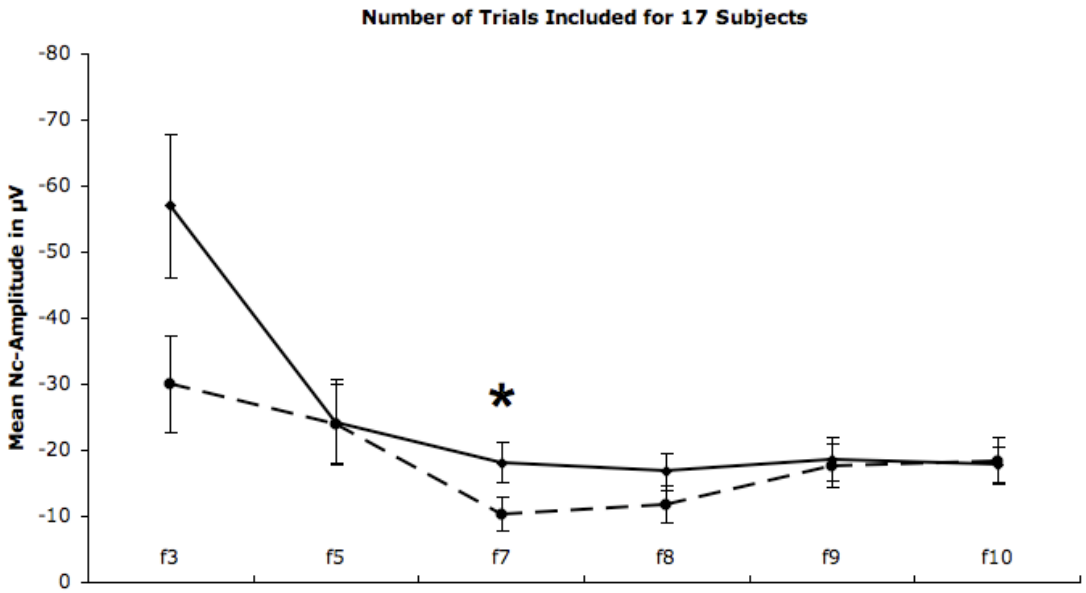

Figure 2. The development of the Negative component in the frontal area (channels F3, F4, Fz, FC3, FC4, C3, C4, and Cz) for 17 subjects according to the number of artifact-free trials included per condition (the solid line depicts the object-directed condition, the dashed line depicts the averted condition).

rent conditions. However, as is indicated by the differences between our dataset with three artifact-free trials and the remaining datasets for directed gaze condition as well as by the lack of a difference between the condition when five trials formed the average, there is still much amplitude variation delivered for each trial. This is mainly apparent from the comparatively large error bars depicted in Figure 2 above – especially in the first two datasets where the likelihood for noise remaining in the signal is larger with such few trials in an average. The apparent collapse in differential processing of the conditions may be due to the slightly different trajectories which the conditions seem to follow with respect to the negativity of the responses that they elicit. The object directed condition shows a steep decline in response negativity from the dataset including the first three artifact-free trials to the one including the first 5

usable trials. From seven trials onwards, the responses remain roughly at the same level of negativity. A similarly steep decline in negativity for the away-condition can only be found when comparing the datasets including the first five trials to the one including the first seven artifact-free trials in the average. Then, the amplitudes are slowly gaining in negativity the more trials are added to the average. However, overall the direction of the main effect of condition is the opposite to that reported in Hoehl et al. [4], with the first seven artifact-free trials revealing a highly significant difference between conditions.

The implications for our understanding of the Nc are substantial. A cognitive explanation for these effects is that the infants initially allocate more attention to the socially more normal situation in which the adult is looking toward an object. This fits with current thought on dyadic information processing, as the infant is primed for the adult to communicate some meaningful information to them with ostentative gaze at the toy (e.g., see [19-20]). Then, the more artifact-free trials are added to the analysis per condition (the first eight, the first nine, or the first 10, respectively) the direction of the condition effect starts to change slowly until the averted gaze condition finally elicits the significantly more negative responses as reported in Hoehl et al. [4]. This effect implies that the infants start to allocate increased attention to the more uncommon situation – a situation in which an adult is communicating ambiguous information by looking away from the object. One interpretation of the changes seen in the Nc amplitudes is that this as an indication that the infants anticipate that the adult will convey meaningful and clearly defined information via gaze. When the infants are familiar with this stream of stimuli, they switch their attention to the more uncommon situation. A further interpretation of these results could also be that the infants start to pay more attention to the averted gaze stimuli because they expect the adult to

communicate ostensive information when looking towards the empty space – meaning that the infants expect the adult to show them something off-screen, with the ambiguity of the stimuli contributing to its novelty.

Investigating infant populations with ERPs has advantages and disadvantages. The big advantages of this technique are that it has a high temporal resolution and that it is non-invasive in nature. One of the disadvantages is, however, that the paradigms are limited in time due to limited attention spans during infancy. Another disadvantage is the relatively high risk of losing data due to potential movement artifacts and a lack of attention to the stimuli. For example, in their classic study from 1981, Courchesne, Ganz, and Norcia [21] excluded 61.5% of their participants aged between 4 and 7 months, and de Haan, Pascalis, and Johnson [22] excluded 80.6% of the 6-month-olds they tested due to too few trials for inclusion in the final sample. Our results suggest that it would actually not be necessary to include that many trials in the analysis to achieve a statistically significant difference between the conditions. In order to make the experimental sessions shorter and, therefore, more suitable for the infant population, we found that concentrating the statistical analysis on a few artifact-free trials from the start of the experimental session could provide meaningful results which were significantly different between conditions. This is in direct contrast to the statements made by Luck [9] and Picton et al. [8]. However, ERP paradigms with infants where all collected useable trials are included already violate these principles as too few trials are included under this technique to fit with adult cognitive neuroscience norms.

As has been described above, we found a strong tendency for the effect of condition to approach significance when the first three artifact-free trials of the original 17 participants were included in the average. This finding led us to believe that

including more participants in the grand average would increase the power of this finding, and thereby support the notion that the trend found in Analysis One was indeed due to cognitive processes. Alternatively, the addition of new participants would reduce the difference between conditions and indicate that the trend for significance in the first three trials was simply due to random noise. We therefore decided to add the data from participants who had been excluded from the final analysis in Hoehl et al. [4] due to too few trials provided by these participants for their individual averages.

Analysis 2

Analysis of the first 3 artifact-free trials with additional participants

Methods

Participants

In addition to the data of the 17 participants originally reported in Hoehl et al. [4], we included the data from seven other participants tested for the original study but excluded from the final analysis. These infants had not met the authors' inclusion criterion of contributing at least 10 artifact-free trials per condition to the average, of which there were 14 excluded in the original paper[1]. However, as we were concentrating on the first three usable trials in this analysis, we randomly selected seven additional participants from this pool of infants. We conjectured that the inclusion of only a small additional number of infants would alter the trend towards significance in Analysis One such that this effect would become highly significant. This would effectively demonstrate the validity of investigating a small number of trials from the beginning of the experimental session when contrasted with conventional ERP averaging techniques derived from all possible trials. It has widely been reported that increased power for statistical tests in EEG studies can be derived from larger sample sizes (see [8]). In fact, during the submission process for [4], reviewers asked for the sample size to be increased to make the effects more clearly

[1] Efforts were made to include the data from the additional seven infants into the analyses outlined in Analyses One and Three. Unfortunately, the directory contaning the EEG-data files for this study were corrupted during a hard-drive crash and could not be recovered. Only the mean values used in this analysis were still available.

delineated [23], and, therefore, more infants had been tested in order to make the Nc effect clearer.

Stimuli, procedure, and data-collection

These were the same as reported in Analysis One.

Data analysis

This was the same as reported in Analysis One. However, for the analysis for Analysis Two, we performed the statistical analysis for the first three usable trials per condition for 24 participants (the original 17 reported in Hoehl et al. [4] and the additional seven infants extracted from the pool of previously excluded participants).

Results

The difference between the Nc for the object-directed condition and the averted gaze condition was significantly different. The object-directed condition elicited a more negative response, $F(1, 23) = 4.940$, $p = .036$ (peak amplitude for the object-directed condition, $M = -55.42$ μV, $SE = 8.9$, peak amplitude for the averted gaze condition, $M = -27.52$ μV, $SE = 7.29$). See Figure 3 for a grand average of the area of interest.

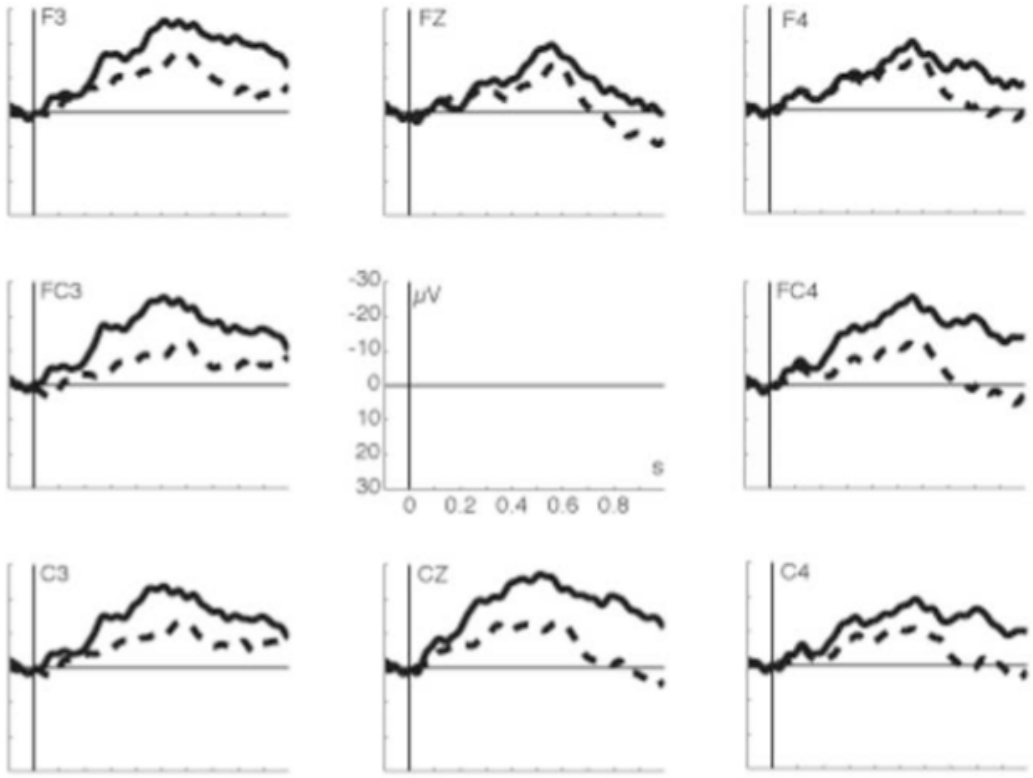

Figure 3. Grand average of the area of interest (channels F3, F4, Fz, FC3, FC4, C3, C4, and Cz) for 24 subjects (the original 17 plus the added seven) with three artifact-free trials included per condition and subject (the solid line represents the object-directed condition, the dotted line represents the averted gaze condition).

Discussion

By including the data of seven additional infants who had been excluded from the final analysis in the original study due to contributing too few trials per condition, we were able to increase the power of the effect with a trend towards significance (see Analysis One) such that it became a statistically significant difference between conditions. This suggests that an analysis of the first few trials may yield results of interest and significance for infancy researchers in addition to the technique of

including all useable trials from each participant for each experiment. Both techniques of analysis violate the principles outlined by Luck [9] and Picton et al. [8].

Recently, researchers investigating infant populations using eye-tracking methodologies have started to analyze their data in an analogous fashion – by investigating the "first look" as well as the overall pattern of looking behavior. Gredebäck and Melinder [22], for instance, found a marked difference in the way that 6- and 12-month-old participants observed feeding scenes over the course of the experimental session. The authors analyzed the data for how well the participants would follow the hand of the feeding person or if they would predict the beneficient of the feeding action by looking at that person before the spoon arrived there. They also calculated the latency of the first fixation on the expected beneficient (time point when spoon entered the head- or hand-area-of-interest surrounding the expected beneficient minus time of infant's first fixation at one of these areas of interest). Gredebäck and Melinder [24] found a marked decrease in latency (especially for the 12-month-olds) for the first trial of the expected condition compared to the second and third trials (see [25] for a review). This way of analyzing eye-tracking data is analogous to our suggestion of examining fewer trials from the beginning of the test session as both cases suggest that a participant's responses to the same kind of stimulus change over the course of the experiment and that they are not stable. The question arises for ERP results if it may be the time that elapses from the start of the stimulus presentation that contributes to this change in amplitude or if it may be the number of trials within the stimulus presentation that contribute to the change in amplitude.

Analysis 3

Does the length of a test session alter the amplitude of the Nc?

Methods

Participants, stimuli, procedure, and data-collection

These were the same as described in Analysis One.

Data analysis

We examined the time it took the 17 4-month-olds reported in Hoehl et al. [4] to collect a specific amount of artifact-free trials per condition. Therefore, we took note of when a subject had seen each specific trial that was classified as artifact free during the course of the experimental session. We took note of the times needed to obtain all the datasets analyzed in Analysis One, which is the first three, five, seven, eight, nine, and 10 artifact-free trials per condition. In order to be able to analyze the condition effects found in Analysis One with time as a co-variate, the infants were then categorized according to the amount of time that had elapsed from the start of the stimulus presentation up until the last trial included in the analysis for that infant. During observation of the time taken to collect these data, we determined that thirty-second time windows would accurately capture differences in the time taken to collect different amounts of data from different infants. To introduce time as a co-variate into the analysis, we weighted the data accordingly for each infant. For instance, one of the faster infants would collect seven artifact-free trials for both conditions within 60 s after the beginning of the stimulus presentation, whereas another participant for whom several trials may have had to be excluded for various reasons would have taken up to

120 s or even more than 150 s to collect the same number of analyzable trials for both experimental conditions. We therefore, classified infants as a function of the speed at which we could obtain data from them during the experimental session.

The fastest participants entered the analyses as Category One-infants. These infants produced useable trials in the shortest possible time (see Table 1). Category Two-participants were those who produced usable trials at a rapid rate but not within the same timeframe as the Category One-infants. They were weighted with a factor of 0.8 (80%), whereas the Category Three-infants took longer and were weighted with a further reduced factor of 0.6 (60%). Category Four-participants took even longer and were weighted with a factor of 0.4 (40%). Category Five-infants were those infants who took the longest time to reach a specified number of trials, and were weighted as 0.2 (20%). The same number of artifact-free trials can occasionally be collected within the same time-windows. For example, for the fastest participants, the first three trials and the first five trials can both be collected within the first 30 seconds of the experiment. This is due to the high variability in the amount of time that it took the infants to collect a specific number of trials per condition, and this variability is due to a number of different reasons such as inattentiveness to the screen or fussiness. We conducted paired-samples t-tests on the mean amplitudes for the two experimental conditions with these weighted factors. For an illustration of the weighting procedure for all the datasets reported in Analysis One, see Table 1. Additionally, a post-hoc one-way ANOVA including a Bonferroni-correction was performed to establish if there were significant differences between the weighted datasets presented in this analysis. Finally, in order to establish if the effect reported in Analysis Two above would hold after being subjected to the same weighting process, we performed a paired-samples t-test on the weighted data of the 24 participants reported above.

Table 1. Rating of the subjects according to the time it took them to collect a specific number of artifact-free trials for the two conditions.

	Category 1 – Subjects weighted with 1.0 (100%)	***Category 2 – Subjects weighted with 0.8 (80%)***	***Category 3 – Subjects weighted with 0.6 (60%)***	***Category 4 – Subjects weighted with 0.4 (40%)***	***Category 5 – Subjects weighted with 0.2 (20%)***
First 3 trials	30 s (*n* = 7)	60 s (*n* = 5)	90 s (*n* = 2)	120 s (*n* = 2)	more than 120 s (*n* = 1)
First 5 trials	30 s (*n* = 2)	60 s (*n* = 6)	90 s (*n* = 4)	120 s (*n* = 1)	more than 120 s (*n* = 4)
First 7 trials	60 s (*n* = 4)	90 s (*n* = 6)	120 s (*n* = 1)	150 s (*n* = 2)	more than 150 s (*n* = 4)
First 8 trials	90 s (*n* = 7)	120 s (*n* = 4)	150 s (*n* = 1)	180 s (*n* = 2)	more than 180 s (*n* = 3)
First 9 trials	90 s (*n* = 7)	120 s (*n* = 4)	150 s (*n* = 1)	180 s (*n* = 1)	more than 180 (*n* = 4)
First 10 trials	90 s (*n* = 4)	120 s (*n* = 5)	150 s (*n* = 2)	180 s (*n* = 2)	more than 180 s (*n* = 4)

Results

The analysis for Analysis Three investigated the amount of time that had elapsed for the individual 17 subjects from the start of the stimulus presentation to the point when they had collected a specific amount of artifact-free trials per condition. In order to investigate the possible influence that time might have on the effects found in Analysis One above and to use time as a co-variate in our analysis, we set up the above rating scale (see Table 1) such that the subjects could be weighted according to how little or how much time they took to collect the artifact-free trials (three, five, seven, eight, nine, or 10 per condition, respectively). The paired-samples t-tests of the mean amplitudes for the 17 participants revealed that the development of the Nc according to the amount of trials included in the average did not change from Analysis One when time was introduced as a co-variate. Figure 4 illustrates the way the Nc developed according to the number of trials in the analysis. However, due to the weighting of the data, all of the datasets tended to lose statistical power when contrasted with the analysis in Analysis One, where all participants had been weighted equally. Therefore, the difference for the Nc in the two conditions no longer approached significance when the first three trials were included, $t(1, 12) = -1.450$, $p = .172$ (object-directed, $M = -51.78$ μV, $SE = 11.28$; averted gaze, $M = -29.20$ μV, $SE = 8.87$). For the first five artifact-free trials the paired-samples t-test revealed that there was no difference, $t(1, 9) = .037$, $p = .972$ (object-directed, $M = -26.04$ μV, $SE = 8.61$; averted, $M = -26.48$ μV, $SE = 8.01$). When seven trials had been included in the analysis, the paired-samples t-test was significant, $t(1, 10) = -2.293$, $p = .045$ (object-directed, $M = -18.81$ μV, $SE = 4.21$; averted, $M = -8.98$ μV, $SE = 3.3$). For the first eight artifact-free trials: $t(1, 11) = -1.726$, $p = .112$ (object-directed, $M = -16.98$ μV, $SE = 3.65$; averted, $M = -10.59$ μV,

SE = 3.39), for the first nine trials: $t(1, 11) = -.378$, $p = .713$ (object-directed, M = -20.58 μV, *SE* = 4.25; averted, M = -19.04 μV, *SE* = 3.87), and when the first 10 artifact-free trials had been included: $t(1, 10) = .431$, $p = .676$ (object-directed, M = -19.24 μV, *SE* = 3.86; averted, M = -20.42 μV, *SE* = 4.84). The post-hoc one-way ANOVA with the Bonferroni-correction revealed that the weighted average derived from the first three usable trials was significantly different from the other datasets for the directed gaze condition. For the averted gaze condition, the dataset with the first three trials was significantly different from the average including the first seven usable trials only. No further differences were found between the datasets. It needs to be noted that variation in the data is likely to become more pronounced in such an analysis based on the loss of statistical power due to the weighting.

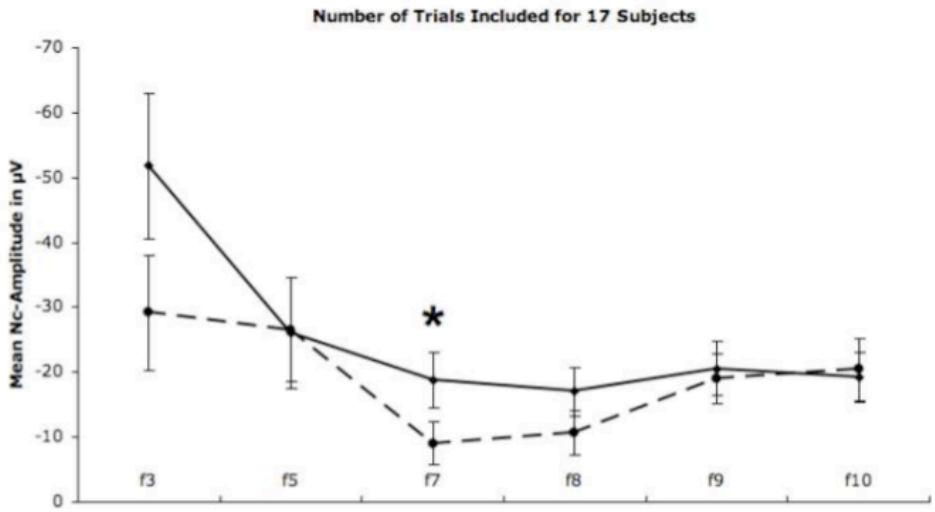

Figure 4. The development of the Negative component in the frontal area (channels F3, F4, Fz, FC3. FC4, C3, C4, and Cz) for 17 subjects according to the number of artifact-free trials included per condition in Analysis Three with time as a co-variate (the solid line depicts the object-directed condition, the dashed line depicts the averted condition).

Further, we conducted paired-samples t-tests with the mean times that the 17 subjects took to accumulate artifact-free trials for the two conditions in order to investigate if one condition would be significantly faster than the other over the time-course of the experiment. There was no significant difference between the conditions for any pair of equal sized trials. Finally, the effect of Condition reported in Analysis Two above could be replicated in a paired-samples t-test for the weighted data of the 24 infants described above; $t(1, 23) = -2.110$, $p = .046$ (object-directed, $M = -37.00$ μV, $SE = 6.52$; averted, $M = -18.98$ μV, $SE = 4.97$).

Discussion

In Analysis Three we concentrated on the time factor as a co-variate contributing to the variability in the data. Therefore, we categorized the infants according to the amount of time that it had taken them to collect a specific amount of artifact-free trials for the two conditions. The differences between the individual subjects were variable largely due to circumstances such as breaks because of an infant's inattentiveness to the screen or because of movement artifacts rendering trials invalid for analysis even when the infant was attending to the stimuli. There was no statistical difference between the mean times the infants needed to collect a specific number of artifact-free trials for the individual conditions. Also, comparing the mean amplitudes for the subjects with specific numbers of trials included per condition and weighting their data accordingly did not change the way the overall difference between the two conditions developed over the course of the experimental session. There was still a significant difference with the first seven usable trials being included in the analysis, as was also found in Analysis One, and the direction of the effect was still the

same with the object-directed condition being significantly more negative than the averted gaze condition in the first half of the experimental session. We interpret this finding such that the amount of artifact-free trials included in the averaging process is sufficient to incorporate the factor of time into the analysis. According to this variable the effect that was found changes across the duration of an experimental session.

General Discussion and Conclusions

Data editing and analysis methodologies for infant ERP studies derive from research techniques developed for work with adult populations. One fundamental element of this method is the assumption that the ERP signal remains constant throughout the time-course of a standard visual ERP study. Therefore, the more trials that are included in the average, the better the signal for the resultant ERP with random noise diminished. In contrast to this concept, we found substantial changes in infant responses to visual stimuli over the course of an experimental session. Specifically, the amplitude of the ERP-component varied from the early stages of the test session to its latter stages. As a function of the number of trials included in the average per condition, the amplitudes did not necessarily reduce in amplitude to a more stable value. Rather, they changed in terms of which condition elicits the more negative amplitude response. Moreover, our findings show that the time taken to collect a specific number of artifact-free trials per condition does not influence this development of the data. Instead, it is the number of trials included per condition and participant that has the major impact on the amplitude of the Nc in this study. There is no linear development in which the conditions elicit the same kind of ERP response at all times. Rather, the reported Nc undergoes a previously unforeseen evolution that

suggests that ERPs have the capacity to detect and track the way infants process social stimuli. However, in order to validate these results and to rule out the possibility that they are a function of the particular set of data that we used, further studies on this particular ERP component, using the proposed way of data analysis, are needed.

In the paper from 2008, Hoehl et al. [4] come to the conclusion that the infants' differential processing of the objects depicted depends on the direction of the adult's eye gaze. The fact that the Nc has a higher amplitude in the averted gaze-condition is explained such that more attentional resources are necessary for this information in order for it to be processed by the 4-month-olds. The condition in which the adult is looking toward the object does not require these additional resources for attention however, because this situation does not violate the infants' social expectations ([4]; see also [16, 26]). These statements are based on the way of analysis generally used in infant ERP-studies. As we now saw, this straightforward interpretation is not applicable under all circumstances. Initially, attentional resources are directed towards the condition, which does not violate the infants' social expectation, with attention switching later in the time-course of the experiment to the non-object directed gaze condition. A possible explanation for this is that the infant focuses first on the stimuli where social communication is present as the gaze towards the object indicates that the object is relevant to the adult, which may in turn make the object relevant to the infant. Only when this information has been processed does the infant switch their attention to the unusual stimuli that depict gaze away from an object, where no social information is being transmitted.

Our findings suggest that focusing the statistical analysis on fewer trials from the beginning of the experimental session in addition to examining all the artifact-free trials available per condition and subject could provide new insights into infant

cognitive development. Instead of surmising that ERP-components are fixed effects that do not change during an experimental session, comparisons with effects found with fewer trials can potentially illuminate cognitive processing of visual stimuli. One consequence of this finding is that researchers working in the field may wish to examine early effects with few trials and compare this analysis with overall effects obtained with all useable trials.

Another implication of this alternative method of analyzing infant ERP data is that it has the potential to make it possible for researchers working with infant populations to still achieve valuable results while reducing the length of the test session. This would also make such studies easier for participants to complete. Further, the number of participants could be reduced who would otherwise need to be excluded from the averaging process due to too few trails obtained of the quality required for inclusion in the average.

One important implication of the finding that differences in conditions are evident with relatively few trials included in the ERP average is that data from infants who are excluded for reasons of fussiness in many other ERP studies may be able to contribute data that can be analyzed and interpreted. Our results from Analysis Two clearly indicate that participants excluded for such factors do contribute meaningful and interpretable data, which is not currently utilized with standard measures of analysis. Implications include the possibility of widening parameters of included data, such that individual differences, such as temperament, can be examined in greater detail than is possible with standard ERP analysis techniques. In order to generalize our results, further studies with different datasets originating from different age groups using the same kind of analysis that we used will be necessary. Moreover, apart from applying the approach to the analysis of the Nc in response to different sets of stimuli

and originating from different age groups, its application to other (infant) ERP components will be necessary as well. We are aware of the preliminarity of this first attempt and would like to invite other researchers in the field to try the strategy in order to potentially strengthen and refine it.

Our results show that infant ERPs are dynamic aspects of brain function that clearly change over time in amplitude within an experimental session. Through showing that the Nc amplitude varies during the experimental session, with implications for our understanding of how infants process eye gaze and object processing, we conclude that the analysis of the first few trials may potentially reveal new insights into issues of cognitive processing during infancy. It could eventually lead to a re-evaluation of the practices currently predominant in ERP-studies with this population.

Acknowledgements

We thank Stefanie Hoehl for providing us with the data and the details about data editing. We thank Dr. Mike Burt and Dr. Daniel Stahl for their comments on a previous draft of this manuscript. Manuela Stets was funded by a ONE Northeast Studentship.

References

[1] Männel C. The method of event-related brain potentials in the study of cognitive processes. In: Friederici AD, Thierry G, editors. Early Language Development: Bridging brain and behaviour (Series: Trends in Language Acquisition Research (TiLAR), Vol. 5. Amsterdam/Philadelphia: John Benjamins; 2008. p. 1-22.

[2] Nelson CA, Monk CS. The use of event-related potentials in the study of cognitive development. In: Nelson CA, Luciana M, editors. Handbook of developmental cognitive neuroscience. Cambridge, MA: MIT Press; 2001. p. 125-36.

[3] Thierry G. The use of event-related potentials in the study of early cognitive development. Infant Child Dev 2005; 14: 85-94.

[4] Hoehl S, Reid V, Mooney J, Striano T. What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. Dev Sci 2008; 11: 10-6.

[5] Parise E, Reid VM, Stets M, Striano T. Direct eye contact influences the neural processing of objects in 5-month-old infants. Soc Neurosci 2008; 3: 141-50.

[6] Farroni T, Mansfield EM, Lai C, Johnson MH. Infants perceiving and acting on the eyes: tests of an evolutionary hypothesis. J Exp Child Psychol 2003; 85: 199-212.

[7] DeBoer T, Scott LS, Nelson CA. Methods for acquiring and analyzing infant event-related potentials. In: de Haan M, editor. Infant EEG and Event-Related Potentials. New York: Psychology Press; 2007. p. 5-37.

[8] Picton TW, Bentin S, Berg P, Donchin E, Hillyard SA, Johnson Jr. R, et al. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 2000; 37: 127-52.

[9] Luck SJ. Ten simple rules for designing ERP experiments. In: Handy TC, editor. Event-Related Potentials: A methods handbook. Cambridge, MA: MIT Press; 2005. p. 17-32.

[10] Reid V, Geangu E. Research methodology and social cognition. In: Striano T, Reid V, editors. Social Cognition: Development, Neuroscience, and Autism, ch. 2. Oxford: Wiley-Blackwell; 2008. p. 17-34.

[11] Nikkel L, Karrer R. Differential effects of experience on the ERP and behavior of 6-month-old infants: Trends during repeated stimulus presentations. Dev Neuropsychol 1994; 10: 1-11.

[12] Snyder K, Webb SJ, Nelson CA. Theoretical and methodological implications of variability in infant brain response during a recognition memory paradigm. Infant Behav Dev 2002; 25: 466-494.

[13] McIsaac H, Polich J. Comparison of infant and adult P300 from auditory stimuli. J Exp Child Psychol 1992; 53: 115-28.

[14] Kushnerenko E, Ceponiene R, Balan P, Fellman V, Huotilaine M, Näätänen R. Maturation of the auditory event-related potentials during the first year of life. Neuroreport 2002; 13: 47-51.

[15] Szücs D. The use of electrophysiology in the study of early development. Infant Child Dev 2005; 14: 99-102.

[16] Reynolds GD, Richards JE. Familiarization, attention, and recognition memory in infancy: an event-related potential and cortical source localization study. Dev Psychol 2005; 41: 598-615.

[17] Richards JE, Reynolds GD, Courage ML. The neural bases of infant attention. Curr Dir Psychol Sci 2010; 19: 41-6.

[18] Sheskin DJ. Handbook of parametric and nonparametric statistical procedures, 5th ed.. CRC Press; 2011.

[19] Hoehl S, Reid VM, Parise E, Handl A, Palumbo L, Striano T. Looking at eye gaze processing and its neural correlates in infancy – implications for social development and autism spectrum disorder. Child Dev 2009; 80: 968-85.

[20] Reid VM, Striano T. The directed attention model of infant social cognition. Eur J Dev Psychol 2007; 4: 100-10.

[21] Courchesne E, Ganz L, Norcia AM. Event-related brain potentials to human faces in infants. Child Dev 1981; 52: 804-11.

[22] de Haan M, Pascalis O, Johnson MH. Specialization of neural mechanisms underlying the face recognition in human infants. J Cogn Neurosci 2002; 14: 199-209.

[23] Reid VM (personal communication) 2010.

[24] Gredebäck G, Melinder A. Infants' understanding of everyday social interactions: a dual process account. Cognition 2010; 114: 197-206.

[25] Gredebäck G, von Hofsten C. Taking an action perspective on infant's object representations. Prog Brain Res 2007; 164, 265-82.

[26] Hoehl S, Striano T. The development of emotional face and eye gaze processing. Dev Sci In press.

The re-analysis of infant ERP-data presented in Study 1 addressed the issue of the nature of data-acquisition within a standard paradigm. The typical data-analysis for infants is currently mainly conducted as would be expected for adult data, with variations such as the number of trials included in each individual average. However, as illustrated by the results of Analysis 2 in this study, the norm in infant ERP-studies often involves the removal of large amounts of data which may in fact contain meaningful information. The main reason for this loss or rejection of potentially usable data is that some participants' artefact-free trials will not be included in the averaging process because an infant failed to provide the minimally required number of analysable trials for all of the experimental conditions. Such lower numbers of artefact-free trials can be the result of multiple factors. For instance, infants may be interested in a stimulus-presentation in the early stages of the test-session but may become bored or distressed later on. As a way of showing their unwillingness to cooperate, an infant may simply not attend towards the stimuli anymore or they may start to get agitated and move around. Alternatively, an infant's data may have to be excluded from a study's analysis because a heritable condition within the family has only been discovered after the data have been collected. Circumstances such as these contribute to a study's attrition rate, which can be relatively high (e.g., Nelson, & Collins, 1991; Parise, Reid Stets, & Striano, 2008). The meta-analysis and the meta-regressions presented in Study 2 aim to examine this particular issue which is present in all infant ERP-studies to varying degrees. By relating different factors that are common features to all ERP-studies to the studies' attrition rates, Study 2 attempts to provide some answers in terms of which of these factors may or may not be contributing to an increase or decrease in attrition. Based on the findings of this study, it may be possible to design studies in such a way as to facilitate infants and their

caregivers behaviour in terms of the requirements for stillness and attention in an EEG-test session, thereby resulting in fewer infants featuring in attrition rates.

References

Nelson, C.A., & Collins, P.F. (1991). Event-related potential and looking-time analysis of infants' responses to familiar and novel events: Implications for visual recognition memory. *Developmental Psychology*, *27*(1), 50-58.

Parise, E., Reid, V.M., Stets, M., & Striano, T. (2008). Direct eye-contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience*, *3*(2), 141-150.

Chapter 3

A Meta-Analysis Investigating Factors Underlying Attrition Rates in Infant ERP Studies

Manuela Stets, Daniel Stahl, and Vincent M. Reid

Developmental Neuropsychology (2012), *37*(3), 226-252.

Abstract

In this meta-analysis, we examined interrelationships between characteristics of infant ERP-studies and their attrition rates. 149 published studies provided information on 314 experimental groups of which 181 provided data on attrition. A random effects meta-analysis revealed a high average attrition rate of 49.2%. Additionally, we used meta-regression for 178 groups with attrition data to analyze which variables best explained attrition variance. Our main findings were that the nature of the stimuli – visual, auditory, or combined as well as if stimuli were animated – influenced exclusion rates from the final analysis and that infant age did not alter attrition rates.

Introduction

Researchers working in the field of developmental psychology have a multitude of tools available to them with which they can investigate their research questions. However, in terms of working with infants and toddlers, the numbers of possibilities at hand are slightly diminished. Some methodologies, such as behavioral reaction-time paradigms and fMRI, are not always fully appropriate for infant populations. This can be due to either ethical issues or simple reasons of practicality. Therefore, techniques such as EEG- and ERP-studies have gained in popularity over the last decades (see Saby, & Marshall, in press, this issue). However, even though EEG-methodology is non-invasive and well-suited for infant populations, the attrition rates reported in research articles can vary dramatically. The reasons for these – sometimes huge – differences in attrition rates have not been investigated to date. In the present meta-analysis, we examine possible explanations for this issue. We are interested in the possible relationships between the features of ERP-studies (see list in Appendix 1) and the attrition rates reported in the papers. We also examine the reasons provided for why data for infants had not been included in the final analyses and if this is related to the design and age of the infant population that is being investigated.

In contrast with ERP-studies, there is a relatively low attrition rate that is usually reported in infant eye-tracking studies, which also investigate infant perception and cognition in the absence of extremely overt behavior. This attrition rate may be related to the apparent ease of data acquisition from this population. For example, Hespos, Gredebäck, von Hofsten, and Spelke (2009) report 52 infants in their sample with the inclusion of 41 of these into their final analysis. Only the data of

11 participants were excluded for reasons of fussiness (n = 5), because of technical problems (n = 4) or because of the infant not completing a behavioral (reaching) task (n = 2). This leaves the authors with data from 78.9% of the total sample and an overall attrition rate of 21.2%. Gredebäck and Melinder (2010) note that in their first experiment only six out of 62 infants failed to contribute data to the final analysis (an attrition rate of 9.7% due to fussiness or failure to calibrate). In their second experiment there was no attrition at all as every 6-month-old who was tested entered the final analysis.

Returning to ERP-methodology, these numbers typically look different. It is relatively unusual for ERP-experiments with infants to feature an attrition rate of under 25%. DeBoer, Scott, and Nelson (2007), for example, state that they expect an attrition rate of between 50 and 75 percent when planning a study (see also Hoehl, & Wahl, in press, this issue). Possible factors leading to such high attrition can be as simple as the challenge of getting the infants to tolerate the electrode mounting system (cap or net) in addition to making the stimulus presentation suitable for the age-group under investigation. For visual stimuli, the paradigm must be created in such a way that they can capture the participants' attention for as long as possible. As has been reported in Stets and Reid (2011; see also Nikkel and Karrer, 1994, and Snyder, Webb, and Nelson, 2002), there is a high potential for infants to habituate to the stimuli such that the amount of attention they allocate to the stimuli can change over the course of a standard experimental session. Such changes can eventually result in major alterations in the ERP-components reported such as a reversal in peak negative amplitude of the conditions under investigation (see Stets and Reid, 2011).

Picton et al. (2000) and Luck's (2005) guidelines on how to conduct an ERP-study attempt to provide researchers with suggestions on how to obtain accurate and

replicable results – mainly from adult populations. That infants and toddlers need to be treated differently in terms of creating the paradigm is also mentioned (see also DeBoer, Scott, and Nelson, 2007), however, Luck (2005) avoids the topic of attrition, whereas Picton et al. (2000) state that the numbers and reasons for attrition should be mentioned as these numbers and reasons may be an indication for the general validity of the reported data.

One study investigating the characteristics of behavioral infant data, which could have implications for our understanding of infant EEG-data, was reported by Slaughter and Suddendorf (2007). These authors collected a total of 187 scientific articles, which had been published between 1985 and 2005, reporting behavioral studies with infants aged 12 months or younger using either a habituation paradigm or a violation-of-expectation paradigm. This review was performed in order to investigate variation in reported attrition rates, taking special note of attrition due to fussiness, and trying to relate this factor to the results reported for the study. Other categories of reasons for attrition which had been included in their analysis were procedural errors, interference by the caregiver, the infant failing to habituate, and a category of "other", which included any other circumstances which could cause a participant's exclusion from the final analysis (e.g., the infant's falling asleep). Despite reporting minor changes in the attrition rates of the behavioral studies over time, Slaughter and Suddendorf (2007) only found a negative correlation between the "other" category of attrition and the age of the infants in days. Their interpretation was that this was due to the infants' lack of self-control in terms of sleepiness – the younger the participant, the higher the chance that the infant will fall asleep before the end of the experimental session and, therefore, to fail to provide enough usable data. All other correlations that were calculated were non-significant: the authors could not

find any relationship between attrition due to fussiness and the outcome of a study. This meta-analysis, while comprising 187 papers, featured 101 of these articles in the final analysis. The remaining 86 articles were excluded due to a number of factors, although the primary reason was that there was insufficient detail, with the total number of participants tested and/or excluded lacking in these articles.

As has been stated above, for ERP-studies, attrition is typically at higher levels than in behavioral studies. When an infant or toddler enters the lab it is often very hard to predict how the experimental session will progress – even for those with many years of experience in laboratory practice. One additional factor may be the participant's temperament. Marshall, Reeb, and Fox (2009) approached this topic by comparing the ERP-responses to an auditory oddball paradigm with 9-month-old infants. One group of 9-month-olds had been rated as highly positive in temperament, a second group had been rated as highly negative in temperament, and a third group had not been rated for temperament prior to participation in the study. The authors could show processing differences toward the same stimuli between the groups. However, what is also apparent from the reported numbers in their methods-section are the differences between the groups in terms of attrition. The group of 9-month-olds who had been rated as being highly negative in temperament by their caregivers was also the group with the highest attrition rate: 52.2% of these participants did not enter the final analysis, primarily due to too many movement artifacts in their EEG-data (the attrition rates for the control group and for the group rated as being highly positive were 32.8% and 40.4%, respectively). This study indicates that ERP-paradigms are generally a mildly stressful experience for infants. Further, this has the consequence that only a subset of infants from a random population will provide sufficient data for analysis and that this is as a function of temperament

characteristics.

Taking Slaughter and Suddendorf's (2007) paper as a model, in the present article we attempted to find relationships between various factors which are necessary parts of an ERP-study and the attrition rates experienced in these studies. Among others, the factors that we investigated were the duration of a single stimulus (i.e., the duration of the presentation of a picture or a sound), the number of experimental conditions, the mean-age of the participants in days, and the duration of a single trial (i.e., the duration of a single stimulus plus the duration of the interval between one stimulus and the next; see Appendix 2 for a complete list of the factors included). To address the limitations in infants' attention span, researchers often try to develop strategies to keep the participants interested in the stimuli for as long as possible. In studies using auditory stimuli only, there is often also some kind of visual stimulation presented to the infants in order to keep them silent, awake and attentive (e.g., a silent puppeteer as in Brannon, Libertus, Meck, & Woldorff, 2008, cf. [6] in Appendix 1; and in Friederici, Friedrich, & Christophe, 2007, cf. [29] in Appendix 1). As a means to keeping his participants alert in a visual recognition memory paradigm, Richards (2003, cf. [114] in Appendix 1) presented his stimuli for 500ms in between sections of the program *Sesame Street* (see also Reynolds, & Guy, in press, this issue). Another relatively common practice in infant ERP-studies is to limit the number of conditions but to have more than one form of stimuli for each of the experimental conditions in order to avoid habituation effects (e.g., Kobiella, Grossmann, Reid, & Striano, 2007). But are some strategies better than others in terms of reducing attrition rates? Do decisions like these have an impact on how cooperative the infants will be in the experimental sessions and on the likelihood for an infant's data to contribute to the final analysis? Finding answers for these questions will be valuable as it increases

our understanding of relations between experimental design, ERP-techniques and attrition rate.

In the current paper we present the results of a meta-analysis on 149 published ERP-studies with infants aged between 0 and 24 months of age. These articles, which had been published between 1978 and 2010, come from a wide variety of theoretical backgrounds. Some investigated cognitive processes in clinical populations compared to typically developing controls (e.g., see papers [7] and [54] in Appendix 1). Some were investigating populations of infants who did not have a clinical condition themselves but their familial background showed some special condition such as the mother suffering from anxiety during pregnancy (e.g., see paper [49] in Appendix 1), having a sibling who is on the Autism spectrum (see [25] in Appendix 1), or the mother being diabetic (e.g., see [22] in Appendix 1). Furthermore, researchers reported studies in which some other condition was present which could potentially affect the infant's performance in a specific task (e.g., the mother having been smoking during pregnancy; see [69] in Appendix 1). Following Slaughter and Suddendorf's (2007) approach, we hypothesized that there would be a link between the specific features of the infants, particularly their age, the features of the ERP-study they participated in (e.g., the duration of the stimuli and trials, the number of conditions, the number of different stimuli per experimental condition), the attrition rate of the study, and the stated reason(s) for attrition. Slaughter and Suddendorf (2007) were investigating connections between a study's reported outcome and attrition rates due to fussiness. We are examining the parameters of the study itself that might act as possible indicators for an expected attrition rate when contrasted with other possibilities such as shorter versus longer trials and the age of the infants.

Methods

For the present meta-analysis we collected data on a wide range of factors from a total of 149 research articles presenting ERP-studies with infants aged between 0 and 24 months. These articles could be separated further into a total of 314 experimental groups consisting of infant populations (some of the papers also reported data from adult control groups but we did not take any note on these). These 314 experimental groups consisted of three different categories of infants: Category 1 (278 experimental groups) were typically developing infants with no further information on any kind of circumstances, which would mark these groups as special in any way (see Category 1 in Table 1). Category 2-groups (25 of the 314) were classified as being typically developing but being marked by a special feature (see Category 2 in Table 1; e.g., a family background of dyslexia or the mother suffering from anxiety for instance; see [145] and [49] in Appendix 1, respectively). The experimental groups belonging to Category 3 (11 out of 314) consisted of clinical populations (see Category 3 in Table 1; e.g., infants born preterm, cf. [133] in Appendix 1; or infants suffering from brain-iron deficiency, see [7] Appendix 1). We took note of characteristics such as these and controlled for them in our analyses, but as the category of the experimental group did not have an impact on the results, we decided to treat the three groups in the same way in our analysis ($F(2, 178) = 0.69$, $p = 0.50$, see Table 3 for multivariate results). In order to include the age of the infants into our analysis, we ordered the experimental groups according to the mean-ages reported in the articles, separating them out per month in most cases. For an overview of the numbers of experimental groups included in our analysis per category and age-group, see Table 1.

Table 1. An Overview of the Numbers of Experimental Groups Separated by Category and Age-Group.

	Newborn	*1 m – 2 m*	*3 m*	*4 m*	*5 m*	*6 m*	*7 m*	*8 m*	*9 m*	*10 m*	*11 m*	*12 m*	*13 m – 16 m*	*17 m – 24 m*	*Random*	*Total*
Number of research articles	10	8	18	27	7	35	18	14	13	5	6	13	6	12	3	195
Number of typically developing groups – Category 1	11	17	30	35	9	51	26	18	20	11	8	18	7	14	3	278
Number of typically developing groups with additional features – Category 2	6	1	0	0	1	4	0	2	3	1	0	4	0	1	2	25
Number of clinical groups – Category 3	3	0	1	1	0	2	0	0	1	0	0	3	0	0	0	11
Total number of experimental groups	20	18	31	36	10	57	26	20	24	12	8	25	7	15	5	314

Note. The total number of research articles reported is higher in this table compared to the total number of articles present in the meta-analysis. This is due to the fact that some articles reported data for more than one age-group.

As has been mentioned previously, our aim was to find potential links between the ages of the infants tested, the features of the studies (e.g., the duration of a single trial and the number of experimental conditions), the overall attrition rates reported, and the attrition due to a specific reason (e.g., attrition due to fussiness as opposed to a lack of artifact-free trials). Therefore, we took note of the information provided in the articles, which were related to these variables. In some cases, when the authors of an article did not provide a full account on all factors, we tried to reconstruct the respective values from the information which had been provided (for a complete list of the factors under investigation, see Appendix 2). For instance, if the duration of a single stimulus, the duration of the inter-stimulus interval, and the total number of stimuli had been given, we recalculated the duration of a single trial and the respective duration of the entire stimulus presentation based on these data.

The same reconstruction of data was conducted, where necessary, for the total number of participants tested when only the numbers were given for the number of participants included in the analysis of the study and the number of participants excluded. However, after this process, we were still left with a relatively large number of experimental groups for which we were unable to reconstruct required information. Therefore, we sent out requests to the corresponding authors of several articles published between 1995 and 2010 asking for clarification or data to fill the gaps for the experimental groups in question. In most of the cases, we were asking for the number of participants who had not entered their final analysis and how many of them had failed to do so for what reason. In some cases, when only a general number for the attrition rate had been given, we were asking if it would be possible to clarify the numbers and reasons for attrition. Occasionally, we also asked for the duration of a single trial in case it could not be reconstructed due to lack of detail in the description

of the stimuli or procedure. We sent out emails to a total of 30 researchers asking for information on 49 of the 149 articles that we had collected. There were more articles for which we would have needed further information, however, we decided not to ask for these data due to the respective articles having been published before 1995 and the common practice in terms of duration of data-storage. About two thirds of the researchers responded to our emails, with about one third of responses contributing data for the present meta-analysis. Finally, due to all of the above-mentioned reasons, the present meta-analysis contains results for 181 out of the total of 314 experimental groups collected (57.6%). The other 133 groups (42.4%) were excluded from further analysis.

Attrition rate

Meta-analyses were carried out using Stata 10.1 and 11.1 (StataCorp, College Station, TX, USA) using the user-contributed commands for meta-analyses: metan (Bradburn, Deeks & Altman, 1998), metainf (Tobias, 1999), metabias (Steichen, 1998), and metareg (Harbord and Higgins, 2008).

The primary outcome was proportion of attrition, defined as the number of infants who were not included in the final analysis of a study divided by the total number of infants who attended the study. Following the recommendations of Lipsey and Wilson (2000), a logit transformation was applied to the proportion estimates to form an unbounded (in contrast to the 0 to 1 bounded nature of proportions) effect size estimator:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$$

where p is the proportion and ln is the natural logarithm. The standard errors were computed as follows:

$$SE_{logit(p)} = \sqrt{\frac{1}{np(1-p)}}$$

where n is the sample size.

The standard errors were used to calculate upper and lower 95% confidence limits of the effect size estimator by logit(p) ±1.96*SE. The inverse of the variance ($1/SE^2$) was used as the basis of the weight for the meta-analysis and meta-regressions. In one study there was no attrition and we added 0.5 so that the logit transformation results in quantities that can be defined. The final pooled logit results and 95% confidence intervals of the meta-analyses were back-transformed to proportions for ease of interpretation.

Meta-analysis

The effect sizes and standard errors of the studies were pooled using random-effect models. A random-effect meta-analysis model assumes that in addition to the presence of a random sampling error, variability of mean effect sizes is also caused by differences of the effect between studies due to differences in study populations and procedures (between-study heterogeneity). Random-effects models result in estimates with wider confidence intervals than fixed-effects models, but are more realistic in this meta-analysis due to the variety of case mix and settings between studies (Everitt, 2003).

Homogeneity between the studies was analyzed using Cochran's Q test and by calculating the measure of heterogeneity or inconsistency I^2 (Higgins et al., 2003). I^2 describes the percentage of total variation across studies that is due to heterogeneity

rather than sampling error and ranges between 0% (no inconsistency) and 100% (high heterogeneity) with values of 25%, 50% and 75% suggested as low, moderate, and high heterogeneity (Higgins et al., 2003).

The analyses were repeated excluding one study at a time to investigate the influence of each individual study on the overall meta-analysis summary using STATA's user written function *metainf* (Tobias, 1999).

Selection bias

It is well-known that statistically significant results are more likely to be published than studies with non-significant results. In our meta-analysis this could mean that a study with a high proportion of attrition and, therefore, smaller sample size was less likely to result in significant findings of their main outcome and, hence, less likely to be published. The presence of a selection bias was assessed informally by visual inspections of funnel plots, scatter plots in which the treatment effects estimated from individual studies are plotted against a measure of study precision. We followed the recommendations in the relevant literature (cf., Sterne & Egger, 2001; Sterne, Egger, & Smith, 2001) and plotted the logarithm of odds rations of the studies (vertical axis) against their standard error (horizontal axis in reversed order; see Figure 1). Odds of attrition rates were plotted on a logarithmic scale (logits), so that effects of the same magnitude but in opposite directions (such as 0.5 and 2) are equidistant from 1.0. We included lines which represent 95% confidence intervals around the estimated summary effect size in the funnel plot to facilitate interpretation. These lines show the expected distribution of studies in the absence of heterogeneity or of selection biases (Sterne et al., 2001). Asymmetrical funnel plots may indicate a selection bias. For instance, the absence of studies in the right bottom corner (low

precision and small effect sizes) of a funnel plot is usually seen as an indication of a classical publication bias. Studies of low precision are usually studies with a small sample size and, consequently, little statistical power. However, for other effect sizes, such as Cohen's d, the effect sizes will not scatter randomly above and below the mean with the spread increasing with an increasing standard error. Because of the mathematical association between the standard error and the logits, we expect the effect sizes to branch away from the mean with increasing standard error. However, branches will be symmetrical in the absence of a selection bias.

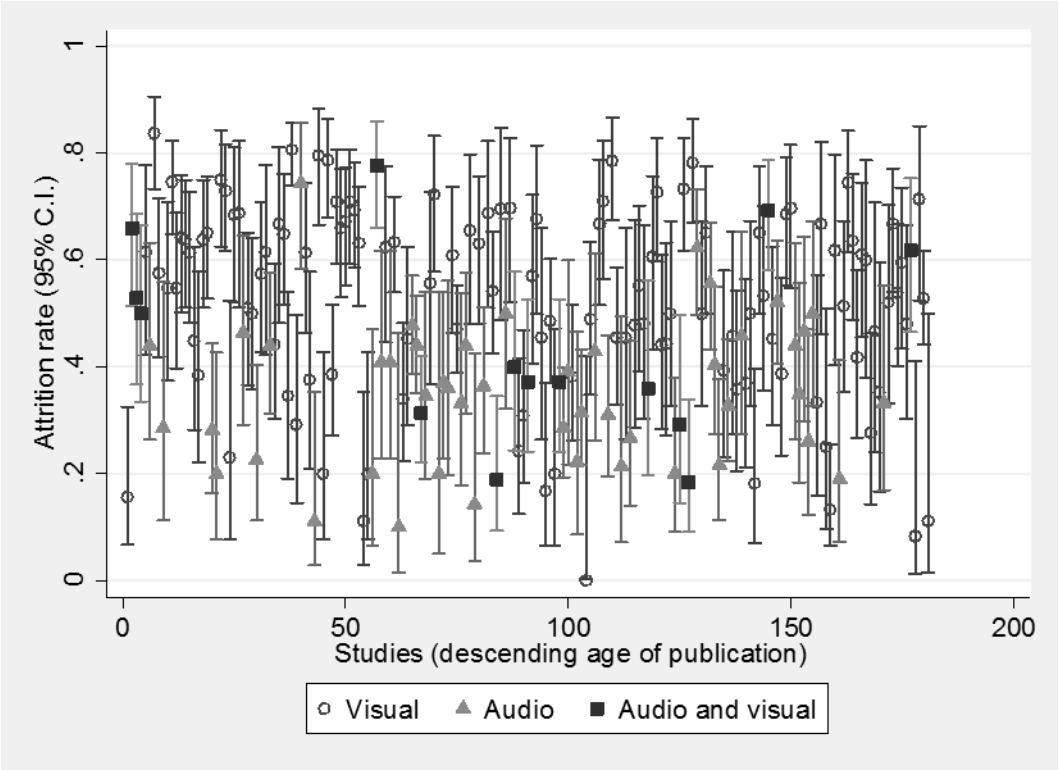

Figure 1. Proportion of attrition (95% C.I) of 181 experimental groups. Data are plotted in decreasing order of publication-age with the most recently published ones depicted on the right. Symbols of means refer to the different kinds of stimuli used: circles for "Visual", triangles for "Auditory", and squares for "Audio *and* visual".

Meta-regression

The meta-analyses were followed up with a meta-regression analysis on 178 experimental groups, an extension of the meta-analysis in which the relationship between attrition rates and explanatory variables (study characteristics) are modeled using weighted regression methods. Random-effects models were used to allow for the residual heterogeneity among attrition rates not modeled by the explanatory variables (Thompson and Sharp, 1999). The explanatory variables are characteristics of studies that might influence the extent of the attrition rate. In our main-analysis we assessed the age of the paper, the type of the experimental group (typically developing, typically developing with an additional special characteristic, or clinical population), the mean age of the infants in months, the total number of infants analyzed in the experimental group, the number of conditions, the type of stimulus (auditory, visual, or visual *and* auditory), and if animation was used (yes/no). Preliminary analyses suggested a nonlinear asymptotic relationship between the total number of infants and attrition, and a quadratic term for the total number of infants was included in the regression model. Because the number of infants can be caused by a high attrition rate (the number of infants was not fixed but more infants were used if attrition rate was higher) we reran the model without this variable.

In a secondary analysis we assessed the effect of variables for which information was available only for a smaller number of studies. These variables were: i) the minimum number of trials per condition required to be included in the final analysis, ii) the mean-duration of a stimulus in ms, iii) the mean-duration of a single trial in ms, and iv) the total mean-duration of the stimulus presentation in min. Each variable was assessed separately in the main model. Again, the regressions were rerun without total number of infants included.

Because the analysis is based on logits of attrition rate, odds ratios and 95% confidence intervals (95% C.I.) are reported. For the main analysis, adjusted *p*-values for multiple testing based on a permutation procedure as described by Higgins and Thompson (2004) are also presented. The assumptions of the meta-regression were confirmed by assessing standardized shrunken residuals as recommended by Harbord and Higgins (2009).

Results

Of the initially 314 identified experimental groups, 181 provided the number of infants analyzed in the study and the number actually tested and, therefore, allowing us to calculate the attrition rate. Table 2 below provides details of these 181 experimental groups. The age of the papers ranged between recently published and 33 years. 120 (66.3%) experimental groups experienced visual, 47 (26.0%) auditory, and 14 (7.7%) both visual *and* auditory stimuli. 14 (7.7%) experimental groups experienced animated visual stimuli.

The average number of infants entering the final analysis of the 181 experimental groups was 19.44 (*SD* = 10.41) with a range from 7 to 72 infants. The stimuli were presented on average for 9.2 min (range: 1.3 to 37.3 min, *n* = 128). The mean total attrition was 47.3%, ranging from 0% to 83.8%. Figure 1 shows the attrition rate and 95% confidence intervals of the 181 experimental groups arranged by the age of the study.

A random-effects meta-analysis revealed an overall pooled attrition rate of 49.16% (95% 46.5% to 51.85%). A sensitivity analysis to assess the robustness of the estimate was examined by sequentially removing each study and reanalyzing the remaining data sets. The estimated effect sizes ranged from 48.95% to 49.46% and, therefore, none of the experimental groups changed the estimate by more than +/- 0.3%.

Table 2. Overview of the 181 Experimental Groups Used in the Meta-Analysis and Key Variables Included.

Continuous Variables					
Variable	*n*	*Mean*	*Std. Dev.*	*Min*	*Max*
Age of paper in years	181	6.58	6.06	1	33
Age of infants in days	173	212.5	119.59	7.6	730
Total number of infants tested	181	41.02	23.92	9	175
Total number of infants included in analysis	181	19.44	10.41	7	72
Relative attrition	181	0.47	0.187	0	0.84
Number of conditions	181	2.39	0.726	1	6
Minimal number of usable trials requires for average	132	20.07	25.15	5	100
Number of trials included in average	124	65.4	182.19	8	1186
Duration of stimulus in ms	156	874.6	839.06	50	4000
Duration of trial in ms	122	2283.6	1126.16	600	6150
Total mean duration of presented stimuli in min	128	9.20	7.07	1.3	37.3
Number of stimuli presented	131	117.8	202.39	1	800
Categorical variables	*n*	*%*			
Nature of experimental group					
Typically developing	166	91.7			
Typically developing with special feature	10	5.5			
Clinical	5	2.8			
Nature of stimuli					
Visual	120	66.3			
Auditory	47	26.0			
Visual and auditory	14	7.7			
Animated stimuli					
No	167	92.3			
Yes	14	7.7			

Heterogeneity between the studies was statistically significant ($Q(180) = 822.18$, $p < 0.0001$) and large in magnitude ($I^2 = 78.1\%$). The following meta-regression tried to explain the large heterogeneity in attrition rate between the studies.

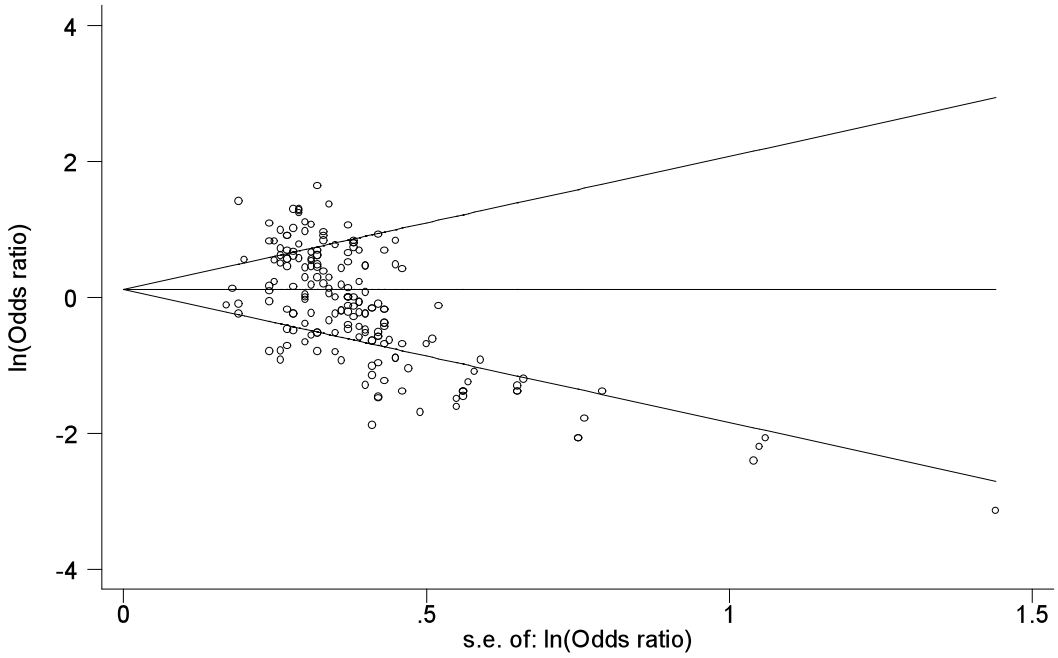

Figure 2. Begg's funnel plot with pseudo 95% confidence limits for meta-analysis. The standard error of the effect sizes (natural logarithm of the odds or attrition or logit) on the horizontal axis are plotted against the effect size on the vertical axis. The dots correspond to the effect sizes of the individual experimental groups and the diagonal lines show the expected 95% confidence interval around the summary estimate (horizontal line). The logits correspond to the percentages of attrition as follows: 2 = 88%, 1.5 = 82%, 1 = 73%, 0.5 = 62%, 0 = 50%, -0.5 = 38%, -1 = 27%, -1.5 = 18%, and -2 = 12%.

Selection bias

Figure 2 shows Begg's funnel plot to assess for a possible selection bias. There is some asymmetry of the distribution of the effect sizes: studies with a larger standard error tend to have disproportionately smaller attrition rates, whereas studies with high attrition rates seem to have smaller standard errors. This suggests that

studies with smaller sample sizes have disproportionately smaller attrition rates than would be expected in the absence of a selection bias.

Meta-regression

178 experimental groups provided data on the main predictor variables and could therefore be used for the meta-regression. The meta-regression model explained 53.5% of the (adjusted) between group variance. The analysis of the main model revealed that older studies showed higher proportions of attrition than more recent ones (see Table 3). Furthermore, studies using visual stimuli were twice as likely to cause attrition than auditory studies, while studies with visual *and* auditory stimuli did not differ significantly from studies with purely auditory stimuli. Studies with visual stimuli were more likely to attrition than studies with visual *and* auditory stimuli (*Odds ratio* = 1.57; 95% *C.I.* = 1.11 - 2.22, p = 0.011; 95% *C.I.* = 1.10 – 2.57, p = 0.016; Marginal mean-proportions of attrition: 35.1% for Auditory, 53.7% for Visual, and 45.0% for Visual *and* Auditory). Finally, the mean-proportion of attrition rises significantly with increasing total number of infants analyzed in the study and tends to rise with the use of animated stimuli. All other variables showed little influence on attrition rate and were not significant. Except age of paper, all significant variables remained significant after corrections for multiple testing.

Rerunning the meta-regressions without the total number of infants analyzed in the study resulted in slightly stronger effects of age of the study and type of stimuli (marginal mean-proportions of attrition: 36.2% for Auditory, 55.0% for Visual, and 45.3% for Visual *and* Auditory). The effect of attrition was slightly reduced and not significant anymore. The overall model explained less between study variance (adjusted r^2 = 28.1%). All non-significant variables remained non-significant.

Table 3. Results of the Meta-Regression of the Main-Model (n = 178) Once With the Number of Analyzed Infants Included in the Model (see Table 3a) and Once Without the Number of Analyzed Infants Included in the Model (see Table 3b).

Table 3a. The Model With the Number of Analyzed Infants Included (Quadratic Term)

Variable	*Odds Ratio (95% C.I.)*	*T*	*p (adjusted p)*
Age of paper in years	1.02 (1 – 1.036)	2.39	.0018 (.16)
Nature of experimental group (F(2, 167) = 0.60, p = .55)			
Typically developing	1		
Typically developing with special condition	0.91 (0.61 – 1.34)	-0.49	.62 (.99)
Clinical	0.72 (0.38 – 1.38)	-0.99	.33 (.97)
Infant mean-age in months	0.98 (0.96 – 1.01)	-1.32	.19 (.86)
Total number of participants tested (F(2, 169) = 31.58, p < .0001)			
Total number of participants	1.03 (1.02 – 1.04)	5.71	< .0001 (< .001)
Total number of participants^2	0.999 (0.999 – 0.999)	-3.39	.001 (.008)
Number of conditions	0.95 (0.84 – 1.08)	-0.79	.43 (.99)
Nature of stimuli (F(2, 167) = 16.69, p < .0001)			
Visual	1.82 (1.47 – 2.27)	5.55	< .0001 (< .0001)
Auditory	1		
Visual and auditory	1.16 (0.80 – 1.70)	0.79	.43 (.99)
Animated Stimuli	1.40 (1.0 – 1.96)	1.98	.04 (.38)
Non-animated stimuli	1		

(Continued)

Table 3. (Continued)

Table 3b. The Model Without the Number of Analyzed Infants Included

Variable	*Odds Ratio (95% C.I.)*	*T*	*p (adjusted p)*
Age of paper in years	1.03 (1 – 1.047)	2.89	.0004 (.025)
Nature of experimental group ($F(2, 169) = 1.57$, $p = .21$)			
Typically developing	1		
Typically developing with special condition	1.24 (0.78 – 1.96)	-0.91	.36 (.96)
Clinical	0.57 (0.27 – 1.20)	-1.49	.14 (.66)
Infant mean-age in months	1.0 (0.98 – 1.03)	0.16	.87 (.99)
Number of conditions	0.97 (0.84 – 1.12)	-0.43	.66 (.99)
Nature of stimuli ($F(2, 169) = 18.1$, $p < .0001$)			
Visual	2.1 (1.63 – 2.70)	5.8	< .0001 (< .0001)
Auditory	1		
Visual and auditory	1.26 (0.81 – 1.98)	1.02	.31 (.93)
Animated Stimuli	1.31 (0.88 – 1.95)	1.35	.18 (.75)
Non-animated stimuli	1		

Note. The effect sizes are Odds ratios (*OR* > 1 means an increased proportion of attrition rate with increasing values).

Secondary meta-regression analysis

The main model was rerun with each of the secondary variables. Each variable was assessed separately for models with and without the number of analyzed infants included as a covariate. None of the variables showed a significant effect on proportion of attrition: i) minimum number of trials per condition required to be included in the final analysis: with number of analyzed infants included: *OR* = 1.003, *p* = 0.31, *n* = 129, without the number of analyzed infants: *OR* = 1.001, *p* = 0.78; ii) mean-duration of stimulus in ms: with number of analyzed infants included: *OR* = 1.00, *p* = 0.56, *n* = 153, without the number of analyzed infants: *OR* = 1.00, *p* = 0.66; iii) mean-duration of single trial in ms: with number of analyzed infants included: *OR* = 1.00, *p* = 0.65, *n* = 119, without the number of analyzed infants: *OR* = 1.00, *p* = 0.56; and iv) total mean-duration of stimulus presented in min: with number of analyzed infants included: *OR* = 1.01, *p* = 0.48, *n* = 126, without the number of analyzed infants: *OR* = 1.01, *p* = 0.65. The additional unique explained variance of the assessed variables was less than 1%.

Furthermore, we explored if some of the effects of age of paper and type of stimuli can be partially explained by an association with one of the other variables. Rerunning the same analyses without type of stimulus (and without the number of infants analyzed) revealed that the total duration of the stimulus-presentation in minutes and the number of trials required for the averaging process had a significant effect on attrition. The longer the mean-duration of the stimulus-presentation, the lower is the attrition rate (*Odds ratio*: 0.97, 95% *C.I.*: 0.95 - 0.99, *p* = 0.002). The total duration of the stimulus-presentations differed significantly between all three conditions with Visual = 6.1 min (*SD* = 3.74, *n* = 83), Auditory = 15.8 min (*SD* =

8.83, n = 34), and Visual *and* Auditory = 11.8 min (SD = 5.28, n = 11). Pair-wise comparisons revealed that all $p < 0.05$.

The second variable, the minimal number of artifact-free trials required per condition for the averaging, correlated strongly with the total mean-duration of the stimulus-presentation ($r = 0.79$, $p < 0.0001$, $n = 96$). Consequently, the more artifact-free trials were required for the averaging process in a specific study, the lower was the attrition rate experienced (*Odds ratio*: 0.99, 95% *C.I.*: 0.98 – 0.995, $p < 0.001$, n = 129). However, this requirement for specific numbers of artifact-free trials was significantly higher for studies using purely auditory stimuli compared to those with purely visual stimulation or those using visual *and* auditory stimuli (mean-numbers of artifact-free trials required for averaging for Visual 10.2 [SD = 2.8, n = 95], for Auditory 52.1 [SD =39.0, n = 29], and for Visual *and* Auditory 20.7 [SD = 10.2, n = 10]). Similar results were obtained when the number of infants analyzed for an experimental group was included in our analyses. The effect of the age of the published article could not be partially explained by any other variable.

Discussion

In this meta-analysis, we attempted to find relationships between the attrition rates reported in published ERP-studies using infant populations and a wide range of variables that are common to all of these studies. Amongst others, these variables were the duration of stimuli and trials, as well as the age of the publication in years or the class of stimuli presented (e.g., animated vs. non-animated). One of the findings of the meta-regression was the lower overall attrition rate for experimental groups who had been presented with auditory stimuli only (18.6% lower in comparison to visual stimuli only), and the slightly lower overall attrition rate for studies using combined visual *and* auditory stimuli (8.7% lower in comparison to purely visual stimuli). The age of the paper also contributed, with older articles reporting higher attrition rates. Furthermore, the number of infants tested and animation affected attrition rate. However, the effect of animation was only a trend and the effect of number of infants tested is not easy to interpret as we will discuss later. This lack of other effects may be due to the number of published articles that were not included in the analysis, due to omission of relevant information. Therefore, care needs to be taken when stating that other factors, such as the duration of the stimuli, do not have any effect on attrition rates. The total duration of the stimuli (or the minimum number of trials which correlated strongly with the former variable) could partially explain the observed differences between the study types. Apparently, studies with auditory stimuli on average featured longer stimulus-presentations than studies with visual stimuli. However, studies using purely visual stimulation usually had higher attrition rates compared to those with purely auditory stimulation. This effect is not seen within the type of study as the non-significant effect of total duration of stimulus-presentation

revealed. This surprising relationship between longer test-sessions with auditory stimuli and lower attrition rates and shorter test-sessions with visual stimuli and higher attrition rates might be due to other factors resulting from the paradigms used. Further analyses might indicate other interrelationships which have not been investigated in the present meta-analysis.

One explanation for the variation between stimuli classes may be that many of the studies using auditory stimuli had been conducted with infants who were sleeping during the experimental session. This reduces the likelihood of movement artifacts to a great extent. Also, in cases where participants were awake during the experimental session, paradigms often included distracting visual stimuli in order to keep the infants calm during the recording of the EEG (e.g., see [6] in Appendix 1). That a combination of auditory *and* visual stimuli does elicit a lower attrition rate than visual stimuli only, irrespective of the age of the infant population tested, may seem surprising. Instead of making the task of attending two streams of stimuli at the same time more complex for the infants, the combination of components seems to help the infants to attend to stimuli for a relatively longer duration of time. A possible explanation for this may be that the synchrony of information from multiple modalities acts to enhance the infant's limited attention capacities by providing extra cues for infants and by acting as a more attractive paradigm than single modality presentations. This is also in line with the intersensory redundancy hypothesis proposed by Bahrick and Lickliter (2000). According to this hypothesis, information that is presented synchronously in more than one modality draws the attention of the infants away from less salient information and toward the more salient one which, in turn, is thought to impact on learning and memory processes. Our data suggest that it is more difficult to collect data with purely visual stimulation when contrasted with

purely auditory stimuli. This could be seen as an indicator for the visual stimuli requiring a relatively high level of explicit attention when contrasted with auditory stimuli, which require relatively less effort in terms of attention.

The static versus dynamic nature of the stimuli also had an impact on reported attrition rates: animated visual stimuli appear to be a larger challenge for infants than static stimuli alone. This result may be due to the increased length of these purely visual stimuli when contrasted with the mainly static visual stimuli alone, although this relationship was not noted to be significant in our analyses. Often such stimuli display short film clips with a longer duration per individual stimulus presentation than the duration for a static image. For example, Marshall and Shipley (2009; see [81] in Appendix 1) featured a mean-stimuli length of 2150 ms whereas studies by Nelson (e.g., Nelson, Henschel, & Collins, 1993; see [94] in Appendix 1) typically feature a stimuli length of 500ms.

Another interesting finding was that attrition rates tended to be higher for experimental groups or studies for which a large number of participants had been tested. Representative examples of this include de Haan, Pascalis, and Johnson (2002; [20] in Appendix 1) with an attrition rate of 80.6% (175 infants were tested of whom 141 were excluded from the final analysis); Hirotani, Stets, Striano, and Friederici (2009; [56] in Appendix 1), attrition rate 69.3% (75 infants tested; 52 excluded); Nelson and Collins (1991; [91] in Appendix 1), attrition rate 83.8% (74 infants tested; 62 excluded); and Parise, Reid, Stets, and Striano (2008; [103] in Appendix 1), attrition rate 78.3% (69 infants tested; 54 excluded). For papers [20] and [91], in which the overall attrition rates lie above 80%, these high percentages are explained with the "strict standard for obtaining sufficient trials to include in ERP averages" (de Haan, Pascalis, & Johnson, 2002, p. 207), and with the "strict requirements on what

constituted satisfactory completion of the study" (Nelson, & Collins, 1991, p. 52). In [103] it is stated that "task demands related to the paradigm" were the reason for the high attrition rate (Parise, Reid, Stets, & Striano, 2008, p. 3). One feature, which is common to all of these three sample studies, is that the participant populations consisted of infants ranging between 5 and 6 months of age. This may be an indication that the paradigm was not age appropriate. Consequently, instead of adjusting either the age of the population tested or the paradigm, more participants were tested in order to obtain a large enough sample for statistical power. This, however, may draw the results of these studies into question. For example, if the paradigm is not age appropriate, then there is the potential for some unmeasured variable to determine which infants have the capacity to attend to a sufficient number of trials for inclusion in the final sample. It may be that those infants who were more developmentally advanced than the norm for the tested age group were disproportionately incorporated into the final sample. It must be stressed that the end result of such experiments is not in doubt, but the age at which infants have the capacity to process the presented stimuli may potentially be drawn into question.

A number of articles indicate that it is seen as acceptable within the field of infant ERPs that an attrition rate of between 50 and 75 percent is to be expected when the study is still in the planning phase (DeBoer, Scott, & Nelson, 2007). Therefore, it is not surprising that sentences stating that the reported attrition rate in a given study is "typical of infant ERP studies" (Snyder, Garza, Zolot, & Kresse, 2010, p. 493) and "comparable to other studies using electrophysiological measurements in young infants" (Halit, Csibra, Volein, & Johnson, 2004, p. 1230) are often found in published studies. The issue then becomes whether this general acceptance of attrition rates of over 50 percent should be treated as something normal and inevitable when

running ERP-studies with infant populations. Recent work suggests that such an acceptance is not required. Stahl, Parise, Hoehl, and Striano (2010) used advanced statistical methods to incorporate a larger proportion of infants into final analyses. Stets and Reid (2011) found that it is possible to extract meaningful data from participants who would not be able to contribute data to the grand average of a given study when using the traditional ways of data analysis. The authors reanalyzed previously published EEG-data from 4-month-old infants and created averages from set numbers of trials for each participant. They found that, when examining only the first three artifact-free trials, results were significant. This result might have implications for the general practice of requiring at least 10 or 15 artifact-free trials per condition per participant in order to include data to a study's final analysis. Stahl et al. (in press, this issue) discuss the possibilities of using machine learning methods to analyze single trials ERP-data to reduce bias and to increase powers.

We also investigated whether a selection bias might be present in our sample. The relationship between sample size and attrition rate would also explain the apparent selection bias in the meta-analysis. Studies with smaller sample sizes usually have larger standard errors and, therefore, provide less precise estimates of the effect size. Our assessment suggests that less precise studies with larger attrition rates are not published. However, this bias could explain why researchers tend to collect more data if the attrition rate is high. This also explains why there are not more studies that are low in precision with low attrition rates than would be anticipated in the field. Those common features between studies that might lead to the inclusion of only a small number of participants' data are currently unknown.

Finally, collecting the data for the present meta-analysis has shown that there is large variation between research labs working in the field of developmental

cognitive neuroscience in terms of reporting attrition rates. It should be emphasized that 42.4% of the experimental groups or studies that should be incorporated in this meta-analysis had to be excluded due to a lack of detail in the method of the published research article. In 2000, Picton et al. stated that articles should be clear about these details, yet the number of papers published over the last decade containing all relevant information was not as high as should be the case. At the very least, reviewers of published articles must be vigilant for missing information related to attrition as these variables have large implications for the validity of reported data.

Recommendations for ERP Paradigm Construction

In this meta-analysis, we did not find clear indications for a relationship between the age range of the population tested for a specific study and the attrition rate reported for that study. Such a finding is in parallel to Slaughter and Suddendorf (2007), who could not draw final conclusions about relationships between the validity of a study's results and the attrition rate reported for behavioral studies with infants. In all likelihood this is most probably due in both cases to the amount of missing variables present in the dataset. However, we could find indications that some essential features of a paradigm are more likely to increase the attrition rate experienced for a specific experimental group. A paradigm with purely visual stimulation is clearly not as engaging as a paradigm in which auditory stimulation is also present. We therefore conclude that it is important to create ERP-studies for infant populations with multimodal stimuli, in order to reduce attrition rates.

Our analyses also revealed that studies or experimental groups with larger sample sizes in general tended to have a higher attrition rate in comparison to those with smaller sample sizes. In most cases the reasons for excluding infants from the final analysis of a given study are movement artifacts and fussiness of the infant, resulting in an inability to collect the required minimum number of trials per condition from the excluded infant. Therefore, considering these high rates of attrition, it draws the validity of these paradigms into question with respect to their appropriateness for the population tested. The practical conclusion that can be drawn from these data is that should such attrition rate patterns start to occur during collecting data, then it is reasonable to cease data collection and attempt to adjust components of the paradigm or alter the age group tested.

We did not find any significant relationships between the attrition rate that was reported for a given study and any other factors that we included in our analysis (e.g., the mean-age of the infants, the duration of a single stimulus or trial, or the number of conditions presented). The question remains why we did not find more relationships. We conclude that – other than the features we investigated here – there must be other factors involved that impact on an infant's likelihood to contribute data to a study's final analysis. We have noted in our laboratory that data collection at 9 months is problematic as infants are more irritable and often come to the laboratory in a tired state. This may be related to the end of maternity leave occurring around this time in our country. Another possible candidate is the extent of experimenter experience, with large variation seen between individuals.

Clearly, researchers working in the field of developmental cognitive neuroscience must meticulously plan and prepare an ERP-study with infants aged between 0 and 24 months. One of their biggest challenges is to produce valuable findings without losing sight of the practical components of data collection. However, these elements of studies are often not mentioned in published articles. This hinders researchers from replicating the findings and makes it difficult for independent evaluation of the validity of the reported experiment. Therefore, our conclusion from the present meta-analysis is that the field would be facilitated if a clear consensus were presented with respect to which variables are to be reported in a research article. Factors that are necessary for replication include specific information on the numbers of participants tested and how many entered the final analysis. Some information on why infants were excluded would also be beneficial as well as descriptions of all aspects of the paradigm. These should include the duration of the stimuli and of the inter-trial intervals and the number different stimuli used per condition, for instance.

Finally, information on the average number of trials per condition contributed by each infant (including the standard deviations) and information on the minimum criteria for inclusion would be beneficial. This will put the results reported in a given article into perspective and make them more transparent for others, thereby facilitating the growth and maturity of the field as a whole.

Acknowledgements

We want to thank all the authors of the research articles used in this meta-analysis for providing us with as much detail as possible, and especially those who provided us with additional information for the extra-effort that they took.

References

Bahrick, L.E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*, 190-201.

Bradburn, M.J., Deeks, J.J., & Altman, D.G. (1998). sbe24: metan – an alternative meta-analysis command. *Stata Technical Bulletin*, *44*, 1-15.

Brannon, E.M., Libertus, M.E., Meck, W.H., & Woldorff, M.G. (2008). Electrophysiological measures of time processing in infant and adult brains: Weber's law holds. *Journal of Cognitive Neuroscience*, *20*(2), 193-203.

DeBoer, T., Scott, L.S., & Nelson, C.A. (2007). Methods for acquiring and analyzing infant event-related potentials. In M. de Haan (ed.), *Infant EEG and event-related potentials* (5-37). Hove: Psychology Press.

de Haan, M., Pascalis, O., & Johnson, M.H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199-209.

Everitt, B.S. (2003). Modern Medical Statistics: A Practical Guide. London: Arnold.

Friederici, A.D., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, *17*, 1208-1211.

Gredebäck, G., & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition*, *114*, 197-206.

Halit, H., Csibra, G., Volein, A., & Johnson, M.H. (2004). Face-sensitive cortical processing in early infancy. *Journal of Child Psychology and Psychiatry*, *45*(7), 1228-1234.

Harbord, R., & Higgins, J.P.T. (2008). Meta-regression in Stata. *Stata Journal*,

8(4), 493-519.

Hespos, S., Gredebäck, G., von Hofsten, C., & Spelke, E.S. (2009). Occlusion is hard: Comparing predictive reaching for visible and hidden objects in infants and adults. *Cognitive Science*, *33*, 1483-1502.

Higgins, J.P.T., & Thompson, S.G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine*, *23*(11), 1663-1682.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., & Altman D.G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557-560.

Hirotani, M., Stets, M., Striano, T., & Friederici, A.D. (2009). Joint attention helps infants learn new words: event-related potential evidence. *NeuroReport*, *20*(6), 600-605.

Hoehl, S., & Wahl, S. (in press). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, this issue.

Kobiella, A., Grossmann, T., Reid, V.M., & Striano, T. (2008). The discrimination of angry and fearful facial expressions in 7-month-old infants: An event-related potential study. *Cognition and Emotion*, *22*(1), 134-146.

Lipsey, M.W., & Wilson, D.B. (2000). Practical Meta-analysis. Thousand Oaks, CA: Sage Publications.

Luck, S.J. (2005). Ten simple rules for designing ERP experiments. In: Handy, T.C. (ed.), *Event-Related Potentials: A methods handbook*. (p. 17-32). Cambridge, MA: MIT Press.

Marshall, P.J., Reeb, B.C., & Fox, N.A. (2009). Electrophysiological responses to auditory novelty in temperamentally different 9-month-old infants. *Developmental Science*, *12*(4), 568-582.

Marshall, P.J., & Shipley, T.F. (2009). Event-related potentials to point-light displays of human actions in five-month-old infants. *Developmental Neuropsychology*, *34*(3), 368-377.

Nelson, C.A., & Collins, P.F. (1991). Event-related potential and looking-time analysis of infants' responses to familiar and novel events: Implications for visual recognition memory. *Developmental Psychology*, *27*(1), 50-58.

Nelson, C.A., & Henschel, M., & Collins, P.F. (1993). Neural correlates of cross-modal recognition memory by 8-month-old human infants. *Developmental Psychology*, *29*(3), 411-420.

Nikkel, L., & Karrer, R. (1994). Differential effects of experience on the ERP and behavior of 6-month-old infants: trends during repeated stimulus presentations. *Developmental Neuropsychology*, *10*, 1-11.

Parise, E., Reid, V.M., Stets, M., & Striano, T. (2008). Direct eye-contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience*, *3*(2), 141-150.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson Jr., R., … Taylor, M.J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, *37*, 127-152.

Reynolds, G.D., & Guy, M.W. (in press). Brain-behavior relations in infancy: Integrative approaches to examining infant looking behavior and event-related potentials. *Developmental Neuropsychology*, this issue.

Richards, J.E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: an ERP study. *Developmental Science*, *6*(3), 312-328.

Saby, J.N., & Marshall, P.J. (in press). The utility of EEG band power analysis

in the study of infancy and early childhood. *Developmental Neuropsychology*, this issue.

Slaughter, V., & Suddendorf, T. (2007). Participant loss due to "fussiness" in infant visual paradigms: A review of the last 20 years. *Infant Behavior and Development*, *30*, 505-514.

Snyder, K.A., Garza, J., Zolot, L., & Kresse, A. (2010). Electrophysiological signals of familiarity and recency in the infant brain. *Infancy*, *15*(5), 487-516.

Snyder, K., Webb, S.J., & Nelson, C.A. (2002). Theoretical and methodological implications of variability in infant brain response during a recognition memory paradigm. *Infant Behavior and Development*, *25*, 466-494.

Stahl, D., Parise, E., Hoehl, S., & Striano, T. (2010). Eye contact and emotional face processing in 6-month-old infants: Advanced statistical methods applied to event-related potentials. *Brain and Development*, *32*, 305-317.

Stahl, D., Pickles, A., Elsabbagh, M., Johnson, M.H., & The BASIS Team (in press). Novel machine learning methods for ERP analysis: A validation from research on infants at-risk for autism. *Developmental Neuropsychology*, this issue.

Steichen, T. (1998). sbe19: tests for publication bias in meta-analysis. *Stata Technical Bulletin*, *41*, 9-15.

Sterne, J.A.C. & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046-1055.

Sterne, J.A.C., Egger, M., & Smith, G.D. (2001). Investigating and dealing with publication and other biases in meta-analysis. *BMJ*, *323*, 101-105.

Stets, M., & Reid, V.M. (2011). Infant ERP amplitudes change over the course of an experimental session: Implications for cognitive processes and methodology.

Brain and Development, *33*(7), 558-568.

Thompson, S.G., & Sharp, S.J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, *18*(20), 2693-2708.

Tobias, A. (1999). Assessing the influence of a single study in the meat-analysis estimate. *Stata Technical Bulletin*, *47*, 15-17.

Appendix 1. Papers collected for the present meta-analysis.

[1] Andrews-Espy, K., Molfese, D.L., Molfese, V.J., & Modglin, A. (2004). Development of auditory event-related potentials in young children and relations to word-level reading abilities at age 8 years. *Annals of Dyslexia*, *54*(1), 9-38.

[2] Balas, B.J., Nelson III, C.A., Westerlund, A., Vogel-Farley, V., Riggins, T., & Kuefner, D. (2010). Personal familiarity influences the processing of upright and inverted faces in infants. *Frontiers in Human Neuroscience*, *4*, 1-6.

[3] Bauer, P.J., Wiebe, S.A., Carver, L.J., Waters, J.M., & Nelson, C.A. (2003). Developments in long-term explicit memory late in the first year of life: Behavioral and electrophysiological indices. *Psychological Science*, *14*(6), 629-635.

[4] Birtles, D.B., Braddick, O.J., Wattam-Bell, J., Wilkinson, A.R., & Atkinson, J. (2007). Orientation and motion-specific visual cortex responses in infants born preterm. *NeuroReport*, *18*(18), 1975-1979.

[5] Braddick, O., Birtles, D., Wattam-Bell, J., & Atkinson, J. (2005). Motion- and orientation-specific cortical responses in infancy. *Vision Research*, *45*, 3169-3179.

[6] Brannon, E.M., Libertus, M.E., Meck, W.H., & Woldorff, M.G. (2008). Electrophysiological measures of time processing in infant and adult brains: Weber's law holds. *Journal of Cognitive Neuroscience*, *20*(2), 193-203.

[7] Burden, M.J., Westerlund, A.J., Armony-Sivan, R., Nelson, C.A., Jacobson, S.W., Lozoff, B. et al. (2007). An event-related potential study of attention and recognition memory in infants with iron-deficiency anemia. *Pediatrics*, *120*, 336-345.

[8] Carver, L.J., Bauer, P.J., & Nelson, C.A. (2000). Associations between the infant brain activity and recall memory. *Developmental Science*, *3*(2), 234-246.

[9] Carver, L.J., & Vaccaro, B.G. (2007). 12-month-old infants allocate increased neural resources to stimuli associated with negative adult emotion. *Developmental Psychology*, *43*(1), 54-69.

[10] Clifford, A., Franklin, A., Davies, I.R.L., & Holmes, A. (2009). Electrophysiological markers of categorical perception of color in 7-month-old infants. *Brain and Cognition*, *71*, 165-172.

[11] Conboy, B.T., & Kuhl, P.K. (2010). Impact of second-language experience in infancy: brain measures of first- and second-language speech perception. *Developmental Science*, 1-7.

[12] Courchesne, E., Ganz, L., & Norcia, A.M. (1981). Event-related brain potentials to human faces in infants. *Child Development*, *52*(3), 804-811.

[13] Csibra, G., Davis, G., Spratling, M.W., & Johnson, M.H. (2000). Gamma oscillations and object processing in the infant brain. *Science*, *290*, 1582-1585.

[14] Csibra, G., Tucker, L.A., & Johnson, M.H. (1998). Neural correlates of saccade planning in infants: A high-density ERP-study. *International Journal of Psychophysiology*, *29*, 201-215.

[15] Csibra, G., Tucker, L.A., & Johnson, M.H. (2001). Differential frontal cortex activation before anticipatory and reactive saccades in infants. *Infancy*, *2*(29), 159-174.

[16] Csibra, G., Tucker, L.A., Volein, A., & Johnson, M.H. (2000). Cortical development and saccade planning: the ontogeny of the spike potential. *NeuroReport*, *11*(5), 1069-1073.

[17] de Haan, M., Belsky, J., Reid, V., Volein, A., & Johnson, M.H. (2004). Maternal personality and infants' neural and visual responsivity to facial expressions of emotion. *Journal of Child Psychology and Psychiatry*, *45*(7), 1209-1218.

[18] de Haan, M., Nelson, C.A. (1997). Recognition of the mother's face by six-month-old infants: A neurobehavioral study. *Child Development*, *68*(2), 187-210.

[19] de Haan, M., & Nelson, C.A. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Developmental Psychology*, *35*(4), 1113-1121.

[20] de Haan, M., Pascalis, O., & Johnson, M.H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199-209.

[21] deRegnier, R.A.O., Georgieff, M.K., & Nelson, C.A. (1997). Visual event-related brain potentials in 4-month-old infants at risk for neurodevelopmental impairments. *Developmental Psychobiology*, *30*(1), 11-28.

[22] deRegnier, R.A., Long, J.D., Georgieff, M.K., & Nelson, C.A. (2007). Using event-related potentials to study perinatal nutrition and brain development in infants of diabetic mothers. *Developmental Neuropsychology*, *31*(3), 379-396.

[23] deRegnier, R.A., Nelson, C.A., Thomas, K.M., Wewerka, S., & Georgieff, M.K. (2000). Neurophysiologic evaluation of auditory recognition memory in healthy newborn infants and infants of diabetic mothers. *The Journal of Pediatrics*, *137*(6), 777-784.

[24] deRegnier, R.A., Wewerka, S., Georgieff, M.K., Mattia, F., & Nelson, C.A. (2002). Influences of postconceptual age and postnatal experience on the development of auditory recognition memory in the newborn infant. *Developmental Psychobiology*, *41*, 216-225.

[25] Elsabbagh, M., Volein, A., Csibra, G., Holmboe, F., Garwood, H., Tucker, L. et al. (2009). Neural correlates of eye gaze processing in the broader autism phenotype. *Biological Psychiatry*, *65*, 31-38.

[26] Farroni, T., Csibra, G., Simion, F., & Johnson, M.H. (2002). Eye contact detection in humans from birth. *PNAS*, *99*(14), 9602-9605.

[27] Farroni, T., Johnson, M.H., & Csibra, G. (2004a). Mechanisms of eye gaze perception during infancy. *Journal of Cognitive Neuroscience*, *16*(8), 1320-1326.

[28] Ferguson, M., & Molfese, P.J. (2007). Breast-fed infants process speech differently from bottle-fed infants: Evidence from neuroelectrophysiology. *Developmental Neurophysiology*, *31*(3), 337-347.

[29] Friederici, A.D., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, *17*, 1208-1211.

[30] Friederici, A.D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *NeuroReprt*, *13*(10), 1251-1254.

[31] Friedrich, M., & Friederici, A.D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, *16*(8), 1465-1477.

[32] Friedrich, M., & Friederici, A.D. (2005a). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *NeuroReport*, *16*(6), 653-656.

[33] Friedrich, M., & Friederici, A.D. (2005b). Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, *17*(11), 1785-1802.

[34] Friedrich, M., & Friederici, A.D. (2005c). Semantic sentence processing reflected in the event-related potentials of one- and two-year-old children. *NeuroReport*, *16*(16), 1801-1804.

[35] Friedrich, M., & Friederici, A.D. (2006). Early N400 development and later language acquisition. *Psychophysiology*, *43*(1), 1-12.

[36] Friedrich, M., & Friederici, A.D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *NeuroReport*, *19*(18), 1757-1761.

[37] Friedrich, M., & Friederici, A.D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain and Language*, *114*(2), 66-71.

[38] Friedrich, M., Herold, B., & Friederici, A.D. (2009). ERP correlates of processing native and non-native language word stress in infants with different language outcomes. *Cortex*, *45*, 662-676.

[39] Friedrich, M., Weber, C., & Friederici, A.D. (2004). Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment. *Psychophysiology*, *41*, 772-782.

[40] Goyet, L., de Schonen, S., & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP-study. *Brain Research*, *1332*, 75-89.

[41] Grossmann, T., Johnson, M.H., Farroni, T., & Csibra. G. (2007). Social perception in the infant brain: gamma oscillatory activity in response to eye gaze. *SCAN*, *2*, 284-291.

[42] Grossmann, T., Johnson, M.H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C et al. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society B*.

[43] Grossmann, T., Striano, T., & Friederici, A.D. (2005). Infants' electric brain responses to emotional prosody. *NeuroReport*, *16*(16), 1825-1828.

[44] Grossmann, T., Striano, T., & Friederici, A.D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science*, *9*(3), 309-315.

[45] Grossmann, T., Striano, T., & Friederici, A.D. (2007). Developmental changes in infants' processing of happy and angry facial expressions: A neurobehavioral study. *Brain and Cognition*, *64*, 30-41.

[46] Gunnar, M.R., & Nelson, C.A. (1994). Event-related potentials in year-old infants: Relations with emotionality and cortisol. *Child Development*, *65*(1), 80-94.

[47] Halit, H., Csibra, G., Volein, A., & Johnson, M.H. (2004). Face-sensitive cortical processing in early infancy. *Journal of Child Psycology and Psychiatry*, *45*(7), 1228-1234.

[48] Halit, H., de Haan, M., & Johnson, M.H. (2003). Cortical specialization for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*, *19*, 1180-1193.

[49] Harvison, K.W., Molfese, D.L., Woodruff-Borden, J., & Weigel, R.A. (2009). Neonatal auditory evoked responses are related to perinatal maternal anxiety. *Brain and Cognition*, *71*, 369-374.

[50] He, C., Hotson, L., & Trainor, L.J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*, *19*(5), 878-892.

[51] He, C., Hotson, L., & Trainor, L.J. (2009a). Maturation of cortical mismatch responses to occasional pitch change in early infancy: Effects of presentation rate and magnitude of change. *Neuropsychologia*, *47*, 218-229.

[52] He, C., Hotson, L., & Trainor, L.J. (2009b). Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *European Journal of Neuroscience*, *29*, 861-867.

[53] He, C., & Trainor, L.J. (2009). Finding the pitch of the missing fundamental in infants. *The Journal of Neuroscience*, *29*(24), 7718-7722.

[54] Hill Karrer, J., Karrer, R., Bloom, D., Chaney, L., & Davis, R. (1998). Event-related brain potentials during an extended visual recognition memory task depict delayed development of cerebral inhibitory processes among 6-month-old infants with Down syndrome. *International Journal of Psychophysiology*, *29*, 167-200.

[55] Hirai, M., & Hiraki, K. (2005). An event-related potentials study of biological motion perception in human infants. *Cognitive Brain Research*, *22*, 301-304.

[56] Hirotani, M., Stets, M., Striano, T., & Friederici, A.D. (2009). Joint attention helps infants learn new words: Event-realted potential evidence. *NeuroReport*, *20*(6), 600-605.

[57] Hoehl, S., Palumbo, L., Heinisch, C., & Striano, T. (2008). Infants' attention is biased by emotional expressions and eye gaze direction. *NeuroReport*, *19*(5), 579-582.

[58] Hoehl, S., Reid, V., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye-gaze. *Developmental Science*, *11*(1), 10-16.

[59] Hoehl, S., & Striano, T. (2008). Neural processing of eye gaze and threat-related emotional facial expressions in infancy. *Child Development*, *79*(6), 1752-1760.

[60] Hoehl, S., & Striano, T. (2009). The development of emotional face and eye gaze processing. *Developmental Science*, 1-13.

[61] Hoehl, S., & Striano, T. (2010). Infants' neural processing of positive emotion and eye gaze. *Social Neuroscience*, *5*(1), 30-39.

[62] Hoehl, S., Wiese, L., & Striano, T. (2008). Yound infants' neural processing of objects is affected by eye gaze direction and emotional expression. *PLoS*, *3*(6), 1-6.

[63] Hofmann, M.J., & Salapatek, P. (1981). Yound infants' event-related potentials (ERPs) to familiar and unfamiliar visual and auditory events in a recognition memory task. *Electroencephalography and Clinical Neurophysiology*, *52*, 405-417.

[64] Hyde, D.C., Jones, B.L., Porter, C.L., & Flom, R. (2010). Visual stimulation enhances auditory processing in 3-month-old infants and adults. *Developmental Psychobiology*, *52*(2), 181-189.

[65] Hyde, D.C., & Spelke, E.S. (2010). Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*(2), 360-371.

[66] Jeschonek, S., Marinovic, V., Hoehl, S., Elsner, B., & Pauen, S. (2010). Do animals and furniture items elicit different brain responses in human infants?. *Brain and Development*, *32*(10), 863-871.

[67] Kaufman, J., Csibra, G., & Johnson, M.H. (2003). Representing occluded objects in the human infant brain. *Proceedings of the Royal Society London B (Supplement)*.

[68] Kaufman, J., Csibra, G., & Johnson, M.H. (2005). Oscillatory activity in the infant brain reflects objects maintenance. *PNAS*, *102*(42), 15271-15274.

[69] Key, A.P.F., Ferguson, M., Molfese, D.L., Peach, K., Lehman, C., & Molfese, V.J. (2007). Smoking during pregnancy affects speech-processing ability in newborn infants. *Environmental Health Perspectives*, *115*(4), 623-629.

[70] Kobiella, A., Grossmann, T., Reid, V.M., & Striano, T. (2008). The discrimination of angry and fearful facial expressions in 7-month-old infants: An event-related potential study. *Cognition and Emotion*, *22*(1), 134-146.

[71] Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, *24*, 109-116.

[72] Kotsoni, E., Mareschal, D., Csibra, G., & Johnson, M.H. (2006). Common-onset visual masking in infancy: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *18*(6), 966-973.

[73] Kuhl, P.K., Conboy, B.T., Coffey-Corina, S. Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society*, *363*, 979-1000.

[74] Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *PNAS*, *105*(32), 11442-11445.

[75] Leppänen, J.M., Moulson, M.C., Vogel-Farley, V., & Nelson, C.A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child Development*, *78*(1), 232-245.

[76] Leppänen, J.M., Richmond, J., Vogel-Farley, V.M., Moulson, M.C., & Nelson, C.A. (2009). Categorical representation of facial expressions in the infant brain. *Infancy*, *14*(3), 346-362.

[77] Lukowski, A.F., Wiebe, S.A., Haight, J.C., deBoer, T., Nelson, C.A., & Bauer, P.J. (2005). Forming a stable memory representation in the first year of life: why imitation is more than child's play. *Developmental Science*, *8*(3), 279-298.

[78] Macchi Cassia, V., Kuefner, D., Westerlund, A., & Nelson, C.A. (2006). A behavioral and ERP investigation of 3-month-olds face preferences. *Neuropsychologia*, *44*, 2113-2125.

[79] Männel, C., & Friederici, A.D. (2009). Pauses and intonational phrases: ERP studies in 5-month-old German infants and adults. *Journal of Cognitive Neuroscience*, *21*(10), 1988-2006.

[80] Marshall, P.J, Reeb, B.C., & Fox, N.A. (2009). Electrophysiological responses to auditory novelty in temperamentally different 9-month-old infants. *Developmental Science*, *12*(4), 568-582.

[81] Marshall, P.J., & Shipley, T.F. (2009). Event-related potentials to point-light displays of human actions in five-month-old infants. *Developmental Neuropsychology*, *34*(3), 368-377.

[82] Mason, A.J.S., Braddick, O.J., Wattam-Bell, J., 6 Atkinson, J. (2001). Directional motion asymmetry in infants VEPs – which direction?. *Vision Research*, *41*, 201-211.

[83] McIsaac, H., & Polich, J. (1992). Comparison of infant and adult P300 from auditory stimuli. *Journal of Experimental Child Psychology*, *53*, 115-128.

[84] Mills, D.L., Coffey-Corina, S., & Neville, H.J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*, *13*(3), 397-445.

[85] Mills, D.L, Pratt, C., Zangl, R., Stager, R.S., Neville, H.J., & Werker, J.F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience*, *16*(8), 1452-1464.

[86] Molfese, D.L. (1990). Auditory evoked responses recorded from 16-month-old human infants to words they did and did not know. *Brain and Language*, *38*, 345-363.

[87] Molfese, D.L., & Molfese, V.J. (1979). Hemisphere and stimulus differences as reflected in the cortical responses of newborn infants to speech stimuli. *Developmental Psychology*, *15*(5), 505-511.

[88] Molfese, D.L., & Molfese, V.J. (1980). Cortical responses of preterm infants to phonetic and nonphonetic speech stimuli. *Developmental Psychology*, *16*(6), 574-581.

[89] Molfese, D.L., & Molefese, V.J. (1985). Electrophysiological indices of auditory discrimination in newborn infants: The bases for predicting later language development?. *Infant Behavior and Development*, *8*, 197-211.

[90] Molfese, D.L., Morse, P.A., & Peters, C.J. (1990). Auditory evoked responses to names for different objects: Cross-modal processing as a basis for infant language acquisition. *Developmental Psychology*, *26*(5), 780-795.

[91] Nelson, C.A., & Collins, P.F. (1991). Event-related potential and looking-time analysis of infants' responses to familiar and novel events: Implications for visual recognition memory. *Developmental Psychology*, *27*(1), 50-58.

[92] Nelson, C.A., & Collins, P.F. (1992). Neural and behavioral correlates of visual recognition memory in 4- and 8-month-old infants. *Brain and Cognition*, *19*, 105-121.

[93] Nelson, C.A., & de Haan, M. (1996). Neural correlates of infants' visual responsiveness to facial expressions of emotion. *Developmental Psychobiology*, *29*(7), 577-595.

[94] Nelson, C.A., Henschel, M., & Collins. P.F. (1993). Neural correlates of cross-modal recognition memory by 8-month-old human infants. *Developmental Psychology*, *29*(3), 411-420.

[95] Nelson, C.A., & Salapatek, P. (1986). Electrophysiological correlates of infant recognition memory. *Child Development*, *57*(6), 1483-1497.

[96] Nelson, C.A., Thomas, K.A., de Haan, M., & Wewerka, S.S. (1998). Delayed recognition memory in infants and adults as revealed by event-related potentials. *International Journal of Psychobiology*, *29*, 145-165.

[97] Nelson, C.A., Wewerka, S.S., Borscheid, A.J., deRegnier, R.A., & Georgieff, M.K. (2003). Electrophysiologic evidence of impaired cross-modal recognition memory in 8-month-old infants of diabetic mothers. *The Journal of Pediatrics*, *142*(3), 575-582.

[98] Nelson, C.A., Wewerka, S., Thomas, K.M., Tribby-Walbridge, S., deRegnier, R.A., & Georgieff, M. (2000). Neurocognitive sequelae of infants of diabetic mothers. *Behavioral Neuroscience*, *114*(5), 950-956.

[99] Nyström, P. (2008). The infant mirror neuron system studied by high density EEG. *Journal of Social Neuroscience*, *3*, 334-347.

[100] Pang, E.W., Edmonds, G.E., Desjardins, R., Khan, S.C., Trainor, L.J., & Taylor, M.J. (1998). Mismatch negativity to speech stimuli in 8-month-old infants and adults. *International Journal of Psychophysiology*, *29*, 227-236.

[101] Pannekamp, A., Weber, C., & Friederici, A.D. (2006). Prosodic processing at the sentence level in infants. *NeuroReport*, *17*(6), 675-678.

[102] Parise, E., Handl, A., & Striano, T. (2010). Processing faces in dyadic and triadic contexts. *Neuropsychologia*, *48*, 518-528.

[103] Parise, E., Reid, V.M., Stets, M., & Striano, T. (2008). Direct eye-contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience*, *3*(2), 141-150.

[104] Parker, S.W., Nelson, C.A., & The Bucharest Early Intervention Project Core Group (2005a). An event-related potential study of the impact of institutional rearing on face recognition. *Development and Psychopathology*, *17*, 621-639.

[105] Parker, S.W., Nelson, C.A., & The Bucharest Early Intervention Project Core Group (2005b). The impact of early institutional rearing on the ability to discriminate facial expressions of emotion: An event-related potential study. *Child Development*, *76*(1), 54-72.

[106] Pascalis, O., de Haan, M., Nelson, C.A., & de Schonen, S. (1998). Long-term recognition memory for faces assessed by visual paired comparison in 3- and 6-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 249-260.

[107] Purhonen, M., Kilpeläinen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2004). Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *International Journal of Psychophysiology*, *52*, 257-266.

[108] Purhonen, M., Kilpeläinen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2005). Four-month-old infants process own mother's voice faster than unfamiliar voices – Electrical signs of sensitization in infant brain. *Cognitive Brain Research*, *24*, 627-633.

[109] Quinn, P.C., Westerlund, A., & Nelson, C.A. (2006). Neural markers of categorization in 6-month-old infants. *Psychological Science*, *17*, 59-66.

[110] Reid, V.M., Hoehl, S., Grigutsch, M., Groendahl, A., Parise, E., & Striano, T. (2009). The neural correlates of infant and adult goal prediction: evidence for semantic processing systems. *Developmental Psychology*, *45*(3), 620-629.

[111] Reid, V.M., Hoehl, S., Landt, J., & Striano, T. (2008). Human infants dissociate structural and dynamic information in biological motion: Evidence from neural systems. *SCAN*, *3*, 1-7.

[112] Reid, V.M., Hoehl, S., & Striano, T. (2006). The perception of biological motion by infants: An event-related potential study. *Neuroscience Letters*, *395*, 211-214.

[113] Reid, V.M., Striano, T., Kaufman, J., & Johnson, M.H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *NeuroReport*, *15*(16), 2553-2555.

[114] Richards, J.E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: an ERP study. *Developmental Science*, *6*(3), 312-328.

[115] Richards, J.E. (2005). Localizing cortical sources of event-related potentials in infants' covert orienting. *Developmental Science*, *8*(3), 255-278.

[116] Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., & Kuhl, P.K. (2005). Neural patterns to speech and vocabulary growth in American infants. *NeuroReport*, *16*(5), 495-498.

[117] Rivera-Gaxiola, M., Silva-Pereyra, J., Klarman, L., Garcia-Sierra, A., Lara-Ayala, L., Cadena-Salazar, C. et al. (2007). Principal component analyses and scalp distribution of the auditory P150-250 and N250-550 to speech contrasts in Mexican and American infants. *Developmental Neuropsychology*, *31*(3), 363-378.

[118] Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P.K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, *8*(2), 162-172.

[119] Schulman-Galambos, C., & Galambos, R. (1978). Cortical responses from adults and infants to complex visual stimuli. *Electroencephalography and Clinical Neurophysiology*, *45*, 425-435.

[120] Scott, L.S., & Monesson, A. (2010). Experience-dependent neural specialization during infancy. *Neuropsychologia*, *48*(6), 1857-1861.

[121] Scott, L.S., & Nelson, C.A. (2006). Featural and configural face processing in adults and infants: A behavioral and electrophysiological investigation. *Perception*, *35*, 1107-1128.

[122] Scott, L.S., Shannon, R.W., & Nelson, C.A. (2006). Neural correlates of human and monkey face processing in 9-month-old infants. *Infancy*, *10*(2), 171-186.

[123] Senju, A., Johnson, M.H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Social Neuroscience*, *1*(3-4), 220-234.

[124] Shafer, V.L., Shucard, D.W., & Jaeger, J.J. (1999). Electrophysiological indices of cerebral specialization and the role of prosody in language acquisition in 3-month-old infants. *Developmental Neuropsychology*, *15*(1), 73-109.

[125] Shirai, N., Birtles, D., Wattam-Bell, J., Yamaguchi, M.K., Kanazawa, S., Atkinson, J. et al. (2009). Asymmetrical cortical processing of radial expansion/contraction in infants and adults. *Developmental Science*, *12*(6), 946-955.

[126] Siddappa, A.M., Georgieff, M.K., Wewerka, S., Worwa, C., Nelson, C.A., & deRegnier, R.A. (2004). Iron deficiency alters auditory recognition memory in newborn infants of diabetic mothers. *Pediatric Research*, *55*(6), 1034-1041.

[127] Simos, P.G., & Molfese, D.L. (1997). Electrophysiological responses from a temporal order continuum in the newborn infant. *Neuropsychologia*, *35*(1), 89-98.

[128] Snyder, K.A. (2010). Neural correlates of encoding predict infant's memory in the paired-comparison procedure. *Infancy*, *15*(3), 270-299.

[129] Snyder, K.A., Garza, J., Zolot, L., & Kresse, A. (2010). Electrophysiological signals of familiarity and recency in the infant brain. *Infancy*, *15*(5), 487-516.

[130] Snyder, K.A., & Keil, A. (2008). Repetition suppression of induced gamma activity predicts enhanced orienting toward a novel stimulus in 6-month-old infants. *Journal of Cognitive Neuroscience*, *20*(12), 2137-2152.

[131] Snyder, K.A., Webb, S.J., & Nelson, C.A. (2002). Theoretical and methodological implications of variability in infant brain response during a recognition memory paradigm. *Infant Behavior and Development*, *25*, 466-494.

[132] Stahl, D., Parise, E., Hoehl, S., & Striano, T. (2010). Eye contact and emotional face processing in 6-month-old infants: Advanced statistical methods applied to event-related potentials. *Brain and Development*, *32*, 305-317.

[133] Stolarova, M., Whitney, H., Webb, S.J., deRegnier, R.A., Georgieff, M.K., & Nelson, C.A. (2003). Electrophysiological brain responses of six-month-old low risk premature infants. *Infancy*, *4*(3), 437-450.

[134] Striano, T., Kopp, F., Grossmann, T., & Reid, V.M. (2006). Eye contact influences neural processing of emotional expressions in 4-month-old infants. *Social Cognitive and Affective Neuroscience*, *1*(2), 87-94.

[135] Striano, T., Reid, V.M., & Hoehl, S. (2006). Neural mechanisms of joint attention in infancy. *European Journal of Neuroscience*, *23*, 2819-2823.

[136] Swingler, M.M., Sweet, M.A., & Carver, L.J. (2007). Relations between mother-child interactions and the neural correlates of face processing in 6-month-olds. *Infancy*, *11*(1), 63-86.

[137] Tew, S., Fujioka, T., He, C., & Trainor, L. (2009). Neural representations of transposed melody in infants at 6 months of age. *The Neurosciences and Music III: Disorders and Plasticity: Ann. N. Y. Acad. Sci.*, *1169*, 287-290.

[138] Thierry, G., Vihman, M., & Roberts, M. (2003). Familiar words capture the attention of 11-month-olds in less than 250ms. *NeuroReport*, *14*(18), 2307-2310.

[139] Trainor, L., McFadden, M., Hogdson, L., Darragh, L., Barow, J., Matsos, L. et al. (2003). Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *International Journal of Psychophysiology*, *51*, 5-15.

[140] Trainor, L.J., Samuel, S.S., Desjardins, R.N., & Soonadara, R.R. (2001). Measuring temporal resolution in infants using mismatch negativity. *NeuroReport*, *12*(11), 2443-2448.

[141] van Elk, M., van Schie, H.T., Hunnius, S., Vesper, S., & Bekkering, H. (2008). You'll never crawl alone: Neurophysiological evidence for experience-dependent motor resonance in infancy. *NeuroImage*, *43*, 808-814.

[142] Vihman, M.M., Thierry, G., Lum, J., Keren-Portnoy, T., & Martin, P. (2007). Onset of word form recognition in English, Welsh, and English-Welsh bilingual infants. *Applied Psycholinguitics*, *28*, 475-493.

[143] von Koss Torkildsen, J., Friis, Hansen, H., Svangstu, J.M., Smith, L., Simonsen, H.G., Moen, I. et al. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language*, *108*, 73-88.

[144] von Koss Torkildsen, J., Sannerud, T., Syversen, G., Thormodsen, R., Simonsen, H.G., Moen, I. et al. (2006). Semantic organization of basic-level words in 20-month-olds: An ERP study. *Journal of Neurolinguistics*, *19*, 431-454.

[145] von Koss Torkildsen, J., Syversen, G., Simonsen, H.G., Moen, I., & Lindgren, M. (2007). Brain responses to lexical-priming in children at-risk for dyslexia. *Brain and Language*, *102*, 243-261.

[146] Webb, S.J., Long, J.D., & Nelson, C.A. (2005). A longitudinal investigation of visual event-related potentials in the first year of life. *Developmental Science*, *8*(6), 605-616.

[147] Webb, S.J., & Nelson, C.A. (2001). Perceptual priming for upright an inverted faces in infants and adults. *Journal of Experimental Child Psychology*, *79*, 1-22.

[148] Weber, C., Hahne, A., Friedrich, M., & Friederici, A.D. (2004). Discrimination of word stress in early infant perception: electrophysiological evidence. *Cognitive Brain Research*, *18*, 149-161.

[149] Weber, C., Hahne, A., Friedrich, M., & Friederici, A.D. (2005). Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment: Neurophysiological evidence. *Cognitive Brain Research*, *25*, 180-187.

Appendix 2. List of factors included in the present meta-analysis.

[1] the age-range used

[2] the mean-age of the participants in days

[3] the Standard Deviation of the mean-age of the participants in days

[4] the total number of infants tested for a specific experimental group

[5] the total number of participants who entered the final analysis for a specific experimental group

[6] the total number of participants who did not enter the final analysis for a specific experimental group

[7] the total number of participants who did not enter the final analysis for a specific experimental group in percent

[8] the number of participants who did not enter the final analysis for a specific experimental group due to “fussiness or inattentiveness”

[9] the number of participants who did not enter the final analysis for a specific experimental group due to “fussiness or inattentiveness” in percent

[10] the number of participants who did not enter the final analysis for a specific experimental group due to “artifacts or an insufficient number of trials in general”

[11] the number of participants who did not enter the final analysis for a specific experimental group due to “artifacts or an insufficient number of trials in general” in percent

[12] the number of participants who did not enter the final analysis for a specific experimental group due to “technical problems”

[13] the number of participants who did not enter the final analysis for a specific experimental group due to “technical problems” in percent

[14] the number of participants who did not enter the final analysis for a specific experimental group due to “experimental error”

[15] the number of participants who did not enter the final analysis for a specific experimental group due to “experimental error” in percent

[16] the number of participants who did not enter the final analysis for a specific experimental group due to “other” (e.g., medical conditions which had been revealed only after the test session)

[17] the number of participants who did not enter the final analysis for a specific experimental group due to “other” in percent

[18] the nature of the experimental group ((a) typically developing infants, (b) typically developing infants with some kind of special condition, and (c) clinical groups)

[19] the number of conditions presented

[20] the number of different stimuli used for one experimental condition

[21] the number of potential presentations per condition

[22] the total number of stimuli which could potentially be presented

[23] the minimal number of artifact-free trials an infant needed to contribute to the average per condition in order to be able to enter the final analysis

[24] the (mean-) number of artifact-free trials included in the averaging process per condition from each of the participants who entered the final analysis

[25] the Standard Deviation of the (mean-) number of artifact-free trials included in the averaging process per condition from each of the participants who entered the final analysis

[26] the (mean-) duration of a single stimulus in ms

[27] the (mean-) duration of the attention getter(s) where applicable in ms

[28] the (mean-) duration of the interstimulus interval in ms

[29] the (mean-) duration of a single trial in ms

[30] the (mean-) duration of the stimulus presentation when all of the stimuli that had been prepared would be presented to the infant without any breaks in s and min (in case it was mentioned in the article that breaks had been built into the procedure, then we included the duration of these breaks into our total duration of the procedure)

[31] the nature of the stimuli ((a) visual, (b) auditory, (c) visual and auditory, (d) visual and animate, (e) visual and animate and interactive, (f) visual and auditory and animate, and (g) visual and auditory and animate and interactive)

[32] the age of the article in years

[33] the type of EEG-system that had been used for data-collection ((a) cap, (b) net, (c) individually applied electrodes, and (d) not specified)

[34] the number of electrodes used for data-collection

As seen in Study 2, the number of factors that appear to influence the size of a study's attrition rate are surprisingly low. Factors such as the duration of a stimulus or trial or of the entire test-session, which would likely be expected to increase or decrease attrition, did not show to have an impact. Moreover, a study-feature that is frequently recommended to be kept at a low level (e.g., DeBoer, Scott, & Nelson, 2007; Picton et al., 2000) – the number of experimental conditions presented – did not impact on the studies' attrition rates either. Therefore, the theoretical basis for the assumption that fewer experimental conditions will help infants to remain attentive and focussed during a test-session and, consequently, will help to keep attrition at an acceptable level, does not appear to have any empirical support. Via the exploration of possibilities surrounding study-design and with the aim to increase the volume of data to be collected from an individual infant, Study 3 has been designed to illustrate a new approach to data-acquisition. In this study, a group of 12-month-olds were presented with eight experimental conditions. In order to ensure that the quality of the study's results can be evaluated and compared to previous findings, the included conditions originated from three earlier research articles that were not theoretically related to each other. However, the main focus of this study is on the methodological insights that the application of such a study-design on an infant population can provide.

References

DeBoer, T., Scott, L.S., & Nelson, C.A. (2007). Methods for acquiring and analysing infant event-related potentials. In M. de Haan (ed.), *Infant EEG and Event-Related Potentials* (pp. 5-38), Hove, England: Psychology Press.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, Jr., R., … Taylor, M.J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*, 127-152.

Chapter 4

Infants Need More Variety – Increased Data Acquisition with Reduced Participant Attrition in Infant ERP Studies

Manuela Stets, Mike Burt, and Vincent M. Reid

(under review).

Abstract

Infant ERP-studies often feature high attrition rates with many trials excluded. The number of experimental conditions is conventionally limited to obtain reasonable trial-numbers. We presented an ERP-study involving eight conditions, based on three previous studies, to 18 1-year-olds. We expected to replicate original results at least partly. We were interested in this novel stimulus-presentation-method's effect on infant attention. Increased trial-numbers were obtained and infant attention was sustained. Attrition rate was decreased. One study was replicated fully, others were replicated in part. Varied stimuli extend infant attention allowing the collection of more data. However, sub-experiments must be cognitively and perceptually distinct.

Introduction

According to general agreement among developmental psychologists working with infant populations, infants' attention span is very limited and hard to sustain over an extended period of time (e.g., DeBoer, Scott, & Nelson, 2007; Thierry, 2005). Furthermore, it is frequently stated that it is easy to overload an infant's cognitive system with too much information or with task demands that are too complex for a specific age group (Parise, Reid, Stets, & Striano, 2008; Nelson & Collins, 1991). Additionally, there seems to be common agreement that substantial attrition rates are to be expected when studies are conducted with infant populations (e.g., DeBoer, Scott, & Nelson, 2007; Picton & Taylor, 2007). Therefore, a common practice – particularly among researchers using ERP-methodologies with infants – is to limit the number of experimental conditions and the duration of the individual stimuli that are presented to the participants. Moreover, these strategies are not only frequently used but also frequently recommended in methodological literature (Picton et al., 2000; DeBoer, Scott, & Nelson, 2007; Hoehl & Wahl, 2012).

One of the most frequent contributors to an infant ERP-study's attrition rate, as stated in published articles, is fussiness. However, as of yet, no comprehensive definition has been provided for this concept. Most likely, fussiness is involved in a marked decline in attention toward the stimuli presented. If an infant displays negative affect prior to observing a minimum number of trials, the study is terminated and the infant is marked as "fussy" for the purposes of the study. However, in their recent meta-analysis of 149 infant ERP-articles, Stets, Stahl, and Reid (2012) could not find any correlation between study-features such as the number of conditions or the stimulus duration and the attrition rate that was attributed to fussiness. This was true

irrespective of the age group tested. The only obvious impact on attrition rate in the examined infant ERP-studies was the nature of the stimuli (purely visual, purely auditory, or auditory and visual combined). The age of the published article itself was also a predictor of attrition rate, as was the number of participants that had been tested in total for a particular study or experimental group. Infants participating in a study using purely visual stimulation were twice as likely to have to be excluded from the study's final analysis than those participating in a study using purely auditory stimulation. Also, the attrition rate tended to be lower if a study had been published recently and was higher if the total number of infants tested was high (see Stets, Stahl, & Reid, 2012).

In an attempt to investigate the impact of study design on infant attention, we created an ERP-study with eight experimental conditions. Currently, the most frequently used strategy is to present infants with stimuli from two, typically very similar, experimental conditions which only differ in the experimental manipulation. However, as can be seen in Stets and Reid (2011), this repetitive presentation of very similar stimuli can eventually result in changes in attention-allocation toward the stimuli, which is likely due to habituation, thereby producing unforeseen changes in the ERP-measures for each condition. We hypothesized that a more variable and, consequently, more engaging stimulus-presentation would help to sustain infant attention over the course of a test-session. We consequently presented a group of 12- to 13-month-old infants with stimuli originating from three theoretically distinct research articles (Csibra, Tucker, Volein, & Johnson, 2000; Halit, de Haan, & Johnson, 2003; and Hoehl, Reid, Mooney, & Striano, 2008). Csibra et al. (2000) and Halit et al. (2003) had conducted their studies with groups of 12-month-old infants and contributed six of our eight conditions. Csibra and colleagues (2000) had investigated

the occurrence of a spike-potential preceding a saccade-onset (i.e., a purely perceptual phenomenon) in three groups of participants. Halit, de Haan, and Johnson (2003) were interested in face-perception and -processing when infants were presented with human and monkey faces in different orientations and without any social components. As these studies had been conducted with infants of approximately the same age-range as our study, we aimed to replicate the results of the original articles at least in part. Finally, as only two studies using visual stimuli have been published with this age group, we included the stimuli originating from Hoehl et al. (2008) even though it had not been used with this age group before. These two conditions were aimed at the perception and processing of social information as they presented the infants with faces looking either towards or away from a toy. Therefore the focus in this study lay less on the perception of the faces as such and more on the processing of the social information conveyed by the person's eye-gaze direction. Again, as such stimuli had not been tested with this age group before, we were cautious to expect a replication of the results of the original study.

Materials

Methods – general

The stimuli depicting the eight conditions used in the present study originated from three theoretically unrelated research articles. Therefore, we decided to treat them as separate studies. In the following, we will first describe the general methods that were common to all of the conditions and then explain the specifics of the separate studies in more detail.

Participants

We tested a total of 18 typically developing infants aged between 367 and 396 days (M = 381.61 days, SD = 9.66 days). All of the participants had been born full-term (between 37 and 42 weeks of gestation) and with a normal birth-weight (< 2500 g). Twelve of the infants were male, six were female. Written consent for study-participation had been provided by the parents or caregivers accompanying the infants prior to the test-session.

Procedure

For the entire duration of the test-session, the infants were seated on their caregiver's lap in a dimly lit, quiet testing-booth. The viewing distance to the 17-inch screen presenting the stimuli was 70 cm. The eight conditions were presented in two blocks of eight stimuli with each block containing one stimulus for each of the conditions. The only difference between these two blocks was the order of the conditions. An uninterrupted run through the entire stimulus presentation lasted for approximately 11 minutes with each condition being presented 34 times (total number

of stimuli = 272). The duration of the inter-stimulus interval was 1000 ms in all cases. When an infant showed obvious signs of fussiness or boredom he or she was given a short break, and the test-session was discontinued when the infant's attention to the screen could not be recaptured. To control for the infants' attention to the screen, their behaviour was video-recorded onto the hard-drive of a DVD-recorder for later off-line coding.

EEG recording and analysis

The EEGs were recorded continuously from 32 scalp locations at a sampling rate of 1000 Hz using Ag-AgCl ring-electrodes, arranged according to the 10-20 system, and a conductive gel. During data-acquisition, the recordings were referenced to FCz. The data were amplified with a NeuroScan SynAmps amplifier, and vertical and horizontal electro-oculargrams were recorded bipolarly to control for artefacts resulting from eye-movements. The data were filtered with a band-pass filter with the upper edge set at 0.3 Hz and the lower edge set at 30 Hz. Baseline-correction was conducted using a baseline of 100 ms.

Study 1 – the Spike-Potential (Csibra, Tucker, Volein, & Johnson, 2000)

Participants

14 of 18 participants entered the final analysis. Four infants had to be excluded from the analysis because they did not reach the minimal criterion of providing at least 10 artefact-free trials to the averaging process due to fussiness or inattentiveness (n = 3) or because the averaged data contained frequency-artefacts (n = 1)[1]. (We define "fussiness" as the state when the infant shows signs of unhappiness with the test-situation such as a negative facial expression, agitated movements, vocalizations, or trying to remove single electrodes or the entire cap.) The overall attrition rate for this analysis was 22.22%. The mean-number of artefact-free trials contributed by the infants was 24.21 (range: 16 to 39). These numbers refer to both of the conditions that contributed data to this analysis in conjunction as the statistical analysis was collapsed for the conditions. This was also the case in the original study (G. Csibra, personal communication, September 25, 2012).

Stimuli

The stimuli were created using Adobe Photoshop based on the descriptions provided in Csibra et al. (2000). One of three different objects (a red rubber-boat, a wooden rattle, or a yellow rubber-duck) appeared in an upright orientation in the centre of the screen set on a gray background. For a total duration of 990 ms, these objects

[1] The data from one participant showed frequency noise particularly at the location of Pz. As opposed to the other analyses in which data are presented as averages derived from at least two channels, this particular analysis is based on Pz only. Therefore, the infant had to be excluded as the noise could not be averaged out of the signal by way of averaging over different channels.

were spun clockwise or anti-clockwise in the centre of the screen while concurrently decreasing in size. In order to create a smooth spinning movement, we presented nine different static images of the same object, each turned by an angle of 90° and slightly reduced in size for 110 ms.

For the gap trials, the last image depicting the spinning object in an upright position and at its smallest size (approx. 2 x 2 cm) was presented for 110 ms and then disappeared leaving the screen blank except for the grey background colour. After 200 ms a black-and-white checkerboard of approximately 5 x 5 cm edge-length appeared on either the left or the right side of the screen for 800 ms with the centre of the screen being empty (see Figure 1a). In the trials without the gap, the last image depicting the spinning object in an upright position and at its smallest size was presented for 110 ms to account for the last 110 ms of the 990 ms-spinning-motion and then for another 200 ms in order to bridge the gap. Again, after these 200 ms a checkerboard appeared to one side of the screen while the previously spinning object was still present in the centre of the screen. This display was presented for 800 ms (see Figure 1b). The spinning-direction and the side on which the checkerboard appeared were counterbalanced across trials and conditions.

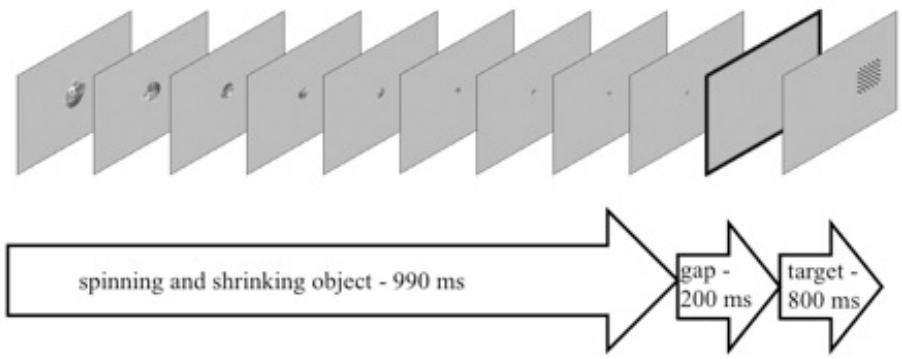

Figure 1a. An example of the setup of a trial depicting the experimental condition with a gap (presented for 200 ms) between the spinning and shrinking objects (presented for 990 ms, 9 individual images presented for 110 ms each) and the target checkerboard (presented for 800 ms).

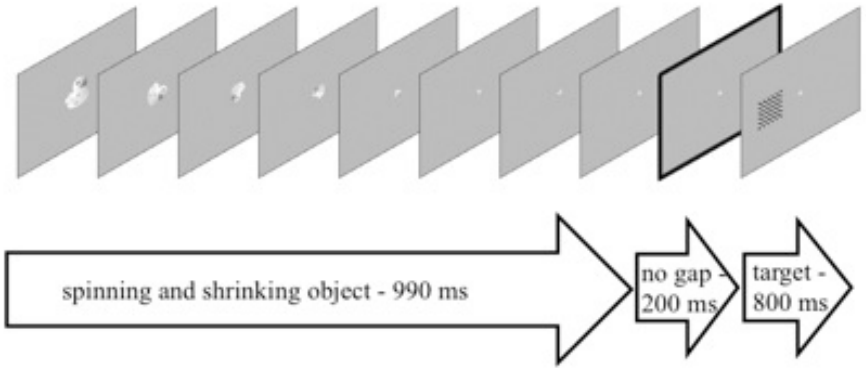

Figure 1b. An example of the setup of a trial depicting the experimental condition without a gap (presented for 200 ms) between the spinning and shrinking objects (presented for 990 ms, 9 individual images presented for 110 ms each) and the target-checkerboard (presented for 800 ms).

Data analysis

Following the methods outlined in Csibra et al. (2000), we re-referenced the data to the average-value of all the channels used. In order to identify the trials that would be usable for the statistical analysis, the video-recordings of each individual

trial of the gap- or no-gap-conditions were manually inspected frame by frame for participants' potential saccades towards the checkerboard. The first frame counted was the one when either the gap or the bridged period started. Each frame had a duration of 40 ms. Therefore, if a shift in a participant's eye-gaze direction was detected in frame 8, for instance, the onset of the saccade was noted as having a latency of 320 ms. According to this approximate latency-value of the saccade-onset, we identified the positive peak closest to this latency in each individual trial. Then, we took note of the exact latency of the respective saccade-onset and of the amplitude at this latency and, as the sampling rate in the original article had been set at 500 Hz rather than at the 1000 Hz that we used, at plus and minus 2 ms around this latency. The same procedure was followed for the amplitudes at 10 and 18 ms prior to the established saccade-onset. Therefore, we had three amplitude-values for each of the three latency-periods (see Figure 2 for an illustration of the sampling-method). Finally, we calculated the average amplitude-values for these three time-points (i.e., saccade-onset, 10 ms and 18 ms before saccade-onset). Following Csibra et al. (2000), this analysis was only conducted for one channel, namely Pz. We discarded any trials in which the participants either did not look at the screen at the onset of the gap- or the bridged-period or when they had seen it but did not shift their eye-gaze toward the checkerboard within 1000 ms or at all. Moreover, trials which were contaminated by movement- or other artefacts and trials when the participants shifted their eye-gaze direction to a point off-screen within the 1000 ms were excluded from the average as well.

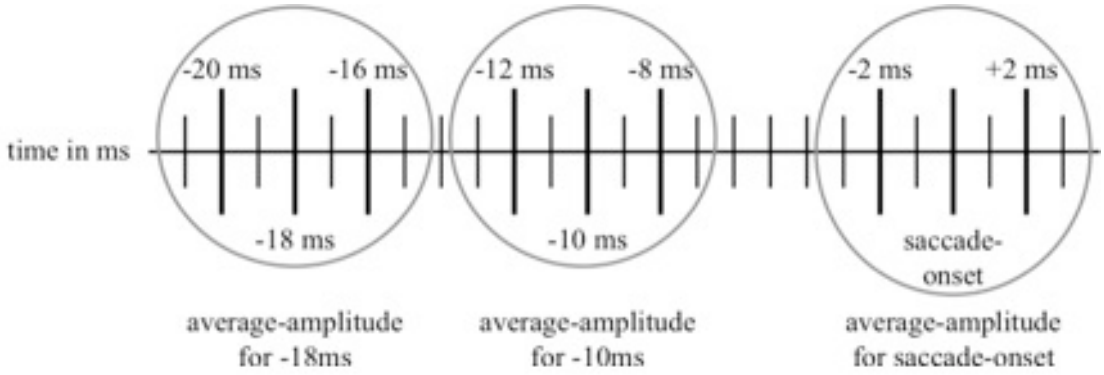

Figure 2. An illustration of our sampling-method used for collecting the data for the amplitude-comparison between saccade-onset and the time-points of -10 and -18 ms before saccade-onset. The highlighted vertical lines stand for the time-points from which the amplitudes have been noted. Afterwards, the amplitudes measured at the encircled lines were used to calculate the mean-amplitudes for the respective time-points.

In the analysis presented in Csibra et al. (2000), no differences were found between the conditions; i.e., both Gap- and No-Gap-trials elicited spike-potentials to the same extents. The only difference that was found between the participants' responses were the latencies at which the spike-potentials and the respective saccades occurred. In the No-Gap-trials, the saccades and the preceding spike-potentials frequently occurred slightly later than those in the Gap-trials which was attributed to the so-called "sticky eyes"-phenomenon. Based on the lack of differences between the condition reported in the original paper and as the majority of our participants contributed more artefact-free trials for the Gap-condition, we decided to collapse trials across conditions which had also been done in later analyses in the original article (G. Csibra, personal communication, September 25, 2012). Consequently, we conducted two paired-samples t-tests: one comparing the average-amplitudes obtained from the latencies at saccade-onsets to those from the time-point 10 ms prior to the

respective saccade-onset and another comparing the average-amplitudes obtained from the latencies at saccade-onsets to those from the time-point 18 ms prior to the respective saccade-onset. However, as mentioned above, the number of artefact-free trials contributed for the Gap-condition tended to be higher than the number of trials obtained for the condition without the gap. Therefore, we also performed post-hoc analyses again using a paired-samples t-test including only data from those infants who had contributed minimally 10 artefact-free trials for both conditions to averaging (n = 5).

Study 2 – the N290 and the P400 (Halit, de Haan, & Johnson, 2003)

Participants

15 of 18 tested participants entered the final analysis. Three infants had to be excluded from the analysis because they did not reach the minimal criterion of providing at least 10 artefact-free trials to the averaging process due to fussiness or inattentiveness (n = 3). Therefore, the overall attrition rate for this analysis was 16.67%. The mean-number of artefact-free trials contributed by the infants across conditions was 13.89 (range: 10 to 25). For individual conditions the mean-numbers of trials contributed was as follows: $M_{HumanUpright}$ = 14.4, SD = 3.94; $M_{HumanInverted}$ = 13.27, SD = 3.13; $M_{MonkeyUpright}$ = 14.4, SD = 4.09; and $M_{MonkeyInverted}$ = 13.47, SD = 2.72.

Stimuli

The raw stimuli used in the original study were provided by Olivier Pascalis. We edited these using Adobe Photoshop according to the descriptions given in Halit, de Haan, and Johnson (2003). We presented faces of humans (n = 3) and of macaque-monkeys (n = 3). Each of these faces appeared in an upright or in an inverted orientation in the centre of the screen for 1000 ms. The background-colour was set to grey. The faces had a neutral expression and their gaze-direction was facing straight ahead toward the infants. The images showed the faces from a fully frontal view.

Data analysis

Following the procedure described in Halit, de Haan, and Johnson (2003), we re-referenced the data to the average-value of all the channels used. The data were inspected manually for artefacts, and trials in which a participant had not been attending to the screen or displayed movement- or eye-artefacts were rejected. For the

original article, the statistical analysis had been conducted on the mean-amplitudes from six topographically distinct channel-groups. As our EEG-system did not feature the same number of electrodes as the one used in the original study, we decided to constrain our analyses to one channel-group per hemisphere that would approximately cover the same posterior areas as the original article, based on 10-20 locations. Therefore, we compared the data from one left and one right medial channel-group featuring the following electrodes: P7, P3, and O1 (left medial) and P8, P4, and O2 (right medial), respectively.

Congruent with Halit, de Haan, and Johnson (2003), we analyzed the data originating from these four conditions for the N290 and for the P400. For the N290-analysis, we extracted the most negative peak for each of the above-mentioned electrodes in the time-window from 140 to 380 ms after stimulus-onset. Then, we calculated one mean per channel-group from the negative peaks that had been obtained for the three electrodes in each channel-group. This mean-amplitude entered the statistical analysis for the N290. Moreover, we took note of the latencies at which the negative peaks had occurred at the individual channels and entered the latency of the most negative value in the analysis for each channel group. For the analysis of the P400, we determined the most positive peak for each of the above-mentioned electrodes for the time-period from 380 to 584 ms after stimulus onset. Then, we calculated one mean for each of the two channel-groups from the positive peaks that had been obtained for each of the three electrodes in the channel-groups. These two means then entered the statistical analysis. For both components, the N290 and the P400, the analysis of the potential difference in amplitude was conducted using a 2 x 2 x 2-repeated measures ANOVA with the factors Species (Human and Monkey), Orientation (Upright and Inverted), and Location (Left medial and Right medial). The

same statistical test was used to establish potential differences in the latencies of the N290 and of the P400.

Study 3 – the Nc and the PSW (Hoehl, Reid, Mooney, & Striano, 2008)

Participants

14 of 18 tested participants entered the final analysis. Three infants had to be excluded from the analysis because they did not reach the minimal criterion of providing at least 10 artefact-free trials to the averaging process due to fussiness or inattentiveness (n = 3). One infant was excluded due to excessive frequency artefacts in the data (n = 1). Therefore, the overall attrition rate for this analysis was 22.22%. The mean-number of artefact-free trials contributed by the infants across conditions was 13.22 (range: 10 to 22). Trial-contribution for the two individual conditions was as follows: M_{Toward} = 12.93, SD = 3.52; and M_{Away} = 13.5, SD = 3.7.

Stimuli

We were provided with the images that had been used in the original study by Stefanie Hoehl. The images depicting a female adult who is looking toward or away from a toy positioned at eye-level to the side of her head were presented for 1000 ms each. Additionally, we presented the same object that Hoehl et al. (2008) had used as a central attractor for 500 ms prior to the target stimulus. The background-colour was set to white both for the central attractor and for the target-stimuli.

Data analysis

Hoehl, Reid, Mooney, and Striano (2008) analyzed the data of their 4-month-olds for potential differences in the peak-amplitudes and -latencies of the Negative component (Nc) between 400 and 600 ms after the onset of the target images. Moreover, they compared the mean-amplitudes of the time-window between 700 and

1000 ms after target-onset for a difference between the conditions in the Positive Slow-Wave (PSW).

After visual inspection of the individual averages in the obtained data for these two conditions, it was apparent that the peaks of the Nc frequently appeared slightly earlier than 400 ms after stimulus onset. This can be explained with the developmental changes that are expected to have taken place between the ages of 4 months and 1 year of age (cf., Rochat, & Striano, 1999; Tomasello, 1995; Taylor & Baldeweg, 2002). Therefore, we decided to shift the time-window for the analyses of the Nc-peak and its latency from 400 to 600 ms after stimulus-onset to the time-window from 350 to 500 ms after stimulus-onset. Moreover, due to a slight difference in the EEG-equipment used, our left-frontal channel-group featured F3, FC5 (instead of FC3), and C3, and our right-frontal channel-group featured F4, FC6 (instead of FC4), and C4, respectively. The fronto-central channel-group consisted of Fz and Cz.

Results

Study 1 – the Spike-Potential (Csibra, Tucker, Volein, & Johnson, 2000)

The cognitive component under investigation in Csibra, Tucker, Volein, and Johnson (2000) was the spike-potential – a positive peak located at 10 and 18 ms prior to a saccade-onset towards a target-stimulus. As the spike-potential was present in both experimental conditions (i.e., trials with or without a gap), the authors had conducted their statistical analysis collapsed across these two conditions. Including the data from 10 out of 26 participants (attrition = 61.5%), they reported significant differences in amplitude between the latencies at which the saccade-onsets appeared compared to the latencies at 10 ms and at 18 ms before the respective saccade-onsets; $F_{\text{-}10ms}(1, 9) = 22.5$, $p < .002$ and $F_{\text{-}18ms}(1, 9) = 8.1$, $p < .02$ (Csibra, Tucker, Volein, & Johnson, 2000, p. 1071). The data from 14 of our 18 1-year-olds replicated these results. The results of our amplitude-comparison between saccade-onsets and the latency at -10 ms to the respective saccade-onsets was $t_{\text{-}10ms}(1, 13) = 5.19$, $p < .001$. The results of our amplitude-comparison between saccade-onsets and the latency at -18 ms to the respective saccade-onsets was $t_{\text{-}18ms}(1, 13) = 4.04$, $p = .001$ (see Table 1 for a list of the respective amplitudes from the individual participants and Figure 3 for an illustration of the waveform at channel Pz).

As mentioned above, we also conducted paired-samples t-tests using data from a subset of five of the 14 infants who entered our final analysis for these conditions. This subset of participants contributed 10 usable trials for both of the conditions whereas the remaining nine contributed unequal trial-numbers to the average which was collapsed across conditions. The paired-samples t-test between saccade-onsets and the latency at -10 ms showed no significant difference between the amplitudes at those

two time-points. However, when comparing the amplitudes at saccade-onsets and at 18 ms prior to that sample-point, we found a strong tendency for a significant difference; $t_{-18ms}(1, 4) = 2.74$, $p =. 052$. Amplitudes were more positive at -18 ms ($M = -0.363$) compared to saccade-onsets ($M = -3.368$).

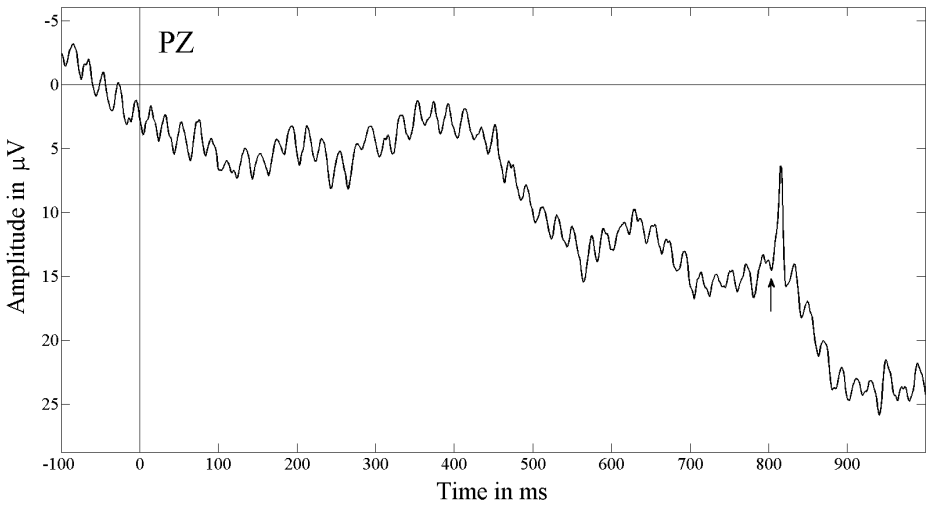

Figure 3. An illustration of the waveform at channel Pz derived from the averages of 14 participants collapsed across the conditions. It needs to be noted that time-point 0 does not represent the onset of the saccades that the infants performed but rather the onset of the target stimuli. Moreover, as the latencies at which the infants started their saccades were extremely variable within and between individual participants, trials, and conditions, great care needs to be taken in interpreting the positive peak at the onset of the sharp negative peak shortly after 800 ms (see arrow) as the true location of the spike-potential.

Table 1. An overview over the mean-amplitudes established from the 14 1-year-olds who contributed data to our analysis of the spike-potential in comparison to the time-points at -10ms and at -18ms before saccade-onset.

Participant-Number	*Mean-Amplitude at Saccade-Onset in µV*	*Mean-Amplitude at -10ms in µV*	*Mean-Amplitude at -18ms in µV*
02	25.235	58.421	32.603
03	4.859	11.582	6.813
04	-7.227	2.329	-6.065
06	-6.038	-1.623	-5.382
07	18.631	30.84	22.056
08	1.694	13.989	4.072
09	-6.362	17.526	-4.409
11	13.566	20.53	16.684
12	-9.325	18.024	-7.336
13	3.298	33.310	5.63
14	-2.818	7.101	-0.57
16	-4.802	23.715	0.673
17	-31.574	47.115	-24.43
18	5.634	14.931	6.454
Mean	0.341	21.271	3.342

Study 2 – the N290 and the P400 (Halit, de Haan, & Johnson, 2003)

The N290

The authors of the original paper analyzed the differences in amplitude for two components: the N290 (between 140 and 380 ms after stimulus-onset) and the P400 (between 380 and 584 ms after stimulus-onset). Using the data from 26 out of 85 participants (attrition = 69.4%), they reported a main-effect of Species on the amplitude of the N290 ($F(1, 24) = 15.57$, $p < .01$; cf. Halit, de Haan, & Johnson, 2003, p. 1184) with the responses to human faces being significantly more negative than those for monkey-faces. Moreover, they found the Orientation of the human and monkey-faces to have an impact on the amplitude of the N290 as well. A post-hoc test revealed that the inverted faces yielded more negative responses than the upright stimuli; $F(1, 24) = 7.21$, $p < .02$ (p. 1185). The data from 15 of our 18 participants showed a tendency towards an interaction between Species and Orientation impacting on the amplitude of the N290 ($F(1, 14) = 3.36$, $p = .088$). For human faces, the amplitude of the N290 was more negative for upright faces (M = -0.226 μV, SE = 1.432) compared to the inverted ones (M = 2.32 μV, SE = 2.074). For monkey-faces, the amplitude of the N290 was more negative for the inverted faces (M = 0.832 μV, SE = 1.727) compared to the upright ones (M = 3.599 μV, SE = 2.091; see Figure 4).

With respect to the latency of the N290, Halit, de Haan, and Johnson (2003) reported a main-effect of Species ($F(1, 24) = 16.00$, $p < .005$; p. 1185) with the latencies for human faces being significantly longer than those for monkey-faces. The Orientation of the faces did not have an effect on the latency of the N290 in the original study. The data from our group of 12- to 13-month-olds did not show any differences in latency for the N290.

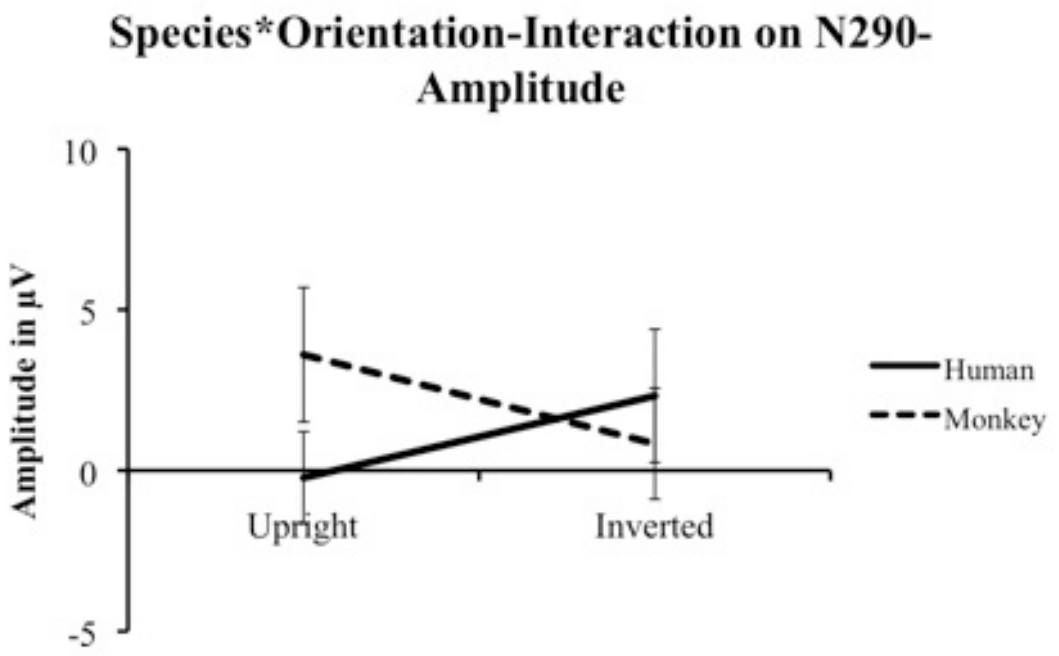

Figure 4. The data from our group of 1-year-olds showed a tendency towards an interaction of Species and Orientation on the amplitude of the N290 ($F(1, 14) = 3.36$, $p = .088$). The upright human faces (± SE) yielded more negative responses than the inverted human faces (± SE). However, the inverted monkey-faces (± SE) yielded the more negative responses compared to the upright monkey-faces (± SE).

The P400

Halit, de Haan, and Johnson (2003) did not find any effects on the amplitude of the P400 in their data. Our data revealed a strong significant interaction between Species and Orientation that impacted on the amplitude of the P400; $F(1, 14) = 5.393$, $p = .036$. In our data, the P400-amplitudes were significantly more positive for the inverted human faces ($M = 32.705$ µV, $SE = 3.898$) compared to the upright human faces ($M = 25.111$ µV, $SE = 2.692$), whereas the upright monkey-faces ($M = 33.327$ µV, $SE = 4.697$) yielded the significantly more positive responses than the inverted monkey-faces ($M = 29.426$ µV, $SE = 2.416$; see Figure 5).

With respect to the latency of the P400, the authors of the original paper reported a main-effect of Species ($F(1, 24) = 11.28$, $p < .01$; Halit, de Haan, & Johnson, 2003, p. 1185) with the P400 for human faces having a shorter latency than the one for monkey-faces. Our data showed a significant effect of Orientation on the latency of the P400. Upright faces ($M = 474.75$ ms, $SE = 15.879$) had a significantly later P400 than inverted faced ($M = 451.63$ ms, $SE = 15.273$) irrespective of species ($F(1, 14) = 6.221$, $p = .026$; see Figure 6). For a comparison of the ERPs reported in Halit, de Haan, and Johnson (2003) and those resulting from our data see Figures 7a and 7b below.

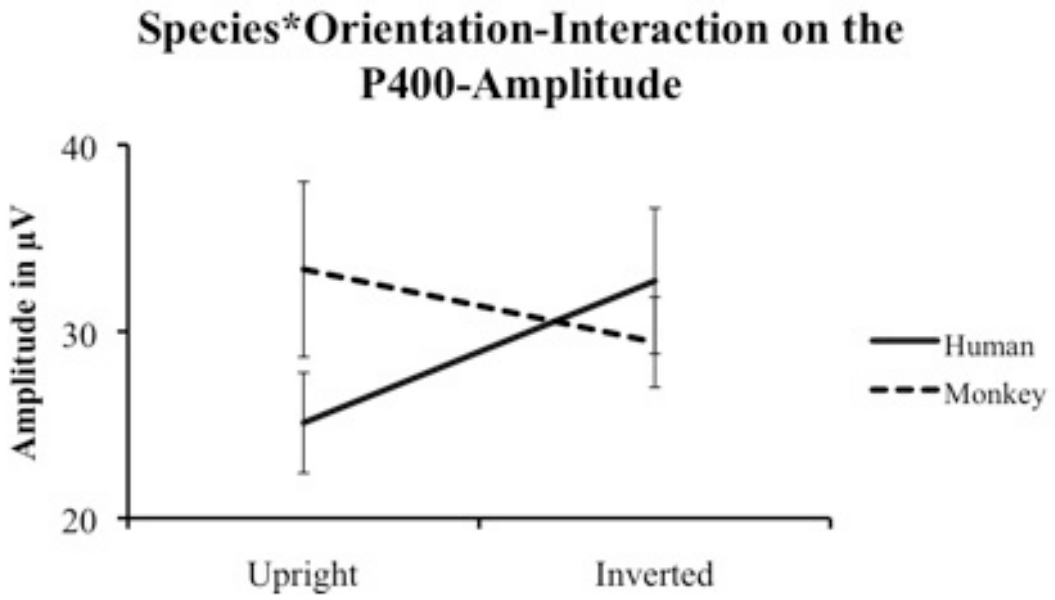

Figure 5. The data from our group of 1-year-olds showed a significant interaction between Species and Orientation impacting on the amplitude of the P400 (F(1, 14) = 5.39, p = .036). The inverted human faces (± SE) yielded the more positive responses than the upright human faces (± SE). However, the upright monkey-faces (± SE) yielded the more positive responses compared to the inverted monkey-faces (± SE).

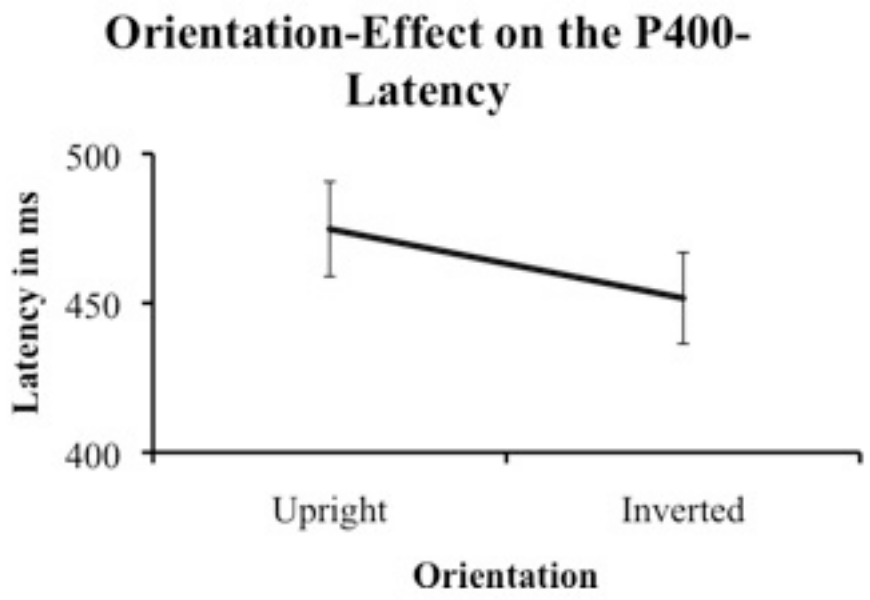

Figure 6. The data from our group of 1-year-olds showed a significant effect of Orientation on the latency of the P400. The faces presented in an upright orientation (± SE) had significantly longer latencies for the P400 than the inverted faces (± SE; $F(1, 14) = 6.22$, $p = .026$). This effect was collapsed across species.

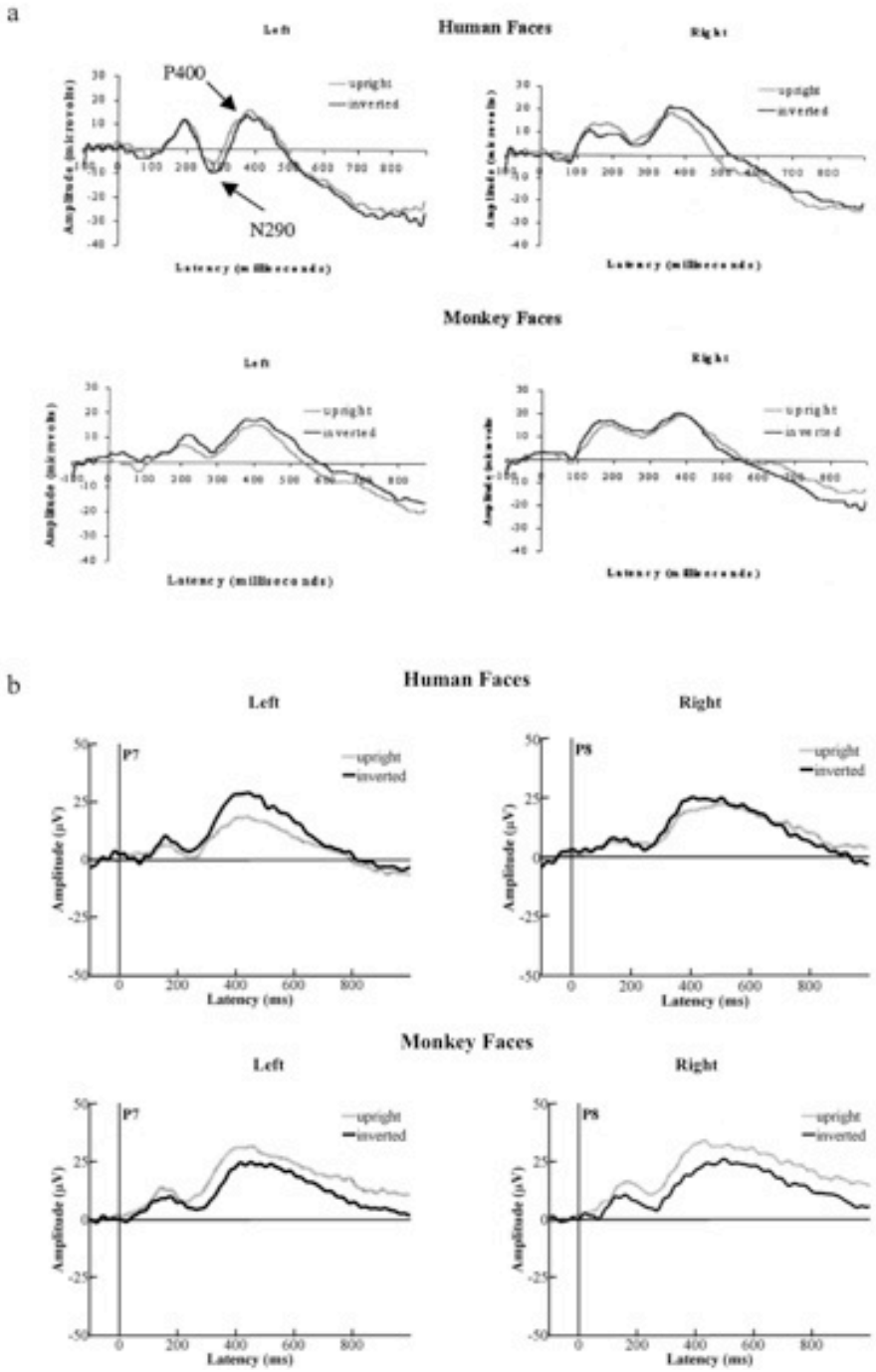

Figure 7. Part (a) depicts the grand-average ERPs that Halit, de Haan, and Johnson's (2003; Figure 1, p. 1183) had deducted from their data. Part (b) illustrates the grand-average ERPs that we deducted from the data from our group of 12- to 13-month-olds. As can be seen, the morphologies of the original ERPs and ours are very similar. However, the directionalities in terms of which condition elicited the more negative or positive responses are reversed.

Study 3 – the Nc and the PSW (Hoehl, Reid, Mooney, & Striano, 2008)

The Negative component (Nc)

The data of 17 of 64 4-month-olds tested in the original study (attrition = 73.4%) revealed a main-effect of Condition on the peak of the Nc for the time-window from 400 to 600 ms after stimulus-onset. The stimuli depicting the condition with the averted gaze showed significantly more negative responses compared to the Toward-condition ($F(1, 16) = 4.8$, $p = .044$; cf. Hoehl, Reid, Mooney, & Striano, 2008, p. 13). Moreover, in the condition in which the adult was looking toward the toy, the Nc-peak had a significantly shorter latency than the Away-condition ($F(1, 16) = 5.52$, $p = .034$; p. 13). The 2 x 3-repeated measures ANOVAs on the data from 14 of 18 participants in our study did not show any effects of Condition on either the amplitude of the Nc-peak or on its latency for the time-window from 350 to 500 ms.

The Positive Slow-Wave (PSW)

Hoehl, Reid, Mooney, and Striano (2008) reported a main-effect of Condition on the mean-amplitude of the PSW for the time-window from 700 to 1000 ms after stimulus onset ($F(1, 16) = 4.68$, $p = .046$; cf. Hoehl, Reid, Mooney, & Striano, 2008, p. 12). The Toward-condition showed a significantly more positive response than the Away-condition. Our data did not show an effect of Condition on the amplitude of the PSW. However, we found a tendency toward an effect of Location; $F(1, 13) = 3.11$, $p = .061$. The right-frontal channels ($M = 21.467$ μV, $SE = 2.7$) had more positive responses compared to the left-frontal ($M = 17.868$ μV, $SE = 2.905$) and fronto-central channels ($M = 21.054$ μV, $SE = 3.841$; see Figure 8). For a comparison of the ERPs reported in Hoehl, Reid, Mooney, and Striano (2008) and those resulting from our data see Figure 9, below.

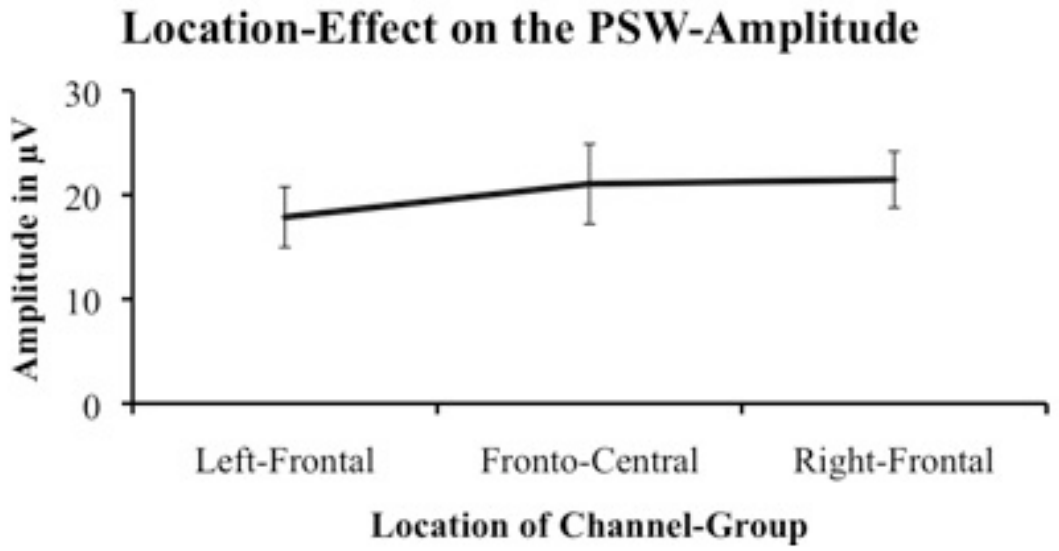

Figure 8. The data from our group of 1-year-olds showed a tendency towards an effect of Location on the amplitude of the Positive Slow-Wave (PSW). The right-frontal channel-group (F4, FC6, and C4; ± SE) had more positive amplitudes compared to the fronto-central channel group (Fz and Cz; ± SE) and the left-frontal channel-group (F3, FC5, and C3; ± SE).

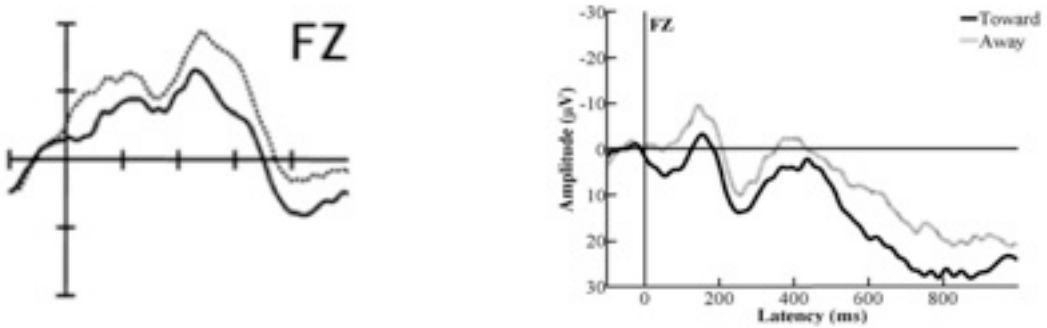

Figure 9. A comparison between the ERPs deducted from the data published in Hoehl, Reid, Mooney, and Striano (2008; Figure 2, p. 13; sample electrode depicted on the left) and those deducted from our data (sample electrode depicted on the right). As can be seen, the morphologies of the original ERPs and those from the present study are approximately similar.

Discussion

The main-objective of the current study was to investigate the consequences of presenting a group of 1-year-old infants with eight highly variable stimuli rather than the normative two experimental conditions which are typically very similar to each other in every way, other than in the experimental manipulation. In the following, we will first assess the effects this eight-condition paradigm had on attrition and put this into perspective compared to typically experienced attrition rates. Then, we will compare the number of trials collected in the present study to those presented in previous research. Finally, we will discuss our findings in comparison to those reported in the respective original articles.

General Discussion of the Methodology

According to common agreement among researchers (e.g., Thierry, 2005; DeBoer, Scott, & Nelson, 2007), infants should not be capable of meeting the high task demands required due to the extended period of time needed for providing a suitable number of trials from viewing eight conditions. It would be expected that the infants' interest in the stimulus-presentation would decrease soon after the start of the experimental session. Consequently, the infants would cease to cooperate prior to the acquisition of data in the volume required for standard ERP-analysis with this population. Moreover, were data to be acquired, the attrition rate experienced with such a paradigm would be substantial, particularly considering that studies with two conditions frequently feature attrition rates between 50 and 75% (see Stets, Stahl, & Reid, 2012). Typical examples are the studies by Kobiella, Grossmann, Reid, and Striano (2008; age group = 7 months, number of conditions = 2, attrition rate =

55.3%), de Haan and Nelson (1997; age group = 6 months, number of conditions = 2, mean-attrition rate across all four experimental groups = 63.58%), and Jeschonek, Marinovic, Hoehl, Elsner, and Pauen (2010; age group = 8 months, number of conditions = 2, attrition rate = 74.6%).

In their meta-analysis of ERP-studies with infants, which had been published between 1978 and 2010, Stets, Stahl, and Reid (2012) reported a mean-attrition rate for infant studies of 49.2% (range: 0-83.8%) for the 181 experimental groups that had been included in their final analysis. This means that, when researchers plan to conduct an ERP-study with an infant population, 50% of the time and resources that will be allocated to data-collection will usually be spent on test-sessions from which no data will be obtained. This is a considerable amount of resources that could be spent on other studies and could, therefore, be used more effectively. However, as the attrition rates reported for our analyses clearly indicate, it is possible to make the process of collecting data from infants much more efficient. We were essentially conducting three studies simultaneously. Rather than the previously predicted effect of inducing fussiness in the participants due to the increased volume of presented stimuli, in actuality, presenting the infants with eight different experimental conditions appeared to help sustain the infants' attention span. Therefore, as opposed to increasing the risk for attrition and for a premature end of the experimental session, the variability of the stimulus-presentation increased the infants' compliance and their ability to attend to the stimuli for a relatively longer period of time. Moreover, not only was it possible to test the infants for approximately 10 to 11 minutes, we were also able to extract at least 10 artefact-free trials for all of the analyses from the infants' data.

This is also in line with Stets, Stahl, and Reid's (2012) finding that there were no correlations in infant ERP-studies between the number of experimental conditions

presented to the participants and either (1) the overall attrition rate reported for a study or (2) the nature of the stimuli or (3) the attrition rate due to fussiness. However, what the authors did report, amongst other effects, was the potential presence of a selection bias (see their Figure 2, p. 235). They found that the number of studies that have a larger standard error (i.e., more variable data coming from included participants from a smaller group of originally tested infants) and are published is much smaller than the number of published studies with less variation in the data (i.e., less variable data coming from few included participants from among a large number of originally tested infants). This likely reflects a potential bias in terms of inclusion criteria and what kind of characteristics, such as a positive temperament, an infant must have in order to be included into a study's data set. As has been mentioned in Stets, Stahl, and Reid (2012), this draws the generalisability of the results into question.

As has been mentioned previously, the total number of stimuli presented in the current study when the stimulus-presentation ran uninterruptedly was 272. A minimum of 70 artefact-free trials (25.7%) were obtained during the 11-minute test-session out of the 272 which could be presented for the infants who were included in our analyses. In infant EEG-research, generally, the mean-number of artefact-free trials included in analyses is around 65 per condition (see Table 2 in Stets, Stahl, and Reid, 2012). However, there was great variation (*SD* =182.19) between studies in the number of trials used per condition. Part of this variation is likely due to the 19 studies which used odd-ball paradigms (e.g., He, Hotson, & Trainor, 2007) in which the often auditory stimuli were of short duration and were very well attended (*M* = 290.43, *SD* = 403.76, range: 8-1186). However, the mean-number of artefact-free trials included in the other 105 experimental groups reported in Stets, Stahl, and Reid (2012) was much

lower, 24.7 (*SD* = 10.8, range: 9-79.8) and is in line with the amount of data obtained in the current study.

Another important ratio is the percentage of usable trials relative to the number of trials presented to the participants per condition. As for the number of included trials, our ratio here (41.73%) is comparable to those reported in other studies using similar standard paradigms. Stets, Stahl, and Reid (2012) reported that the mean-number of stimuli presented to a participant per condition was 117.8 (*SD* = 202.39) for 131 out of the 181 experimental groups. As for the number of included trials, the standard deviation is very high due to the groups originating from odd-ball paradigms.[2] Therefore, when calculating the percentage for how many of the mean-number of trials that have been presented per condition typically enter the final analysis on average, the result is 55.52%. However, calculating the same value based on the data from only the 70 experimental groups which did not originate from a study using an odd-ball paradigm and provided data on both the number of presented and included trials per condition, the value goes down to 44.18% ($M_{presented}$ = 53.51, *SD* = 35.21, range: 1-108; $M_{included}$ = 23.64, *SD* = 9.36, range: 9 to 53) per condition. For the 18 experimental groups with odd-ball paradigms that provided data on both presented and included trials, the percentage of number of trials included on average from the number of trials presented on average is 73.17% ($M_{presented}$ = 400.98, *SD* = 352.28, range: 1 to 800; $M_{included}$ = 293.38, *SD* = 415.25, range: 8 to 1186). In the present

[2]It needs to be noted that Stets, Stahl, and Reid (2012) only presented their analysis collapsed for all types of experimental tasks and did not differentiate between experimental groups originating from standard and odd-ball paradigms. Therefore, the above numbers for the respective variables and types of experimental tasks were calculated for the present study.

study, the mean-number of trials presented per condition and analysis was 36.35 (*SD* = 4.06, range: 28.86 to 42) and the mean-number of trials included per condition and analysis was 15.17 (*SD* = 4.03, range: 10.86 to 24.29). This results in a percentage of usable trials per condition and analysis of 41.73% – again, a value that is comparable to the 44.18% from the 70 non-odd-ball experimental groups from Stets, Stahl, and Reid's (2012) meta-analysis. Therefore, despite presenting a larger array of experimental conditions, we were able to acquire the comparable amount of data per condition as is seen in two-condition paradigms. Moreover, as we had eight conditions, the method used in the present study – if validated by future studies – can be four times more efficient for data acquisition than when contrasted with standard paradigms.

Discussion of the Results of the Individual Sub-Studies

Study 1 – the Spike-Potential (Csibra, Tucker, Volein, & Johnson, 2000)

Csibra, Tucker, Volein, and Johnson (2000) showed that the spike-potential could be reliably measured at 10 and 18 ms before a saccade-onset towards a target-stimulus in both a group of adults and in a group of 12-month-olds. The results from our group of 12-month-old infants replicated Csibra et al.'s (2000) results. There were significant differences in amplitude between the time-points at saccade-onsets and at 10 and 18 ms prior to the respective saccade-onsets. However, as described above this was only the case when all infants were analysed. When only analysing the data from the infants who contributed minimally 10 trials for both gap- and bridged trials, only the amplitude-comparison between -18 ms and at saccade-onset was showing a strong trend towards a significant difference. This is likely due to the loss in statistical power

as nine of the 14 infants were removed from the dataset for this analysis allowing for a larger impact from variations between individual participants and trials.

Finally, we could replicate the so-called "sticky-eyes" phenomenon which had also been reported in Csibra et al. (2000). As in the original study, infants took longer to disengage their attention from the small object depicted in the centre of the screen and to shift their eye-gaze towards the checkerboard appearing on the side in the condition when the gap was bridged. Overall, we interpret our replication of the original findings such that the presence of the other six experimental conditions did not impact on the way the gap- or bridged trials were processed by our group of 12- to 13-month-olds.

Study 2 – the N290 and the P400 (Halit, de Haan, & Johnson, 2003)

Halit, de Haan, and Johnson (2003) reported significant main-effects of species and orientation on the amplitude of the N290. They reported more negative amplitudes for human faces relative to the monkeys and more negative responses for inverted human faces compared to upright human faces. Moreover, a main-effect of species on the latency of the N290 was found. The data from our group of 1-year-olds did not show separate main-effects of species and orientation on either the amplitude or the latency of the N290. However, we found a tendency towards an interaction between these factors that was impacting on the amplitude of the N290 only. Additionally, while Halit, de Haan, and Johnson (2003) could not find significant differences in N290-amplitude for upright and inverted monkey-faces, our data showed a tendency towards a difference in N290-amplitude which was the opposite to our results for the human faces. The latency of the N290 remained unaffected by these four experimental conditions in our study. With respect to the P400, our results also differed from the

ones presented in Halit, de Haan, and Johnson (2003). While the authors of the original paper could not report differences in the amplitudes of the P400, our data revealed a significant interaction of species and orientation for the P400-amplitude. Moreover, contrary to the main-effect of species on the P400-latency reported in the original paper, we found a main-effect of orientation, not species, on the latency of the P400.

As can be seen in Figure 7b above, despite the differences in the effects found, the ERPs sourced from our data are morphologically similar to those presented in Halit, de Haan, and Johnson (2003). Therefore, in terms of the morphology of the resulting waveform, our overall results are broadly in alignment with those reported in the original article. However, the fact that the number of conditions in our study was greater than the original study might have contributed to the discrepancies between the results of the original study and our own. It is possible that the cognitive processes involved in the processing of the other experimental conditions may have interfered with the processing of the upright and inverted human and monkey-faces, and vice versa. This may be the case given that the other stimuli also involved faces (i.e., the Toward- and Away-conditions). This may have had an impact on the way the human and monkey-faces were processed by our group of participants – even though the social contexts that are depicted in the stimuli are very different from one another.

Alternatively, given the potential changes in ERP components as shown in Stets and Reid (2011) or Nikkel and Karrer (1994), the differences in the effects and interactions found may be partly due to the minimum criterion applied in the original study and here. The infants who had been included in the analysis in Halit et al. (2003) had to contribute at least 25 artefact-free trials for each of the four conditions. Here, due to the naturally lower number of trials that could be presented per condition, the minimum criterion was set to 10. Consequently, it could be argued that the separate

main-effects reported in Halit, de Haan, and Johnson (2003) were derived from interactions between the factors of species and orientation in the earlier stages of the experimental session. This argumentation may also be supported by the similarity between the waveform morphologies that were derived from our data and those presented in the respective original study (see Figure 7 above).

Study 3 – the Nc and the PSW (Hoehl, Reid, Mooney, & Striano, 2008)

Hoehl, Reid, Mooney, and Striano (2008) reported their group of 4-month-old infants to show significantly more negative responses towards the stimuli depicting the Away-condition. Following previous suggestions from earlier studies, this was interpreted as suggesting that the infants had allocated more attention to these stimuli. However, even though more attention had been paid to the Away-condition, the processing of the Toward-stimuli was faster as shown in the significantly shorter latencies of the Nc for this condition. The results from our group of 1-year-olds did not show any differences in the amplitude or in the latency of the Nc. This may well be due to the age of our participants compared to the ones reported in Hoehl, Reid, Mooney, and Striano (2008) and the developmental changes in terms of social information processing that will have taken place during these eight months (e.g., Rochat, & Striano, 1999).

Stets and Reid (2011) argued that the evolution of the morphology during the testing session in the Negative component in Hoehl et al.'s (2008) data might be due to the infants' familiarity with the situation in which an adult is looking toward or away from an object. The authors suggested that the more negative responses would have been found for the Toward-condition in the early stages of an experimental session because this would be the social situation with which the infants would be more

familiar. Someone looking away from a toy, however, is likely to be a situation that the infants would not have encountered during the first four postnatal months. Stets and Reid (2011) further argued that the shift in the directionality of the Nc-responses as a function of the time-course of the experiment might therefore be explained with the infants' loss of interest in, or habituation to, the Toward-stimuli. Consequently, the infant would then allocate more attention to the Away-stimuli as this condition became more interesting relative to the Toward-condition. Following this logic, it is possible to conjecture that we did not find differences in the way the Toward- and Away-conditions have been processed by our group of 12-month-olds because by this age the infants would have had multiple experiences of both these social situations. Consequently, they could be expected to allocate equal amounts of attention to both of these conditions.

Finally, Hoehl, Reid, Mooney, and Striano (2008) found an effect of condition on the amplitude of the PSW. They interpreted this as an indication that the memory trace for the toys, which had been highlighted by the adult's gaze towards them, was relatively more imprinted than the trace for the objects that were not being looked at. We did not find an effect of condition on the amplitude of the PSW. However, we found a tendency toward an effect of location that was modulating the mean-amplitude of the PSW. The right-frontal channels showed the more positive responses compared to the other two channel-groups. Comparing this effect with the morphologies of the PSWs presented in Figure 2 in Hoehl, Reid, Mooney, and Striano (2008; p. 13), one might argue that already at the age of 4 months a tendency towards this effect has been present. The ERPs of the right-frontal channels (presented on the right side of their Figure 2) were appeared to be more positive compared to the ERPs for the fronto-central channel group and reliably more positive than those for the left-frontal channel-

group. However, apparently the difference was not a significant one in the group of 4-month-olds, with the current data suggesting that eight additional months would produce location effects for this task.

As was the case for the experimental conditions originating from Halit, de Haan, and Johnson (2003), we found different results compared to the original article for the Toward- and Away-conditions (Hoehl, Reid, Mooney, & Striano, 2008). However, parallel to what we could show for the upright and inverted human and monkey-faces, the ERPs sourced from our set of data were morphologically similar to the ones presented in Hoehl et al. (2008; for a comparison see Figure 9). This indicates that our results are generally representative of the cognitive processes involved in processing such stimuli at an age of 12 to 13 months.

Implications

Scientific work with infant populations requires large amount of resources, due, in part, to high attrition rates. Attrition is frequently caused by limited infant attention capacities, coupled with the requirements for minimum trial-contributions to individual ERP-averages. Here, the variation in stimuli appears to decrease attrition rates. In our analyses, we experienced attrition rates that were between 25 and 30% lower than the average attrition rate typically reported (Stets, Stahl, & Reid, 2012) Furthermore, we demonstrated that it is possible to increase the amount of usable ERP-data obtained from infant participants. Our results show that by presenting 12- to 13-month-olds with eight as opposed to the more typical two experimental conditions, the amount of artefact-free data to be obtained may be increased three to four times. We attribute this to the variation present in the stimulus-presentation, allowing the infants' attention to be sustained for an extended period in comparison to previous studies. The approach outlined in the present study can therefore facilitate data acquisition in an infant population when contrasted to current approaches.

Another positive aspect of such a decrease in attrition rates is that a study's results can potentially be seen as being more representative of the population. For example, Marshall, Reeb, and Fox (2009) conducted an ERP-study with three groups of 9-month-olds who had been rated on their temperament by their parents prior to testing. Apart from their finding that there were processing-differences between the three groups, as a by-product, the authors noted that the attrition rates experienced for these groups differed substantially as well. The 9-month-old infants who had received negative ratings were much more likely to have to be excluded from the study's analysis compared to those rated positively and to a control group which had not been

rated at all prior to the test-session. Based on these data and on the likely selection bias reported in Stets, Stahl, and Reid (2012), it seems apparent that representativeness of a study's results cannot be assumed in all cases – especially, when the attrition rate is comparatively high. Consequently, using a paradigm that helps to decrease the likelihood for attrition can be instrumental in ensuring the generalisability of a study's results.

With respect to replicating the results of the three original research studies, our data show varying results. While some of the effects were replicated (Csibra, Tucker, Volein, & Johnson, 2000), others were not (e.g., our significant interaction between Species and Orientation which was affecting the amplitude of the P400 for the upright and inverted human and monkey faces where Halit, de Haan, & Johnson [2003] had not found any effects). Particularly in the case of the conditions originating from Hoehl, Reid, Mooney, and Striano (2008), the difference to the original findings might be due to the developmental changes between the ages of 4 and 12 months. An additional factor that might potentially have caused differences in results was interference between the face stimuli used in different conditions derived from different original experiments: both Halit, de Haan, and Johnson (2003) and Hoehl et al. (2008) included facial stimuli. Therefore, we recommend researchers to be very cautious in selecting experimental conditions when attempting to employ a paradigm similar to the one outlined in this article. The conditions presented to participants should cover distinct perceptual or cognitive domains in order to ensure that the processing of one set of conditions will not interfere with the processing of another set of conditions. For instance, it might be possible to mix two conditions on colour-perception with stimuli involving social cognition, and a task on saccades. An alternative approach would be to present the participants with a small amount of

experimental conditions but a large variety of different stimuli for each of the conditions. However, the variability of the stimuli would still be relatively limited as the stimuli could only be different to each other to a certain degree. Too much variability between the stimuli could potentially lead to unwanted differences in the neural responses towards those stimuli and, ultimately, in the responses towards the experimental conditions. Therefore, we would recommend caution with such an approach.

Conclusions

In the present infant ERP-study we tested a group of 12- to 13-month-olds with eight rather than the more typical two experimental conditions. These conditions originated from three distinct research articles addressing theoretically unrelated aspects of cognitive development. As a consequence of the high variability in the stimulus presentation, we were able to collect a sufficient number of artefact-free trials to conduct statistical analyses comparable to each of the original analyses. Moreover, for each of these three analyses, we had improved attrition rates when contrasted with studies which utilized two conditions. The effects and interactions that we could draw from our data did not entirely match those reported in the original articles. However, the morphologies of our waveforms did match those in the original publications. We conjecture that discrepancies in the reported effects and interactions were a consequence of cross-experiment interference between those conditions that included facial and, therefore, social stimuli. Therefore, we advise researchers to take great care in their choice of experimental conditions should they decide to attempt applying the outlined method for presenting stimuli in a study. However, our results suggest that the process of collecting infant ERP-data can be improved and that infants can benefit from it. Given that we could not replicate all of the original results, we are aware however that our methodological results are preliminary and require refinement and validation through further studies – also from other laboratories. Therefore, we would encourage the research community to apply similar paradigms with varying components (e.g., intermixing auditory and visual stimuli) in order to validate the approach that has been taken in the current study.

Acknowledgements

We would like to thank all the infants and their caregivers for kindly participating in our study. This work was funded by a ONE Northeast Studentship.

References

Csibra, G., Tucker, L.A., & Johnson, M.H. (1998). Neural correlates of saccade planning in infants: A high-density ERP-study. *International Journal of Psychophysiology*, *29*, 201-215.

Csibra, G., Tucker, L.A., Volein, A., & Johnson, M.H. (2000). Cortical development and saccade planning: the ontogeny of the spike potential. *NeuroReport*, *11*(57), 1069-1073.

DeBoer, T., Scott, L.S., & Nelson, C.A. (2007). Methods for acquiring and analyzing infant event-related potentials. In M. de Haan (ed.), *Infant EEG and event-related potentials* (5-37). Hove: Psychology Press.

de Haan, M., & Nelson, C.A. (1997). Recognition of the mother's face by six-month-old infants: A neurobehavioral study. *Child Development*, *68*(2), 187-210.

de Haan, M., Pascalis, O., & Johnson, M.H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199-209.

Halit, H., de Haan, M., & Johnson, M.H. (2003). Cortical specialisation for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*, *19*, 1180-1193.

He, C., Hotson, L., & Trainor, L.J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*, *19*(5), 878-892.

Hoehl, S., Reid, V., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science*, *11*(1), 10-16.

Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, *37*(3), 187-209.

Jeschonek, S., Marinovic, V., Hoehl, S., Elsner, B., & Pauen, S. (2010). Do animals and furniture items elicit different brain responses in human infants?. *Brain and Development*, *32*(10), 863-871.

Kobiella, A., Grossmann, T., Reid, V.M., & Striano, T. (2008). The discrimination of angry and fearful facial expressions in 7-month-old infants: An event-related potential study. *Cognition and Emotion*, *22*(1), 134-146.

Marshall, P.J., Reeb, B.C., & Fox, N.A. (2009). Electrophysiological responses to auditory novelty in temperamentally different 9-month-old infants. *Developmental Science*, *12*(4), 568-582.

Nelson, C.A., & Collins, P.F. (1991). Event-related potential and looking-time analysis of infants' responses to familiar and novel events: Implications for visual recognition memory. *Developmental Psychology*, *27*(1), 50-58.

Nikkel, L. & Karrer, R. (1994). Differential effects of experience on the ERP and behavior of 6-month-old infants: trends during repeated stimulus presentation. *Developmental Neuropsychology*, *10*, 1-11.

Parise, E., Reid, V.M., Stets, M., & Striano, T. (2008). Direct eye contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience*, *3*(2), 141-150.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson Jr., R., … Taylor, M.J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, *37*, 127-152.

Picton, T.W., & Taylor, M.J. (2007). Electrophysiological evaluation of human brain development. *Developmental Neuropsychology*, *31*(3), 249-278.

Rochat, P., & Striano, T. (1999). Social-cognitive development in the first year. In P. Rochat (ed)., *Early social cognition: understanding others in the first months of life* (3-34). Mahwah, NJ: Lawrence Erlbaum.

Stets, M., & Reid, V.M. (2011). Infant ERP amplitudes change over the course of an experimental session: Implications for cognitive processes and methodology. *Brain and Development*, *33*(7), 558-568.

Stets, M., Stahl, D., & Reid, V.M. (2012). A meta-analysis investigating factors underlying attrition rates in infant ERP-studies. *Developmental Neuropsychology*, *37*(3), 226-252.

Taylor, M.J., & Baldeweg, T. (2002). Application of EEG, ERP and intracranial recordings to the investigation of cognitive functions in children. *Developmental Science*, *5*(3), 318-334.

Thierry, G. (2005). The use of event-related potentials in the study of early cognitive development. *Infant and Child Development*, *14*(1), 85-94.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore and P.J. Dunham (ed.s), *Joint attention: its origins and role in development* (103-130). Hillsdale, NJ: Lawrence Erlbaum.

Counter to what would generally be expected to happen when infants are presented with a comparatively high number of experimental conditions, Study 3 showed that the increased variability in the stimulus-presentation helped the infants to remain attentive. This led to a decrease in the attrition rate experienced in this study. Moreover, the numbers of usable trials per condition were comparable to typical infant ERP-studies that feature fewer conditions and include a minimum criterion of 10 artefact-free trials per condition. However, due to the high number of experimental conditions in Study 3, the overall number of analysable trials that could be collected from an infant was increased in general. This means that the approach to study-design outlined in Study 3 may well be more time- and cost-effective that the strategies that are currently used by most developmental psychologists. Aiming to further explore the findings of both Study 1 and Study 3, the final study (Study 4) combines the strategies outlined in the earlier two. Therefore, the data collected for the eight experimental conditions in Study 3 are analysed along the lines described in Study 1, with an examination of how individual trials and their order of acquisition may interact with the development of morphological aspects of the ERP. As shown in Chapter 2, a more detailed analysis of infant ERP-data may provide new insights into infant cognitive processes. Moreover, with respect to the interferences between the conditions found in Study 3, a clearer image of the dynamics underlying the impact of the study-design on the infants' neural responses may emerge. This study seeks to bring these issues together and, in so doing, merge those factors into a final experimental culmination within this thesis.

Chapter 5

The Fine Print of Infant ERPs: The Effect of Additional Trials on Components

Manuela Stets, Mike Burt, and Vincent M. Reid

(in preparation).

Abstract

Extracting sufficient numbers of artifact-free trials per condition from an infant EEG to compute meaningful ERP-averages is challenging in standard paradigms, which typically feature two experimental conditions. The common practice of excluding data from participants providing fewer than typically 10 or 15 per condition only serves to intensify this challenge. Recent research has shown both that meaningful information can be obtained from participants who do not meet the standard minimal criterion and that infants are capable of meeting a minimal criterion of 10 artifact-free trials when varied stimuli are presented and, therefore, cognitive workload is higher. We therefore investigated the contribution of individual ERP-trials to the amplitudes and latencies of infant ERP-components across an array of conditions that were obtained in a within-subjects design in the same testing session.

In order to investigate the potential dynamics underlying the acquisition of different infant ERP-components, analyses were performed including only the first trial in each condition and the results were compared with different numbers of usable trials, up to a maximum of 10 trials. We hypothesized that amplitude-changes across an experimental session may be found in all components, whereas previously this has been found only in attention-related components. Additionally, we theorized that the latencies of individual aspects of ERP-morphology might be affected by the number of trials included in an individual's average. Our results show that amplitude modulations including a change in component-directionality are restricted specifically to attention-related components whereas latencies can undergo changes on a more regular basis dependent on how much exposure an infant has had to a certain kind of stimulus. We conclude that investigating the time-course of the development of an infant's EEG-

signal can uncover new insights into infant cognition as such dynamics have previously been largely unknown.

Introduction

In order to obtain a reasonable number of artifact-free trials per condition from an infant participant in an ERP-study, infants are normally tested for as long as possible (Hoehl, & Wahl, 2012). In a standard paradigm in which conditions are presented with equal frequency, the minimum criterion for entering a participant's data into the final analysis can be as low as five artifact-free trials per condition (e.g., Grossmann, Oberecker, Koch, & Friederici, 2010) or as high as 30 (e.g., Csibra, Davis, Spratling, & Johnson, 2000). In infant odd-ball paradigms which typically present brief auditory stimuli, the minimal criterion for data-inclusion is usually much higher (e.g., 65 deviant trials in Weber, Hahne, Friedrich, & Friederici, 2004; 100 odd-ball trials in He, Hotson, & Trainor, 2007). In order to obtain high numbers of analyzable trials per condition from an infant, researchers tend to try and use a great diversity of methods to keep the participants happy and interested in the stimulus-presentation. For example, Richards (2003) presented his test-trials embedded in an episode of *Sesame Street* in order to draw and maintain infants' attention. Csibra, Tucker, and Johnson (1998) used various background-sounds presented in tandem with the visual stimuli to help their 6-month-olds maintain interest. It is also common to give the infants short breaks from the stimulus-presentation at the first signs of distress or inattentiveness, i.e., when an infant becomes "fussy". Despite the lack of a solid definition for this state from the literature, infants who cry or show negative affect, are often labeled "fussy". The issue of how and when to intervene in such a case is left to the discretion of the researcher in any given test-session. However, making such a crucial decision, which is likely key to which infants will provide data to the statistical analysis, requires extensive training and experience.

If an infant cannot provide the requisite number of trials in all conditions involved in a standard-paradigm ERP-study, then their data will not be included in the study's final analysis and the infants' only contribution to the study will be to increase its attrition rate. In a meta-analysis of published articles, Stets, Stahl, and Reid (2012) investigated the possible connections between study-features such as the duration of stimuli or the nature of the stimuli used in a study and the reported attrition rate for the respective infant ERP-study. According to the results of their meta-regression, the majority of the study-features under investigation did not impact on attrition rates – one study feature that surprisingly did not impact was the number of conditions presented (Stets, Stahl, & Reid, 2012). Within infant EEG-research the number of conditions is usually small in number in order to try and avoid overloading the infants' cognitive systems. Recommendations along these lines can be found frequently in methodological literature (e.g., Hoehl, & Wahl, 2012; Picton et al., 2000). Consequently, the mean-number of conditions reported in the meta-analysis by Stets, Stahl, and Reid (2012) was 2.39 with a range from 1 to 6 (see Table 2, p. 234). However, in their recent ERP-study with a group of 12-month-olds, Stets, Burt, and Reid (under review) presented their participants with eight experimental conditions originating from three distinct research articles (Csibra, Tucker, Volein, & Johnson, 2000; Halit, de Haan, & Johnson, 2003; Hoehl, Reid, Mooney, & Striano, 2008). Stets, Burt, and Reid (under review) obtained a minimum of 10 artifact-free trials per condition from at least 14 of their 18 participants. It appeared that the infants were attracted by the comparatively high number of experimental conditions that gave rise to the high variability of the stimuli presented. Rather than losing interest in the stimuli shortly after the start of the presentation, the infants' attention was sustained, yielding an increase in obtained data by a factor of three or four while simultaneously

decreasing the overall attrition rate to as low as 15%. This is in stark contrast to the more typical attrition rates of between 61.5% and 73.4% in the original studies (Stets, Burt, & Reid, under review).

Another study-feature that was included in the aforementioned meta-analysis and found not to have any impact on attrition rate was the mean-number of artifact-free trials per condition included in the analyses (Stets, Stahl, & Reid, 2012). On average 65.4 trials were found to be included per condition. However, as Stets, Burt, and Reid (under review) pointed out, this average is inflated by the very high numbers of trials used in odd-ball paradigms, with the mean for standard (non-oddball) experimental paradigms equaling 24.7. Despite – or because of – the comparatively high number of experimental conditions that were presented to the 12-month-olds in Stets, Burt, and Reid (under review), the majority of infants provided the minimum of 10 artifact-free trials for all of the conditions and analyses in that study (M = 15.2; Stets, Burt, & Reid, under review). The proportion of useable trials (which averaged at around 42%) in Stets, Burt, and Reid's (under review) experiment was similar to other studies using standard paradigms (for which 44.2% of trials were useable as calculated for the 70 standard-paradigm groups presented in Stets, Stahl, and Reid's [2012] meta-analysis).

Based on these methodological findings, it is apparent that currently held assumptions about how ERP-paradigms interact with infant participants should be re-evaluated. Additionally, as has been shown in Stets and Reid (2011), the assumption that ERP-components are stable in terms of component-directionalities should be treated with caution since, according to their re-analysis of ERP-data from a group of 4-month-olds (originally published by Hoehl, Reid, Mooney, & Striano, 2008), the number of trials included in the averaging process can impact on the results obtained. For example, with 10 or more trials included in the average, Hoehl et al. (2008)

reported the Negative component (Nc) to have a more negative peak in the condition where the adult was looking away from an object compared to when the adult was looking toward the object. In comparison, using the same data, Stets and Reid (2011) found the effect to be reversed when the first seven trials for each of the two conditions were analyzed. Moreover, analyzing only the first three trials from the 17 originally analyzed participants and from seven infants previously excluded for not contributing a minimum of 10 trials per condition, found a significant difference between conditions in the Nc-peak (Stets, & Reid, 2011) at this early point in the experimental paradigm.

Finally, another frequently reported aspect of infant-derived ERP-components is their latency, which provides information on how fast or slow a stimulus is processed by a participant. This is usually interpreted as an indication for the amount of effort an individual needs to make to process a particular stimulus in its entirety (Rugg, & Coles, 1995). Whereas alterations in ERP-amplitudes are occasionally reported (e.g., Nikkel, & Karrer, 1994), ERP-latencies have so far been reported to be stable over the course of an experimental session (e.g., Polich, 1986). When significant effects related to latency have been reported within infancy literature, they have been typically related to developmental processes as shown by different processing-speeds in groups of participants representing different developmental stages (e.g., Little, Thomas, & Letterman, 1999). This phenomenon was also apparent at a gross level in Stets, Burt, and Reid's (under review) data: the time-window for the analysis of the Nc had to be moved for the 12-month-old participants compared to the one used in Hoehl et al. (2008). For their 4-month-olds, Hoehl et al. (2008) used the time-window between 400 and 600 ms after stimulus-onset. Stets, Burt, and Reid (under review) moved the time-window to between 350 and 500 ms in order to accurately capture the

peak of the Nc component as, when visually inspecting the individual averages, it appeared that it often occurred very close to and earlier than at 400 ms. Therefore, for a study with highly variable visual stimuli as those presented in Stets, Burt, and Reid (under review), the question arises whether this generally experienced stability in latency in response to stimuli could be replicated or should even be expected. Based on the latency-differences between the effects reported for the N290- and the P400-latencies in Halit, de Haan, and Johnson (2003) as compared to those reported in Stets, Burt, and Reid (under review), this assumption might not be met under all circumstances – especially so given that the effects found for the amplitudes also did not always match those reported in the original studies. One could reason that the amplitude- and latency-differences between the original study and Stets, Burt, and Reid (under review) result from the apparent differences between the paradigms (i.e., four experimental conditions in Halit, de Haan, & Johnson [2003] versus eight conditions in Stets, Burt, & Reid [under review]).

The analysis presented in this paper aims to understand the behavior of ERP-components when increasing numbers of trials are used to form the average waveform for individual participants. ERPs were recorded from participants viewing multiple conditions within the same experimental session. As noted by Nikkel and Karrer (1994) and de Haan, and Nelson (1999, p. 1114, Footnote 1), Stets and Reid (2011) found changes in the Nc-amplitude depending on how often an infant had been presented with an experimental condition. In this paper, we will follow the same routine for these conditions. Moreover, to our knowledge, no attempts have been made to investigate potential changes in the morphologies of other infant ERP-components as a function of the number of trials present in the average.

Having presented their 12-month-olds with eight experimental conditions, Stets, Burt, and Reid (under review) reported on five separate ERP-components: (a) the Spike-Potential (SP), (b) the N290, (c) the P400, (d) the Nc, and (e) the Positive Slow-Wave (PSW). The SP is related to an individual's preparation to shift eye-gaze direction and was recorded in response to stimuli presenting a gap or a bridged period between an attractor and a target checkerboard (Csibra, Tucker, Volein, & Johnson, 2000). The N290 and the P400 are reported to be involved in face processing and face recognition and were recorded in response to human and monkey faces presented in an upright or inverted orientation, looking straight at the infant (de Haan, & Nelson, 1999; Halit, de Haan, & Johnson, 2003). The Nc has previously been related to attention-allocation (Courchesne, Ganz, & Norcia, 1981), and the PSW is frequently reported to be related to embedding information about a certain type of stimulus in memory or in updating environmental context (e.g., Webb, Long, & Nelson, 2005). These two components were recorded in response to females directing their eye-gaze either toward or away from a toy, never at the infant. Both the N290 and the P400 and the Nc and the PSW seem to be closely related and frequently co-occur. We hypothesize that amplitude-changes, as seen in Stets and Reid (2011), might also be present in other infant components (apart from the Nc) depending on the number of artifact-free trials included in the analyses, i.e., how far an experimental session has progressed and, ultimately, how much exposure an individual had to an experimental condition. Therefore, we used the first 10 trials which were of an analyzable quality for each of the conditions from the datasets presented in Stets, Burt, and Reid (under review). The current study attempts to replicate Stets and Reid's (2011) finding of changes in the Nc due to the number of trials included in the average. Moreover, we hope to assess whether amplitude-modulations caused by the increasing amount of experience a

participant has with the conditions, are only to be found in components connected to attention-processes or, alternatively, if such modulations can also be found in infant ERP-components that are not thought to be related to attention. Furthermore, we aim to clarify if similar changes can be found in the components' latencies or if such modulations are solely to be seen in the amplitudes of the components.

Methods

Participants, stimuli, procedure, EEG recording and data-analysis

The data analyzed in the current study were presented in Stets, Burt, and Reid (under review). Therefore, for the specifics concerning all experimental details, please refer to Stets, Burt, and Reid (under review). In the following, we will outline the particulars of the new analyses that we conducted on these data.

Data Re-Analysis

We conducted the same statistical tests that had been performed in Stets, Burt, and Reid (under review). However, instead of using all available artifact-free trials in the average waveforms, they were created by including increasing numbers of trials from trial 1 through to trial 10 (i.e., one average with only the first trial, one average with the first two trials, etc.). In the following, we will refer to them as "F1" for "first trial only", "F2" for "first two trials only", "F3" for "first three trials only", and so on. Stets, Burt, and Reid's (under review) experiment combined the stimuli and conditions from three separate papers each of which contributed data to a separate analysis. Analysis 1 followed the statistical analysis described in Csibra et al. (2000) with trials being collapsed across the two conditions (G. Csibra, personal communication, September 25, 2012). Paired-samples t-tests were performed in which (a) the mean-amplitudes measured at three sample-points from around the saccade-onsets to the mean-amplitudes from three sample-points from around 10 ms prior to the respective saccade-onsets and (b) the mean-amplitudes measured at three sample-points from around saccade-onsets to those from three sample-points from around 18 ms prior to the respective saccade-onsets were compared.

Analysis 2 used the data from the conditions including human and monkey faces presented in upright and inverted orientations (see Halit, de Haan, & Johnson, 2003). First, we created averages including only specific numbers of trials. Then, we performed 2 x 2 x 2-repeated measures ANOVAs with the factors Species (Human and Monkey), Orientation (upright and inverted), and Location (a Left medial channel group and a Right medial one) on both the amplitudes and latencies for both the N290 (in the time-window from 140 to 380 ms after stimulus-onset) and the P400 (in the time-window from 380 to 584 ms after stimulus-onset).

In Analysis 3, the data originating from the Toward- and Away-conditions were assessed (see Hoehl, Reid, Mooney, & Striano, 2008). We conducted 2 x 3-repeated measures ANOVAs with the factors Condition (Toward and Away) and Location (a left-frontal channel group, a fronto-central group, and a right-frontal group) for the amplitudes of the Nc (in the time-window from 350 to 500 ms after stimulus-onset) and the PSW (in the time-window from 700 to 1000 ms after stimulus-onset) as well as for the latency of the Nc. Again, these analyses were conducted for datasets including decreasing numbers of trials from 10 down to the first usable trial.

As has been explained above, the averages used in these analyses consist of increasing numbers of trials by adding one artifact-free trial to the previous average. This means that the datasets are not independent from each other and, on the contrary, strongly related to each other. Consequently, the possibility arises that, in our analyses with multiple comparisons, significant results may be found by chance rather than based on a true difference between the investigated factors, i.e., a Type I error or a false positive may occur (Sheskin, 2011). Therefore, we performed post-hoc one-way ANOVAs on the datasets which are building on each other in order to establish if there were significant differences between them that may give rise to a false positive. We

performed these one-way ANOVAs including a post-hoc Bonferroni-correction. Moreover, we attempted to incorporate data from excluded participants where possible – following the approach outlined in Analysis Two in Stets and Reid (2010). In the cases in which this was possible, the respective results will be indicated below. However, it was not possible to include all 18 participants' data in any of the analyses as the data of one infant in particular were too contaminated with movement-artifacts while in all other cases, not more than one artifact-free trial could be obtained from some of the conditions. Therefore, in order to adhere to the common strategy to include infants only to analyses to which they can contribute equal numbers of trials, further participants had to retain their excluded status.

Results

The analysis below is organized to follow similar conventions to those of the three experiments whose stimuli were interleaved with each other to create the single experiment. These data were initially presented in Stets, Burt, and Reid (under review) to assess whether infants would attend more to experiments with a highly variable stimulus-presentation.

Analysis 1 – the Spike-Potential

Csibra, Tucker, Volein, and Johnson (2000) reported significantly more positive amplitudes at 18 and at 10 ms prior to a saccade-onset towards a target-checkerboard in their group of 10 12-month-olds (M_{SacOns} = 0.8 μV, M_{-10ms} = 3.2 μV, $F_{-10ms}(1, 9) = 22.5$, $p < .002$; M_{-18ms} = 2.1 μV, $F_{-18ms}(1, 9) = 8.1$, $p < .02$; p. 1071). Stets, Burt, and Reid (under review) replicated these results with their group of 14 12-month-olds (M_{SacOns} = 0.34 μV, M_{-10ms} = 21.27 μV, $t_{-10ms}(1, 13) = 5.19$, $p < .001$; M_{-18ms} = 3.34 μV, $t_{-18ms}(1, 13) = 4.04$, $p = .001$; p. 13). These effects appear to be highly robust and are present in the data from the first presentations onwards (i.e., from F1 through to F10). All of the amplitude-comparisons between the sampling points at saccade-onsets and at -10 ms showed significant differences in the F-datasets; $p \leq .001$ (see Table 1 and Figure 1). Additionally, when comparing the amplitudes at the sampling-points at saccade-onsets and at -18 ms, all datasets from F1 up to F10 showed significant differences between the amplitudes; $p \leq .01$ (see Table 1 and Figure 1). Moreover, in an additional analysis of F1 including data from two of the previously four excluded participants, amplitude differences between -18 ms and saccade onsets and between -10 ms and saccade onsets were found to be significant as well; $t_{-18ms}(1, 15) = 3.34$, $p =$

.004 and $t_{-10ms}(1, 15) = 4.478$, $p = .000$, respectively. In this analysis the amplitudes at -18 ms and at -10 ms were significantly more positive compared to those at saccade onsets as well ($M_{-18ms} = 3.33$ μV, $M_{-10ms} = 18.62$ μV, and $M_{Sac} = -1.61$ μV).

Table 1. The Mean-amplitudes and *p*-values for the comparisons between saccade-onsets and -10 ms and between saccade-onsets and -18 ms in the F-datasets.

Dataset	***Mean-Amplitude at Saccade-Onset in μV***	***Mean-Amplitude at -10 ms in μV***	***p-value of the Amplitude-Comparison between Saccade-Onsets and -10 ms***	***Mean-Amplitude at -18 ms in μV***	***p-value of the Amplitude-Comparison between Saccade-Onsets and -18 ms***
F1	-6.31	15.04	.001	-1.52	.014
F2	-4.98	15.55	.001	-1.47	.013
F3	-4.32	16.65	.001	-0.60	.003
F4	-5.53	15.68	.001	-2.27	.002
F5	-2.89	17.59	< .001	0.71	.001
F6	-3.57	16.39	< .001	-0.54	.001
F7	-1.83	18.26	.001	1.33	.001
F8	-1.48	18.00	.001	1.77	< .001
F9	-0.54	19.66	.001	2.51	.001
F10	-0.40	19.7	.001	2.62	.001
Stets, Burt, & Reid (under review)	0.34	21.27	< .001	3.34	.001
Csibra, Tucker, Volein, & Johnson (2000)	0.80	3.20	< .002	2.10	< .02

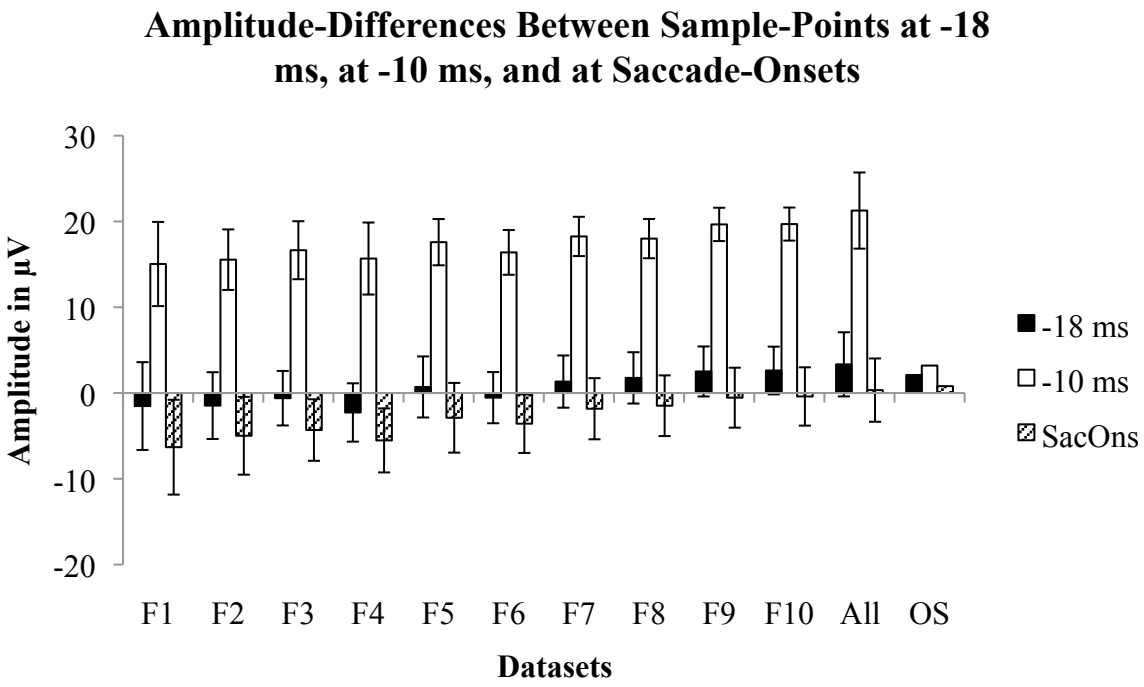

Figure 1. An illustration of the mean-amplitudes (± SE) calculated from the amplitudes from three sample-points around -18 ms before their respective saccade-onset, from around -10 ms before their respective saccade-onset, and from around the saccade-onset (n = 14). Depicted are the Spike-Potentials for our F-datasets from F1 through to F10 as well as the data presented in Stets, Burt, and Reid (under review; "All" for "all available artifact-free trials included") and the data from Csibra, Tucker, Volein, and Johnson (2000; "OS" for "original study"). All datasets feature significant differences between saccade-onsets and -10 ms ($p \leq .002$) as well as between saccade-onsets and -18 ms ($p \leq .02$).

Finally, the post-hoc one-way ANOVA revealed that there were no significant differences between the F-datasets either for the saccade onset or for the sample-point at -10 ms or for the sample-point at -18 ms.

Analysis 2 – The N290 and the P400

The N290

Halit, de Haan, and Johnson (2003) reported that the Species of the faces had a significant impact on the peak-amplitude of the N290 with human faces eliciting more negative responses than monkey faces – irrespective of orientation; M_{Human} = -8.5 μV, SD = 5.6; M_{Monkey} = 7.3 μV, SD = 5.4; $F(1, 24) = 15.57$, $p < .01$ (Halit, de Haan, & Johnson, 2003, p. 1184). Moreover, these authors reported an effect of Orientation from a post-hoc analysis with the N290 peaking more negatively for inverted as compared to upright faces – irrespective of species; $F(1, 24) = 7.21$, $p < .02$ (see p. 1185). These effects were not replicated either in Stets, Burt, and Reid (under review) where all available trials had been included per condition or in our analyses with the F-datasets. In the current analyses, we found significant interactions between Species and Orientations in datasets F2, F6, F7, F8, F9, and F10 ($p \leq .05$) as well as tendencies towards significant differences in datasets F1, F3, F4, and F5 ($p \leq .08$; see Table 2a for Mean-amplitudes, *SD*s and respective *p*-values; see Figure 2a).

In the post-hoc one-way ANOVAs with the Bonferroni-adjustment, we found significant differences between the F1-datasets for Human Upright. For the averages derived from the channel-group over the left hemisphere, we found F1 to be significantly different from F5 through to F10 as well as the dataset presented in Stets, Burt, and Reid (under review); $p \leq .027$. For the averages derived from the channel-group over the right hemisphere, we found F1 to be significantly different from F7, F8, F9, F10, and the datasets including all available artifact-free trials per condition; $p \leq .019$. No further differences were found for this condition between the datasets. For inverted human faces, no significant differences were found between any of the datasets. The same was true for the averages for upright monkey faces derived from

the channel-group over the left hemisphere. However, for the averages for Monkey Upright which were derived from the right channel-group, F1 was significantly different from F5 through to F10 as well as from the dataset presented in Stets, Burt, and Reid (under review); $p \leq .017$. The difference between F1 and F4 showed a tendency towards significance; $p = .069$. For the datasets for Monkey Inverted that were derived from the left channel-group, F1 showed to be significantly different from F3 through to F10 as well as from the average with all artifact-free trials included; $p \leq .013$. No other differences were found for these datasets. None of the F-datasets for inverted monkey faces derived from the right channel-group were significantly different from each other.

Finally, in an attempt to lend support to the data-analysis approach outlined in Analysis Two of Stets and Reid (2010), we performed additional analyses trying to include the data from the excluded participants when possible. For the conditions presenting upright and inverted human and monkey faces, this was only possible for the F1-dataset and only for two out of the three excluded infants. The 2x2x2 repeated measures ANOVA with 17 of 18 tested participants revealed that the results mirrored those of the ANOVA described above. No effect of species or orientation was found. However, the interaction between these two factors was significant; $F(1, 16) = 6.344$, $p = .023$. As seen in the analysis including 15 out of 18 tested participants, Human Upright elicited the most negative responses ($M = -21.088$ μV, $SD = 18.967$) followed by Monkey Inverted ($M = -16.050$ μV, $SD = 13.368$), Monkey Upright ($M = -10.476$ μV, $SD = 11.729$), and with Human Inverted eliciting the least negative responses ($M = -6.723$ μV, $SD = 16.769$).

Table 2a. The Mean-amplitudes and Standard Deviations for Human and Monkey faces presented in Upright and Inverted orientations in the F-datasets, in the data presented in Stets, Burt, and Reid (under review), and in Halit, de Haan, and Johnson (2003) as well as the *p*-values for the respective Species*Orientation-Interactions for the N290.

Dataset	***M-Amplitude for Human Upright in µV***	***SD***	***M-Amplitude for Human Inverted in µV***	***SD***	***M-Amplitude for Monkey Upright in µV***	***SD***	***M-Amplitude for Monkey Inverted in µV***	***SD***	***p-values for the Species*Orientation-Interaction***
F1	-21.56	25.87	-7.08	22.08	-12.41	14.99	-16.15	19.41	.057
F2	-13.43	22.6	-1.39	18.26	-4.63	12.70	-8.74	15.3	.038
F3	-6.76	16.10	-0.30	14.57	-1.11	13.94	-6.51	11.66	.064
F4	-6.5	14.12	0.5	13.26	-0.57	16.66	-2.03	12.63	.08
F5	-4.3	11.68	1.74	12.08	0.71	13.98	-1.94	10.26	.073
F6	-3.20	10.75	2.45	11.35	3.44	10.78	-1.65	10.23	.011
F7	-1.71	10.17	2.36	11.10	3.80	10.40	-0.42	9.26	.05
F8	-1.06	9.07	3.38	10.08	4.98	11.35	.032	10.10	.014
F9	-1.49	8.2	2.66	10.23	4.69	12.26	0.52	9.41	.013
F10	-0.89	8.41	2.64	9.83	3.90	11.98	0.24	9.19	.025
Stets, Burt, & Reid (under review)	-0.23	6.91	2.32	9.83	3.6	9.27	0.83	9.18	.088
Halit, de Haan, & Johnson (2003)	-6.5	6.4	10.7	6.6	-0.5	5.9	0.5	5.9	.005

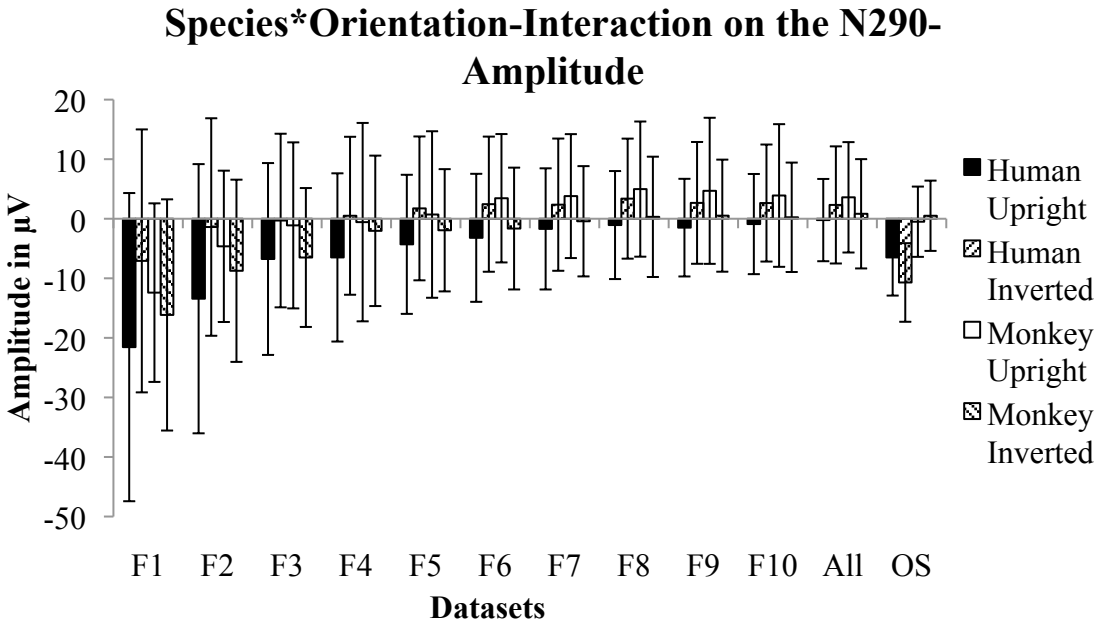

Figure 2a. An illustration of the mean-N290-peaks (± SD) for conditions Human Upright, Human Inverted, Monkey Upright, and Monkey Inverted (n = 15). Depicted are our F-datasets from F1 through to F10 as well as the data presented in Stets, Burt, and Reid (under review; "All" for "all available artifact-free trials included") and the data from Halit, de Haan, and Johnson (2003; "OS" for "original study"). Datasets F2, F6, F7, F8, F9, F10, and OS feature significant differences between the conditions, i.e., an interaction between species and orientation ($p \leq .05$). Datasets F1, F3, F4, F5, and All feature tendencies towards significant interactions ($p \leq .088$).

In Halit, de Haan, and Johnson's (2003) data, Species were found to have a significant effect on the latency of the N290 with humans yielding a later peak than monkeys; M_{Human} = 296, SD = 19; M_{Monkey} = 268, SD = 17; $F(1, 24) = 16.00$, $p < .005$ (p. 1185). However, Stets, Burt, and Reid (under review) did not replicate this effect when including all available artifact-free trials per condition into their analyses. In the

data-analyses in the present study, apart from dataset F6, which shows a significant effect of Species on the N290-latency ($F(1, 14) = 8.46$, $p = .011$), only F5 shows a trend in this direction ($F(1, 14) = 3.5$, $p = .082$). In both cases, humans elicited an earlier N290-peak compared to monkeys, which is the opposite effect to that reported in Halit, de Haan, and Johnson (2003). All other F-datasets showed no effect of species. However, as can be seen in Table 2b and Figure 2b below, our datasets including decreasing numbers of trials revealed more varied results. Instead of the latency being shorter for humans compared to monkeys in all F-datasets, the trend of directionality varies over the course of the test-session.

Additionally, apart from datasets F3 and F6, we found significant latency effects for Location in datasets F1, F2, F4, F7, F8, F9, and F10 ($p \leq .046$) with a strong tendency in dataset F5; $p = .059$ (see Table 2c and Figure 2c). In all cases, the latency of the N290-peak was shorter in the channel-group in the right hemisphere compared to the one in the left hemisphere. Csibra, Tucker, Volein, and Johnson (2003) also reported an effect of location in their data. However, these authors used a different EEG-system and had three channel-groups per hemisphere. Moreover, their location-effect was not reflected in the hemispheres but in the laterality of the different channel-groups within the hemispheres. Therefore, a meaningful comparison between the effects found in our data and those reported in the original study is not possible. The effect of location on the N290-latency could also be found in an additional analysis of the F1-dataset including data from two of the previously three excluded participants; $F(1, 16) = 7.368$, $p = .015$. Again, the N290-latency was shorter over the right channels ($M = 241.088$ ms, $SD = 32.727$) as compared to the left channel-group ($M = 265.108$ ms, $SD = 28.181$).

The post-hoc one-way ANOVAs with the respective corrections revealed that there were no significant differences between any of the F-datasets for any of the conditions or channel-groups for the latency of the N290.

Table 2b. The Mean-latencies and Standard Deviations for Human and Monkey faces in the F-datasets, in the data presented in Stets, Burt, and Reid (under review), and in Halit, de Haan, and Johnson (2003) as well as the *p*-values for the respective effects of Species for the N290.

Dataset	***M-Latency for Humans in ms***	***SD***	***M-Latency for Monkeys in ms***	***SD***	***p-value for the Species-Effects***
F1	261.17	69.34	250.93	70.86	*ns*
F2	260.22	66.96	256.55	75.11	*ns*
F3	251.17	63.88	250.67	80.58	*ns*
F4	248.4	68.75	255.68	79.46	*ns*
F5	242.87	67.21	263.13	80.57	.082
F6	234.72	69.84	263.13	78.75	.011
F7	233.62	70.50	251.1	77.70	*ns*
F8	242.88	71.73	246.8	77.49	*ns*
F9	249.85	74.46	247.35	77.23	*ns*
F10	252.22	68.88	243.93	72.54	*ns*
Stets, Burt, & Reid (under review)	253.77	76.57	256.62	74.76	*ns*
Halit, de Haan, & Johnson (2003)	296	19	268	17	.005

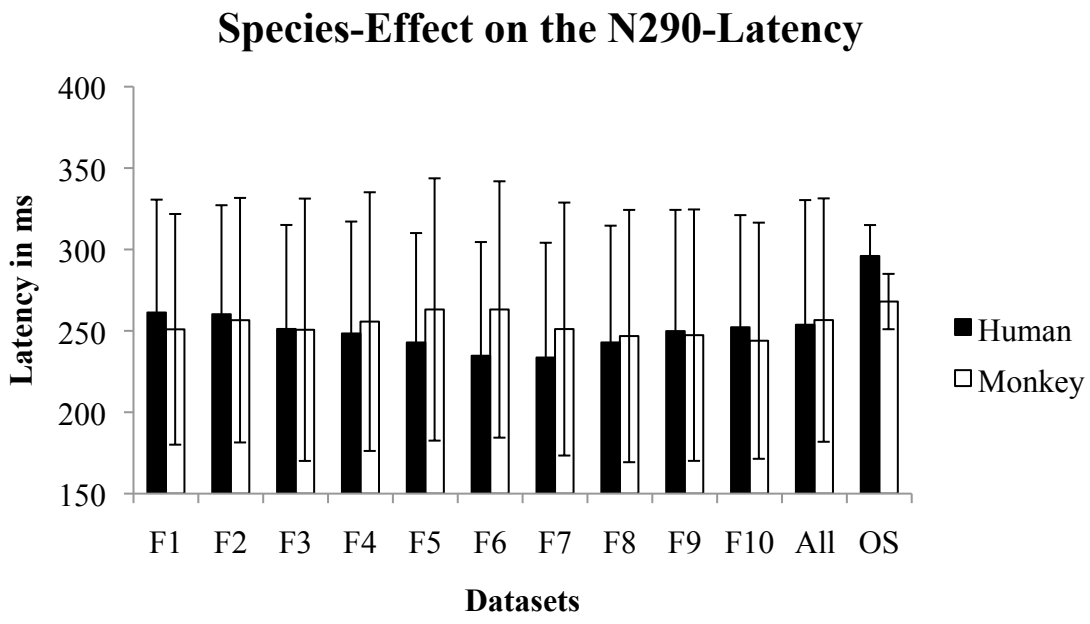

Figure 2b. An illustration of the mean-N290-latencies (± SD) for conditions human and monkey-faces (n = 15). Depicted are our F-datasets from F1 through to F10 as well as the data presented in Stets, Burt, and Reid (under review; "All" for "all available artifact-free trials included") and the data from Halit, de Haan, and Johnson (2003; "OS" for "original study"). Datasets F6 and OS feature significant differences between the species ($p \leq .011$). Dataset F5 features tendencies towards a significant effect ($p = .082$).

Table 2c. The Mean-latencies and Standard Deviations for the Left and Right channel-groups in the F-datasets, in the data presented in Stets, Burt, and Reid (under review), and in Halit, de Haan, and Johnson (2003) as well as the *p*-values for the respective effects of Location for the N290.

Dataset	***M-Latency for the Left Hemisphere in ms***	***SD***	***M-Latency for the Right Hemisphere in ms***	***SD***	***p-value for the Location-Effects***
F1	266.37	71.02	245.83	68.01	.041
F2	274.95	71.13	241.82	67.19	.003
F3	255.47	73.7	246.37	71.42	*ns*
F4	265.6	79.65	238.48	65.94	.007
F5	265.82	77.00	240.18	70.39	.059
F6	259.05	81.67	238.8	67.90	*ns*
F7	257.35	82.15	227.37	62.91	.018
F8	258.17	76.89	231.52	69.9	.002
F9	260.95	79.92	236.25	69.39	.046
F10	262.12	76.43	234.03	61.63	.009
Stets, Burt, & Reid (under review)	262.38	81.06	248	69.15	*ns*

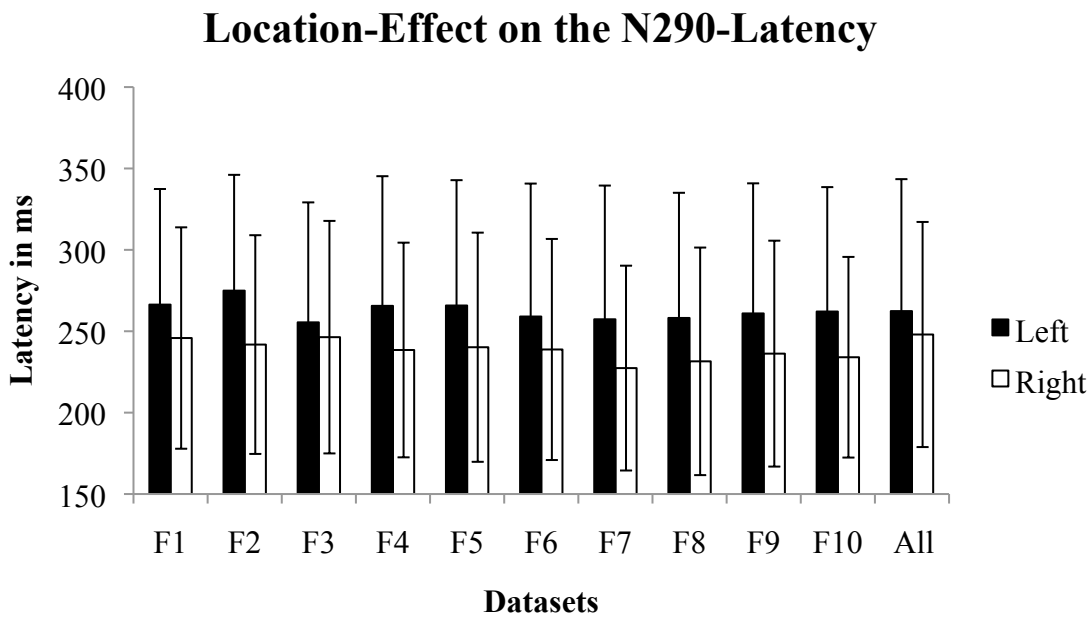

Figure 2c. An illustration of the mean-N290-latencies (± SD) for conditions the left and the right channel-group (n = 15). Depicted are our F-datasets from F1 through to F10 as well as the data presented in Stets, Burt, and Reid (under review; "All" for "all available artifact-free trials included"). Datasets F1, F2, F4, F7, F8, F9, and F10 feature significant differences between the hemispheres ($p \leq .046$). Dataset F5 features a tendency towards a significant difference ($p \leq .059$). (Due to differences in the EEG-systems used, the location-effect presented in Halit, de Haan, and Johnson [2003] is not comparable to those found in our F-datasets.)

The P400

Whereas Halit, de Haan, and Johnson (2003) did not find any effects on the peak-amplitude of the P400, Stets, Burt, and Reid (under review) reported a significant interaction between Species and Orientation; $F(1, 14) = 5.39$, $p = .036$ (p. 13). This interaction could also be found in our F-datasets. However, as can be seen in Table 3a

and in Figure 3a, the strength of the effects varies between the different datasets. In F1 and F2, a marginal tendency towards a species*orientation-interaction was found; $p \leq .089$. Monkey Upright elicited the most positive P400-amplitudes followed by Human Inverted, and Monkey Inverted. Human Upright yielded the least positive P400-amplitudes. Datasets F3, F4, and F5 did not show significant differences between the conditions. However, datasets F6 through to F10 feature significant differences in the P400-peaks; $p \leq .044$. Again, Monkey Upright had the most positive responses followed by Human Inverted, Monkey Inverted, and Human Upright eliciting the least positive P400-peaks (see Table 3a and Figure 3a). The above-mentioned marginal tendency towards a species*orientation-interaction in F1 was also found in the additional analyses including the first artifact-free trial available per condition from two of the three excluded participants; $F(1, 16) = 3.763$, $p = .070$. Again, Monkey Upright was most positive ($M = 55.985$ µV, $SD = 22.255$) followed by Human Inverted ($M = 51.048$ µV, $SD = 31.751$), Monkey Inverted ($M = 45.847$ µV, $SD = 21.858$), and Human Upright ($M = 38.902$ µV, $SD = 19.659$).

For the P400-amplitude, the post-hoc one-way ANOVAs with the Bonferroni-correction revealed that there were no significant differences between any of the averages derived for upright and inverted human faces in either of the two channel-groups. For Monkey Upright, F1 showed to be significantly different from F10 and the dataset presented in Stets, Burt, and Reid (under review) over the left hemisphere only; $p \leq .05$. No differences were found between datasets for upright monkey faces over the right hemisphere or for either channel-group for Monkey Inverted.

Table 3a. The Mean-amplitudes and Standard Deviations for Human and Monkey faces presented in Upright and Inverted orientations in the F-datasets, in the data presented in Stets, Burt, and Reid (under review), and in Halit, de Haan, and Johnson (2003) as well as the *p*-values for the respective Species*Orientation-Interactions for the P400.

Dataset	***M-Amplitude for Human Upright in µV***	***SD***	***M-Amplitude for Human Inverted in µV***	***SD***	***M-Amplitude for Monkey Upright in µV***	***SD***	***M-Amplitude for Monkey Inverted in µV***	***SD***	***p-values for the Species*Orientation-Interaction***
F1	39.74	31.81	52.28	38.37	57.79	32.07	46.57	30.65	.089
F2	34.73	28.57	44.32	27.83	48.08	26.70	39.15	20.74	.078
F3	33.68	23.23	40.27	24.09	44.04	26.11	37.07	20.52	*ns*
F4	32.13	22.5	38.71	23.17	38.74	26.35	37.12	15.63	*ns*
F5	31.83	16.93	38.55	22.61	39.56	25.68	35.83	14.63	*ns*
F6	31.51	17.31	39.05	20.46	39.74	20.25	35.80	14.55	.044
F7	30.34	16.49	39.11	19.65	39.5	19.45	34.56	13.67	.034
F8	29.94	15.31	38.44	17.46	38.66	19.85	34.05	13.98	.02
F9	28.93	14.09	36.45	16.59	38.62	23.99	32.28	13.12	.035
F10	27.89	14.81	35.27	16.46	37.08	23.36	30.89	11.64	.03
Stets, Burt, & Reid (under review)	25.11	12.75	32.71	17.08	33.33	19.32	29.43	11.47	.036
Halit, de Haan, & Johnson (2003)	15.6	8	16	8	16.5	7.9	18.2	7.9	*ns*

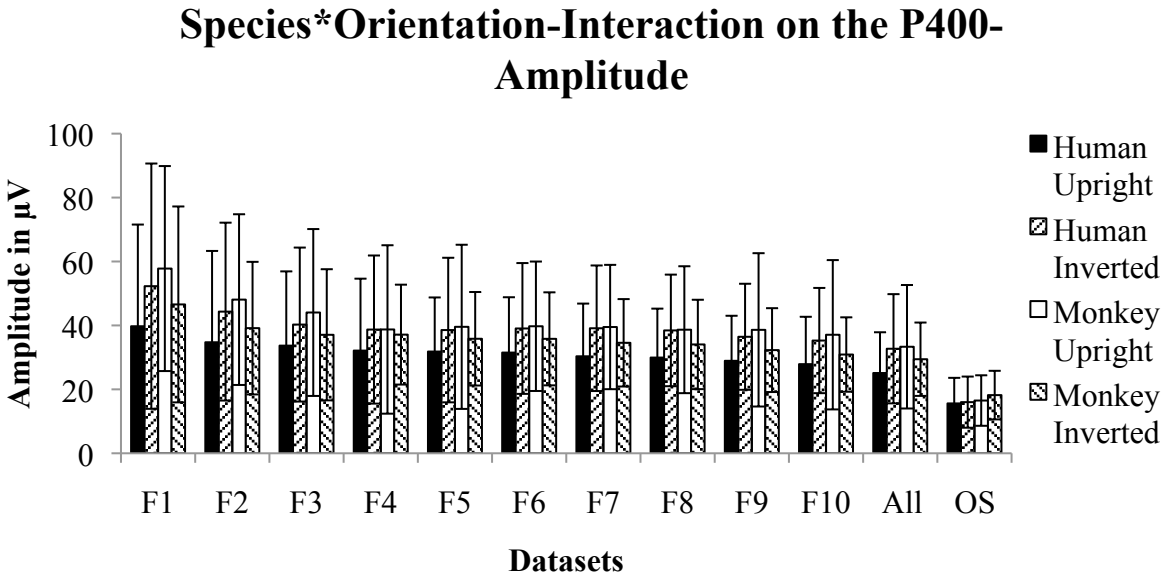

Figure 3a. An illustration of the mean-P400-peaks (± SD) for conditions Human Upright, Human Inverted, Monkey Upright, and Monkey Inverted (n = 15). Depicted are our F-datasets from F1 through to F10 as well as the data presented in Stets, Burt, and Reid (under review; "All" for "all available artifact-free trials included") and the data from Halit, de Haan, and Johnson (2003; "OS" for "original study"). Datasets F6, F7, F8, F9, F10, and All feature significant differences between the conditions, i.e., an interaction between species and orientation ($p \leq .044$). Datasets F1, and F2 feature tendencies towards significant interactions ($p \leq .089$).

Whereas Halit, de Haan, and Johnson (2003) found a significant effect of Species on the latency of the P400, neither Stets, Burt, and Reid's (under review) analysis nor our F-datasets showed significant differences due to species. The effect of Orientation reported in Stets, Burt, and Reid (under review), which showed inverted faces to be processed faster than upright faces ($F(1, 14) = 6.22$, $p = .026$; p. 15), was

replicated in our dataset including the first usable trial per condition. F1 also showed upright faces to be processed more slowly than inverted ones – irrespective of species; $M_{Upright}$ = 496.02 ms, SD = 68.07; $M_{Inverted}$ = 467.02 ms, SD = 68.23; $F(1, 14) = 5.75$, $p = .031$. This effect was not replicated when data from two of the previously three excluded infants were added to the analysis of the F1-dataset.

Additionally, our F-datasets showed significant interactions between Species and Orientation on the latency of the P400. In F2, F3, and F4, Human Upright had the longest P400-latency followed by Monkey Inverted, Monkey Upright, and Human Inverted; $p \leq .043$. In F6, Monkey Inverted had the longest P400-latency followed by Human Upright, Monkey Upright, and Human Inverted; $F(1, 14) = 7.46$, $p = .016$. In F7, Human Upright had the longest latency followed by Monkey Inverted, Human Inverted, and Monkey Upright; $F(1, 14) = 8.11$, $p = .013$. Additionally, F5 showed a tendency to a significant interaction between species and orientation; $F(1, 14) = 3.93$, $p =. 068$. All other datasets did not show significant interactions (see Table 3b and Figure 3b).

The post-hoc one-way ANOVAs with the Bonferroni-correction revealed that there were no differences between the F-datasets for the P400-latency for either upright or inverted human faces or for upright or inverted monkey faces. This was true for both hemispheres.

Table 3b. The Mean-latencies and Standard Deviations for Human and Monkey faces presented in Upright and Inverted orientations in the F-datasets, in the data presented in Stets, Burt, and Reid (under review), and in Halit, de Haan, and Johnson (2003) as well as the *p*-values for the respective Species*Orientation-Interactions for the P400.

Dataset	***M-Amplitude for Human Upright in µV***	***SD***	***M-Amplitude for Human Inverted in µV***	***SD***	***M-Amplitude for Monkey Upright in µV***	***SD***	***M-Amplitude for Monkey Inverted in µV***	***SD***	***p-values for the Species*Orientation-Interaction***
F1	500.43	66.91	472.53	68.02	491.6	70.07	461.5	69.15	*ns*
F2	517.57	60.86	476.03	69.19	484.83	71.92	508.13	64.09	.043
F3	505.2	61.22	465.93	74.89	475.87	67.08	500.1	59.38	.022
F4	500.67	70.5	470.93	77.33	474.6	68.88	498.73	59.31	.018
F5	497.37	72.42	464.63	74.31	477.63	72.34	489.93	64.57	.068
F6	487.47	74.74	462.5	73.53	474.07	76.68	496.57	69.60	.016
F7	490.23	71.10	463.47	75.08	461.03	69.31	483.03	67.07	.013
F8	481.8	76.83	468.77	73.91	470.47	78.34	472.37	66.38	*ns*
F9	486.67	76.11	460	75.37	473.67	82.23	464.17	63.36	*ns*
F10	482.6	72.51	456.17	70.86	472.93	78.36	469.03	65.37	*ns*
Stets, Burt, & Reid (under review)	483.57	73.34	446.67	73.04	465.93	75.17	456.6	63.42	*ns*
Halit, de Haan, & Johnson (2003)	395	23	406	19	419	25	419	24	*ns*

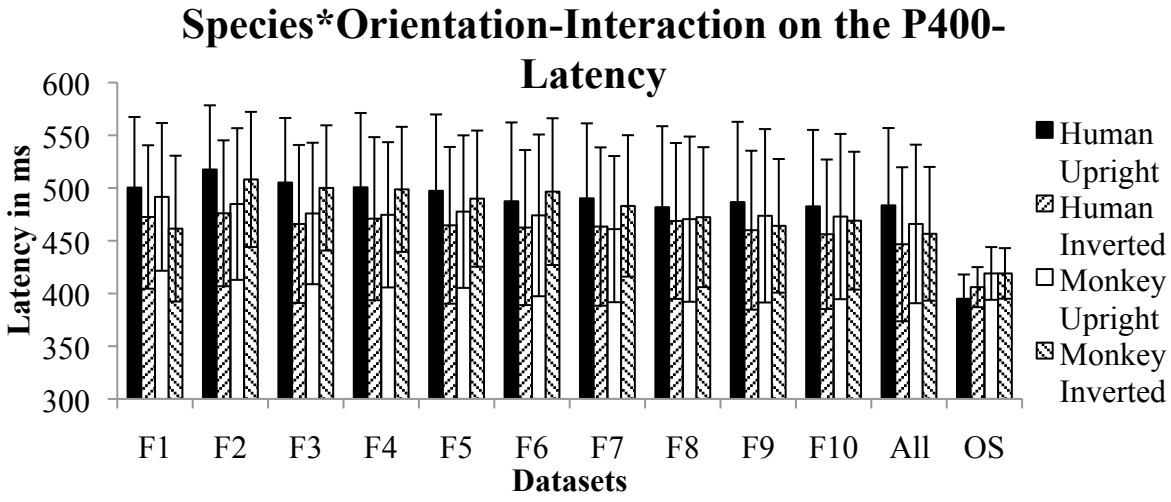

Figure 3b. An illustration of the mean-P400-latencies (± SD) for conditions Human Upright, Human Inverted, Monkey Upright, and Monkey Inverted (n = 15). Depicted are our F-datasets from F1 through to F10 as well as the data presented in Stets, Burt, and Reid (under review; "All" for "all available artifact-free trials included") and the data from Halit, de Haan, and Johnson (2003; "OS" for "original study"). Datasets F2, F3, F4, F6, and F7 feature significant differences between the conditions, i.e., an interaction between species and orientation ($p \leq .043$). Dataset F5 features a tendency towards a significant interaction ($p \leq .068$).

Analysis 3 – The Nc and the PSW

The Nc

Hoehl, Reid, Mooney, and Striano (2008) reported a significant effect of Condition on the Nc-peak with the stimuli of a person looking away from a toy eliciting a more negative amplitude compared to the Toward-stimuli, between 400 and 600 ms after stimulus-onset. Stets and Reid (2011) re-analyzed the data from these 4-month-olds by including only the first three, five, seven, eight, nine, or 10 usable trials into their averages. They found a significant difference between the conditions when including the first seven trials, however, it was the Toward-condition that elicited the most negative responses. Moreover, their Figure 2 (Stets & Reid, 2011, p. 562) provides an insight into the underlying processes with regard to the formation of the Nc over the course of an experimental session. Stets, Burt, and Reid (under review) did not find significant differences in the Nc-peak for their group of 12-month-olds for these two conditions between 350 and 500 ms after stimulus-onset. In our re-analysis of the data initially presented in Stets, Burt, and Reid (under review), we found a significant effect of Condition in F1; $F(1, 13) = 5.11$, $p = .042$. As was found in the data from the first seven trials in Stets and Reid (2011), the condition in which the actor was looking towards the toy yielded the more negative Nc-peak compared to the amplitude derived from the first presentation of the Away-condition; $M_{Toward} = -43.80$ μV, $SE = 6.47$; $M_{Away} = -10.67$ μV, $SE = 15.9$. The remaining F-datasets did not show significant differences between the conditions. Moreover, when adding data from three of four previously excluded participants to the analysis of F1, the significant difference between the condition was replaced by a marginally tendency towards a significant difference with the directed gaze condition eliciting the more negative responses ($M = -32.450$ μV, $SE = 8.448$) compared to the averted gaze condition ($M = -8.152$ μV, $SE =$

13.221); $F(1, 16) = 3.345$, $p = .086$. However, in general we still found the same pattern of directionality-modulation that was reported in Stets and Reid (2011). Specifically, an initial trend for the Toward-condition to elicit the more negative responses progressively shifted towards the Away-condition eliciting the more negative responses as the experimental session continued. In Stets and Reid (2011), such a change in effect-directionality was found between the averages including nine and 10 trials. Our sample was eight months older than Hoehl et al.'s (2008) infants which might lead to the assumption that this change should happen in a dataset comprising fewer trials as the infants would have had eight months more exposure to these two social situations. However, the characteristics of the data in terms of which of the conditions would elicit the more negative responses did not occur substantially earlier in our data when compared with the data from Hoehl et al.'s (2008) 4-month-olds. Here, the Toward-condition elicited the more negative responses from F1 through to F8 and only from F9 onwards did the Away-condition show the more negative responses (see Table 4). No other significant differences for the Nc-amplitudes or -latencies were found in the other F-datasets. Also, no effects were found for the Nc-latency when including data from three of the four previously excluded participants into the analysis of the F1-dataset.

With respect to the amplitude of the Nc, the post-hoc one-way ANOVAs with the above-mentioned corrections revealed differences between F1 and the remaining datasets for the Toward-condition all three channel-groups under investigation. For the left-frontal channels, F1 was significantly different from F3 through to F10 and the data presented in Stets, Burt, and Reid (under review); $p \leq .002$. For fronto-central channels, there was a mild tendency towards a significant difference between F1 and F3; $p = .082$. Moreover, F1 was significantly different from F4 through to F10 as well

as from the data including all artifact-free trials for this channel-group; $p \leq .006$. For the channel-group over the right hemisphere, F1 for the Toward-condition was significantly different from all other datasets; $p \leq .004$. No significant differences were found between any of the datasets originating from the Away-condition. This was true for all three channel-groups. Finally, with respect to the Nc-latency, the post-hoc one-way ANOVAs revealed that there were no significant differences between any of the datasets for either condition or channel-group.

Table 4. The Mean-amplitudes and Standard Errors of the Negative component for conditions Toward and Away in the F-datasets, in the data presented in Stets, Burt, and Reid (under review), and in Hoehl, Reid, Mooney, and Striano (2008) as well as the p-values for the respective effects of Condition.

Dataset	***M-Amplitude for Toward in μV***	***SE***	***M-Amplitude for Away in μV***	***SE***	***p-value for the Condition-Effects***
F1	-43.80	6.47	-10.67	15.9	.042
F2	-20.94	5.83	-10.93	8.88	*ns*
F3	-13.64	4.74	-6.25	6.56	*ns*
F4	-8.56	4.86	-7.56	5.15	*ns*
F5	-2.88	5.86	-0.35	4.82	*ns*
F6	-2.29	5.64	-0.67	3.57	*ns*
F7	-1.57	5.06	-1.32	3.86	*ns*
F8	-2.16	4.39	-1.96	3.94	*ns*
F9	-2.13	4.17	-2.71	3.65	*ns*
F10	-2.87	4.12	-3.31	3.45	*ns*
Stets, Burt, & Reid (under review)	-3.04	4.2	-3.99	2.91	*ns*
Hoehl, Reid, Mooney, & Striano (2008)	-14.3	2.26	-19.6	2	.044

The PSW

As was the case for the Nc, Hoehl, Reid, Mooney, and Striano (2008) reported significant differences between the Conditions for the mean-amplitudes of the Positive Slow-Wave. In the time-window from 700 to 1000 ms after stimulus-onset, the stimuli presenting the directed gaze elicited the more positive responses compared to those presenting the averted gaze-condition. Neither Stets, Burt, and Reid's (under review) analysis nor our current re-analysis replicated this effect. However, the directionality of the PSW, i.e., which of the conditions elicited the more negative responses relative to the other one, undergoes a similar evolution as that which has been seen in the Nc.

Stets, Burt, and Reid (under review) found a tendency towards an effect of location on the PSW-latency that had not been found in the original study. The most positive responses were reported in the right-frontal channel-group compared to the left-frontal and the fronto-central channel-groups. An effect of Location was also found in our datasets from F5 through to F9 ($p \leq .046$) with a tendency towards significance in F10, ($F(1, 13) = 3.27$, $p = .054$). As can be seen in Table 5 and in Figure 4, in these datasets, it was the fronto-central channel-group that elicited the most positive responses followed by the right-frontal and the left-frontal channels. No effects were found in a re-analysis of F1 with data included from three of four previously excluded participants.

The post-hoc one-way ANOVAs with the Bonferroni-adjustment revealed that there were no significant differences between the datasets for either of the two conditions or for either of the three channel-groups under investigation.

TABLE 5.

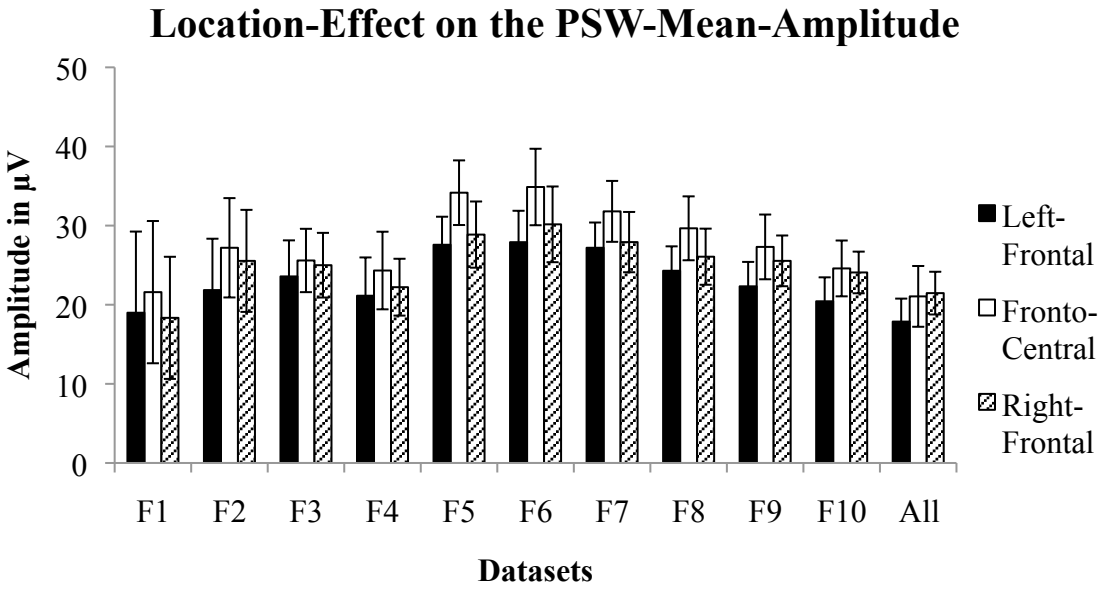

Figure 4. An illustration of the mean-PSW-amplitude (± SE) for conditions the left-frontal, the fronto-central, and the right-frontal channel-groups (n = 14). Depicted are our F-datasets from F1 through to F10 as well as the data presented in Stets, Burt, and Reid (under review; "All" for "all available artifact-free trials included"). Datasets F5, F6, F7, F8, and F9 feature significant differences between the locations ($p \leq .046$). Datasets F10 and All feature tendencies towards significant differences ($p \leq .061$).

Discussion

The theory behind averaging over as many trials as possible for an experimental condition is that noise originating from artifacts and high variability in the individual responses will be averaged out of the signal (e.g., Luck, 2005; Picton et al., 2000). The resultant ERP-average is thought to present a clearer image of the neural responses within the sample to a certain kind of stimulus. However, capturing high numbers of artifact-free trials per condition is not without issues as it can be hard to achieve in infant populations (see Marshall, Reeb, & Fox, 2009). Not only do large numbers of infants need to be tested, with only a subsample contributing enough trials to the analysis, but, additionally, the basic assumption that a component will remain stable over all trials while the level of noise in the individual trials will differ may not be well-understood in developmental populations. This theory that noise will be averaged out of a signal when an increased number of artifact-free trials are included into the analysis is commonly referred to as the signal-to-noise ratio. The perfect signal-to-noise ratio of 10.0 (i.e., a pure ERP-component without any noise) mentioned in Luck (2005) could only be achieved if hundreds of useable trials would be available for all of the conditions. Even aiming for a much less optimal signal-to-noise ratio of 0.4 (i.e., an average consisting of fewer trials with the same component-amplitude but with much noise left in the signal) would require infants to provide a minimum of 30 artifact-free trials per condition as calculated by Stets and Reid (2011; see also Luck, 2005). It seems apparent that current guidelines for conducting ERP-studies in general are clearly designed for adult populations and are not easily applicable to infant samples (Reid, 2012). Moreover, as has been shown in Stets and Reid (2011), averaging 10 or more artifact-free trials per condition together can potentially hide

changes in the components over time which may be of interest in terms of infant cognitive or perceptual processing. On the one hand, the Nc-amplitudes for the conditions presenting a directed or averted gaze changed slightly and response-variability was averaged out of the signal when the number of trials in the average increased. On the other hand, the peak-amplitudes showing which condition elicited the more negative responses changed dramatically depending on the number of trials Stets and Reid (2011) included in their analyses. The Nc is related to attention-allocation and it has been previously shown that infants' attention to stimuli can change over the course of an experimental session (e.g., Nikkel, & Karrer, 1994). However, so far it had remained unclear if other components, which were not thought to be related to attention, would undergo similar or different changes as a function of the number of trials included in the average.

With respect to amplitude, we found varied results for the different components in our F-datasets. Comparing upright and inverted human and monkey-faces, the noise and the amplitude-differences for both the N290 and the P400 are averaged out between individual trials when more trials are added to the averages. The relative amplitudes, however, remain stable in their directionalities over the different datasets, i.e., which condition elicits the more negative or positive responses compared to the others. This is in accordance with the assumptions for the signal-to-noise ratio described above (Luck, 2005; Picton et al., 2000). For the Nc and as a trend for the PSW, we found a pattern which approximated the amplitudes and we also replicated the change between conditions dependent on the number of trials for the Nc as reported in Stets and Reid (2011). Our findings for the Nc indicate that components related specifically to attention-processes are more prone to be modulated by the duration of a test-session and the number of trials included in the statistical analyses.

In comparison, a perceptual component such as the SP, which is related to planning a saccade towards a stimulus (Csibra et al., 2000), and both the N290 and the P400, which are related to the processing of faces as opposed to objects and other types of stimuli (e.g., de Haan, Pascalis, & Johnson, 2002; de Haan, & Nelson, 1999), appear to be less affected by these factors. Therefore, investigations of attention-related components might be more likely to find high component-variability and changes in the ERPs over the course of the experimental session.

So far, the latencies of ERP-components have been seen as indicators for the ease of processing a certain kind of stimulus. However, breaking up the data into datasets with different numbers of trials and analyzing the latencies based on these different averages, reveals a much more complex picture. Despite only showing a significant difference in processing-speed in F6 and a tendency in F5, the N290-latency cycles back and forth between the response to human faces peaking earlier and monkey faces eliciting an earlier N290-peak. The questions arise why the human condition shows later responses than the monkey condition considering that, based on the increased exposure to this species, human faces should be easier to process, and why this cycling appears. It is entirely plausible that, at the beginning of a test-session, human faces take longer to show a N290 because more details are processed in human than in monkey face stimuli. However, after seeing human faces repeatedly (both as stimuli in this condition and other experimental conditions) the change-over to the monkey-faces eliciting the longer N290-latencies might indicate that the infants study the monkey-faces more closely. As has been shown in previous studies, older infants tend to process human and monkey-faces in a more "adult-like" fashion and are not as good at distinguishing between faces of different monkeys as younger infants (see Pascalis, de Haan, & Nelson, 2002). This short interval in which the N290-latencies

are longer for monkeys than for humans could be an indication of the infants attempting to discriminate between individual monkeys before going back to doing so for the human faces. This might also be an explanation for the latency-cycling seen in Figure 3b in which Human Upright and Monkey Inverted are taking turns in showing the longest P400-latency. Moreover, whereas the directionality for the P400-latencies for human faces stays the same in all F-datasets (i.e., Human Upright peaking later than Human Inverted), the latencies for Monkey Upright are longer than those for Monkey Inverted in F1. Following this, they are then shortest in F2 through to F8, and then they are longer again from F9 onwards. Considering the high variability of information being directed to the infants (i.e., eight experimental conditions) and, as suggested by the low attrition rate reported in Stets, Burt, and Reid (under review), the infants were most likely in a more alert state compared to when experiencing a standard ERP-paradigm with the more typical two experimental conditions. Following the same kind of reasoning as presented above, infants might try to extract meaningful information from the stimuli depicting inverted monkeys. However, realizing that no valuable information is delivered by these stimuli, they lose interest and no longer attempt to process these stimuli in the later stages of the experimental session. This is parallel to the claims made in Stets and Reid (2011), who proposed that Hoehl et al.'s (2008) 4-month-olds would start to attend more to the Away-stimuli later in the test-session as they might infer a communicational intent from the person repeatedly looking away from the toy towards an unknown location in space.

The latency-cycling described above is not present in all of the components under investigation. For instance, when looking at a component-feature such as channel-location which is less related to the contents of the stimuli and more to the cognitive processes themselves, latency-cycling cannot be found in all instances. As

seen in Figure 2c, irrespective of the condition presented, the channel-group in the left hemisphere always elicited the later N290-peaks compared to those in the right hemisphere. However, for the P400 and for the Nc, results are mixed and slight shifts or opposing latency-patterns between two neighboring datasets can be found. This indicates that location-effects for latencies might not always be very informative for all possible ERP-components. Moreover, for such analyses the EEG-system used to record the data will need to feature enough electrodes to be able to draw meaningful conclusions from potential effects. Consequently, results for the same infant ERP-component coming from different laboratories using different EEG-systems with varying numbers of electrodes might not always be comparable and should be treated with care. This is especially the case, when the component in question is related to attention and criteria for inclusion (i.e., the minimal criterion) are differing between labs.

Finally, it needs to be noted that our datasets feature comparatively large amplitudes and standard deviations – especially so, when only few trials form the average for a condition. We acknowledge that averages derived from two trials for instance are likely to still contain a certain amount of noise. Moreover, such variability in the data is more likely to impact on a study's results if the number of participants is relatively small. Therefore, it may be necessary to include more participants in an analysis, when averages are calculated based on fewer trials. Following the logic outlined in Analysis Two in Stets and Reid (2010), the variability originating from individual participants should help to clean out the average overall – as additional trials would in an average derived from larger volumes of trials per condition. As described above we attempted to perform additional analyses along those lines. However, the number of participants that could be added in the additional analyses

here may not be sufficient to make claims with respect to this. As mentioned above, further studies applying similar strategies of data-analysis are needed to refine and validate this approach.

Conclusion

Stets and Reid (2011) were the first to systematically analyze and report the changes that an infant ERP-component undergoes depending on the number of artifact-free trials included in an average. Here, as well as replicating this finding for the Nc-component, we attempted to assess whether other ERP-components, both attention-related and -unrelated, showed changes based on the number of trials included. We found that the amplitudes of attention-related ERP-components seem to undergo similar processes displaying directionality-shifts in terms of which condition elicited the most negative Nc-peak or the most positive PSW-response. For the other components, the SP, the N290, and the P400, amplitudes do not change in directionality but might undergo shifts in terms of polarity with increased number of trials included in the average. Moreover, apart from the SP and the PSW, all components showed changes in latency effects. In these particular cases, the latencies cycled back and forth so that different conditions took shorter or longer to show a specific peak across the different F-datasets. Based on our findings that latencies may undergo regular changes as a function of the number of trials included in the average (i.e., cycling), new insights into cognitive processes may be gained from investigating infant ERP-data in this manner. Looking into latencies can likely provide us with additional clues as to why and how infants process certain kinds of stimuli in a specific way. We were able to analyze the latencies of three different (infant) ERP-components. However, it would be useful for developmental scientists to apply this manner of analyzing ERP-amplitudes and -latencies to a wider range of ERP-components. This is particularly true as infant populations are very different from adult populations. Whereas adults can be instructed to behave in very specific ways during

an EEG-session, infants cannot. Additionally, while older children as well as clinical populations can potentially be instructed, they might not comply at all times. Considering such minor challenges as well as the bigger issues discussed above, the need for more improved strategies and tools for data-collection and -analysis is given in the wider research community. We believe that, through further refinement and development of the strategies, the method of data-analysis outlined here and, earlier, in Stets and Reid (2011) combined with the method for stimulus-presentation outlined in Stets, Burt, and Reid (under review) can be helpful tools for researchers in general and especially for those investigating more challenging populations such as infants or clinical samples. Therefore, we would like to invite the wider research community from a wider range of disciplines to adopt of this new method of data-collection in order to more fully understand its usefulness for a broader research base.

References

Courchesne, E., Ganz, L., & Norcia, A.M. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535-543.

Csibra, G., Davis, G., Spratling, M.W., & Johnson, M.H. (2000). Gamma oscillations and object processing in the infant brain. *Science*, *290*, 1582-1585.

Csibra, G., Tucker, L.A., & Johnson, M.H. (1998). Neural correlates of saccade planning in infants: A high-density ERP study. *International Journal of Psychophysiology*, *29*, 201-215.

Csibra, G., Tucker, L.A., Volein, A., & Johnson, M.H. (2000). Cortical development and saccade planning: the ontogeny of the spike potential. *NeuroReport*, *11*(57), 1069-1073.

de Haan, M., & Nelson, C.A. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Developmental Psychology*, *35*(4), 1113-1121.

de Haan, M., Pascalis, O., & Johnson, M.H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199-209.

Grossmann, T., Oberecker, R., Koch, S.P., & Friederici, A.D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, *65*, 852-858.

Halit, H., de Haan, M., & Johnson, M.H. (2003). Cortical specialization for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*, *19*, 1180-1193.

He, C., Hotson, L., & Trainor, L.J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*, *19*(5), 878-892.

Hoehl, S., Reid, V., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science*, *11*(1), 10-16.

Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, *37*, 187-209.

Little, V.M., Thomas, D.G., & Letterman, M.R. (1999). Single-trial analyses of developmental trends in infant auditory event-related potentials. *Developmental Neuropsychology*, *16*(3), 455-478.

Luck, S.J. (2005). Ten simple rules for designing ERP experiments. In: Handy, T.C. (ed.), *Event-related potentials: a method handbook* (pp. 17-32). Cambridge, MA: MIT Press.

Marshall, P.J., Reeb, B.C., & Fox, N.A. (2009). Electrophysiological responses to auditory novelty in temperamentally different 9-month-old infants. *Developmental Science*, *12*(4), 568-582.

Nikkel, L., & Karrer, R. (1994). Differential effects of experience on the ERP and behavior of 6-month-old infants: trends during repeated stimulus presentation. *Developmental Neuropsychology*, *10*, 1-11.

Pascalis, O., de Haan, M., & Nelson, C.A. (2002). Is face processing species-specific during the first year of life?. *Science*, *296*, 1321-1323.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, Jr., R., … Taylor, M.J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37*, 127-152.

Polich, J. (1986). Attention, probability, and task demands as determinants of P300 latency from auditory stimuli. *Electroencephalography and clinical Neurophysiology*, *63*, 251-259.

Reid, V.M. (2012). Infant EEG comes of age. *Developmental Neuropsychology, 37*(3), 185-186.

Richards, J.E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants an ERP study. *Developmental Science*, *6*(3). 312-328.

Rugg, M.D., & Coles, M.G.H. (1995). The ERP and cognitive psychology: conceptual issues. In M.D. Rugg, & M.G.H. Coles (eds.), *Electrophysiology of mind – Event-related brain potentials and cognition. Oxford psychology series, No. 25* (pp. 27-39). New York, NY: Oxford University Press.

Sheskin, D.J. (2011). *Handbook of parametric and nonparametric statistical procedures, 5th ed.*. CRC Press.

Stets, M., Burt, M., & Reid, V.M. (under review). Infants need more variety – Increased data acquisition with reduced participant attrition in infant ERP-studies. *FRONTIERS in Developmental Psychology*.

Stets, M., & Reid, V.M. (2011). Infant ERP amplitudes change over the course of an experimental session: Implications for cognitive processes and methodology. *Brain and Development*, *33*(7), 558-568.

Stets, M., Stahl, D., & Reid, V.M. (2012). A meta-analysis investigating factors underlying attrition rates in infant ERP studies. *Developmental Neuropsychology*, *37*(3), 226-252.

Webb, S.J., Long, J.D., & Nelson, C.A. (2005). A longitudinal investigation of visual event-related potentials in the first year of life. *Developmental Science*, *8*, 605-616.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A.D. (2004). Discrimination of word stress in early infant perception: electrophysiological evidence. *Cognitive Brain Research*, *18*, 149-161.

Chapter 6

General Discussion

In all stages of an infant ERP-study, i.e., study-design, data-acquisition, and, especially data-analysis, common methodological strategies leave large room for interpretation by individual researchers. Both Luck (2005) and Picton and colleagues (2000) reported what they believed to be good guidelines for analysing infant ERP-data. These included the steps for the preparation of the raw EEG-data for further analyses of the extracted ERPs, how many trials would be sufficient or insufficient to form a relatively clean ERP-average, as well as descriptions of which statistical tests are commonly conducted in this field of research with such data. Moreover, Picton et al. (2000) repeatedly mention how all these steps of analysis should be reported in a research article resulting from an ERP-study in order to allow other researchers in the field to understand all the parameters of a study and, potentially, be able to easily replicate the respective study. However, as has been mentioned in Chapter 1, many of these guidelines presented in methodological literature concerning ERP-studies (e.g., Luck, 2005; Picton et al., 2000) are well-suited for work with adult participants and less applicable with infant populations. Attempts have been made to describe the methodological challenges faced by ERP-researchers working with infants and possible solutions for these challenges have been suggested (e.g., DeBoer, Scott, & Nelson, 2007; Hoehl, & Wahl, 2012). However, with respect to the large variability between the strategies of data-analysis used in the different labs and even between individual researchers in the same laboratory, no stable solution has to date been found. Therefore, with the studies presented in Chapters 2 to 5, I attempted to address the aforementioned challenges appearing in the different stages of an experiment and provide potential solutions for them. In the following, I will first discuss the results and methodological implications of the separate studies individually. In addition, I will

discuss the results and draw conclusions for future infant ERP-studies as a result of issues and implications arising from multiple studies presented in this thesis.

General Overview of Available Approaches for Study-Design and Data-Collection

The methodological issues related to infant ERP-studies are numerous and cover every single stage of a research study (see Chapter 1). Starting with the preparation of the stimulus-presentation, developmental psychologists have multiple options to design a study in such a way as to ensure that infants will not lose interest in the stimuli too early. For example, Csibra, Tucker, and Johnson (1998) had their visual target-stimuli accompanied by various background sounds. Richards (2003) embedded his target-stimuli in snippets of a children's programme. During a test-session, again, there are various possibilities to tailor the session to an infant's needs and intervene when necessary. As a first measure, and in order to avoid too much intervention by the experimenter, parents are often instructed to try and redirect their infant's attention back to the screen (e.g., Nelson, & Salapatek, 1986). Pang, Edmonds, Desjardins, Khan, Trainor, and Taylor (1998) had an assistant play with hand-puppets during the presentation of their auditory stimuli when a participant showed signs of boredom. Secondly, the infants will commonly be given a break from the stimulus-presentation when they show signs of distress. In such a case, the presentation could be interrupted by simply turning the screen off and entertaining the infant with blowing bubbles (L. Tucker, personal communication, October 18, 2011). Another strategy, used in Parise, Handl, and Striano (2010), is to replace the target-stimuli on the monitor by an animated stimulus accompanied by background sounds to maintain or regain the infants' focus (see also Clifford, Franklin, Davies, & Holmes, 2009). Alternatively,

researchers such as Mason, Braddick, Wattam-Bell, and Atkinson (2001) and Scott and Nelson (2006) opted for closely monitoring the infants' behaviour and only presenting the stimuli when participants were looking at the screen. Additionally, if trials are judged to not be attended towards by the infant, despite attempting to display them when the infant is looking in the direction of the monitor, some researchers also delete these trials during the test-session and may repeat the trial when attention has been re-established (e.g., Nelson, Henschel, & Collins, 1993; Webb, Long, & Nelson, 2005). Eventually, when the infants have been given breaks repeatedly and their attention cannot be re-established, the session will be discontinued (e.g., Hoehl, & Striano, 2008; Jeschonek, Marinovic, Hoehl, Elsner, & Pauen, 2010). Csibra, Tucker, and Johnson (2001, p. 162) as well as Csibra, Tucker, Volein, and Johnson (2000, p. 1070) state that this was the case when the session "had to be interrupted a second time".

Approaching Data-Analysis from a Different Angle

The first study (Stets, & Reid, 2011, see Chapter 2) was a re-analysis of ERP-data collected from a group of 17 four-month-old infants and published in Hoehl, Reid, Mooney, and Striano (2008). Presenting their participants with images of an adult either looking toward or away from a toy, these researchers had analysed, among others, the minimal peak-amplitude and the latency of the Negative component (Nc) between 400 and 600 ms after the onset of the stimuli. Following traditional ways of data-analysis, Hoehl et al. (2008) had included only the data from those four-month-olds from whom they could obtain minimally 10 artefact-free trials per condition for the averaging process. However, there are substantial differences between the

recommendations in the methodological literature and what is and can be practised by the researchers with respect to how many trials should be and are included minimally in the statistical analyses to obtain meaningful averages for the respective experimental conditions. In an attempt to address this issue, the rationale of the re-analysis presented in Stets and Reid (2011; see Chapter 2) was to investigate if there was an optimal minimum number of analysable trials that is required in an infant ERP-study. Moreover, the study addressed the question if there was also a time-factor impacting on the components when forming the averages. Therefore, in a first step, Stets and Reid (2011) re-analysed Hoehl et al.'s (2008) data for the peak-amplitudes of the Nc including only the first three, five, seven, eight, nine, and 10 artefact-free trials per condition when calculating the ERP-averages for the two conditions.

Comparing the negative peak-amplitudes between the conditions in which the adult had been directing her gaze toward the toy and the one in which she had averted her gaze, Hoehl and colleagues (2008) had found more negative responses between 400 and 600 ms after stimulus-onset toward the latter condition. This was interpreted as an indication that more of the infants' attention had been directed to the Away-stimuli compared to the Toward-ones. Including varying numbers of usable trials in the analyses, Stets and Reid (2011) could show that the infants' attention towards the stimuli from the two experimental conditions can shift from one to the other. When comparing the peak-amplitudes including three and seven artefact-free trials per condition, the directed gaze condition showed to elicit the more negative responses compared to the averted gaze condition. In case of the analysis including seven trials the difference between the conditions reached a significant level. In the analysis including the first three trials, the difference showed a tendency towards significance. As can be seen in Figure 2 in Stets and Reid (2011; p. 562), the Toward-condition still

yields the more negative Nc-peaks in the dataset including the first nine usable trials in the average and only when the first 10 trials are forming the ERP-average, the directionality of the difference between the conditions is changing towards what had been reported in Hoehl et al. (2008). Consequently, as Hoehl and colleagues (2008) conducted their analyses in a conventional manner and did not compare ERPs for fewer than 10 artefact-free trials per condition, the shift in attention from one condition to the other could not be detected in the original study. However, finding such underlying forces working on an ERP-component could help us better understand infant cognition in general and, as the Nc has repeatedly been related to attention (e.g., Nikkel, & Karrer, 1994; Richards, 2003), attention-related processes in particular. Stets and Reid (2011) conclude that, counter to previous assumptions that ERPs are stable entities which remain unchanged in response to a certain class of stimuli, the relevant processes are much more dynamic. However, the question remained if this is only true for attention-related components or for a wider range of ERP-components. Moreover, it seems apparent that if attention can shift between conditions in a group of infants as young as four months of age, more varied results could be expected in populations representing other developmental stages when applying this in-depth way of data-analysis as compared to a traditional approach.

Based on the finding that the amplitude comparison for the first three usable trials of the 17 original four-month-olds revealed a tendency to significance, Stets and Reid (2011) took another unconventional step. It was assumed that the majority of those infants who had been excluded from Hoehl et al.'s (2008) final analysis due to not providing minimally 10 artefact-free trials per condition, would at least have three analysable trials per condition. Conducting the same statistical tests on the peak-amplitudes for the first three trials from 24 infants (i.e., the data from the original 17

infants plus the data from seven participants who had been excluded from Hoehl et al.'s [2008] final analysis) revealed a statistical significance between the conditions. As had been found in the first analysis, the stimuli presenting the directed gaze condition – the condition which has been described as being socially more normal and familiar for the infants at such a young age (see Stets, & Reid, 2011) – produced a more negative Nc-amplitude compared to the stimuli presenting the person looking away from the toy. This finding lends further support to the claims made based on the findings in Stets and Reid's (2011) first analysis. Moreover, considering the additional issue of typically high exclusion rates in infant ERP-studies due to infants failing to meet the minimal criterion, this analysis illustrates that, often, large amounts of usable and informative data are discarded in a traditional analysis.

Finally, it was attempted to establish if the shift in attention between the conditions resulted from the number of trials that had been included in the analyses only or if there was an additional time-factor involved. Therefore, the analyses were re-run for the datasets presented in Analysis 1 (see Chapter 2) with the infants' data being weighted according to the amount of time that had passed from the beginning of an individual's test-session up to the point when the respective number of artefact-free trials had been collected for both conditions. As can be seen in Figure 4 in Stets and Reid (2011, p. 566; see Chapter 2), the dynamics that the Negative component is subject to in this third analysis were the same as those seen in Analysis 1. Again, the Toward-condition elicited the more negative Nc-responses with this difference between the conditions being significant when the first seven artefact-free trials had formed the average. Moreover, again, there was a change towards the Away-condition eliciting the more negative responses when 10 trials were included in the averaging process. Stets and Reid (2011) interpreted this as an indication that the amount of time

that passed until a specific number of usable trials are collected for all experimental conditions is already incorporated into the analysis through the inclusion of the specific numbers of trials. However, in order to gain clarity on this, it would be desirable to apply such analyses to other datasets from slightly older infants for instance or from studies that did not use stimuli including faces. Additionally, such analyses could be applied to data originating from an auditory paradigm.

Apart from investigating the possible alterations of the Nc based on how many trials are included in the average, the study presented in Chapter 2 (Stets, & Reid, 2011) aimed to test a strategy for data-analysis that may be more suitable for the use with ERP-data originating from infant populations. Apparently, the currently used strategies for data-analysis are typically derived from those used with adult-data. However, as has been mentioned in Chapter 1, this causes difficulties for researchers working with infants as these two populations are very different from each other in many respects (e.g., Hoehl, & Wahl, 2012; Reid, & Geangu, 2008). Therefore, the assumption that downscaling a measure that is typically used with adults will be sufficient to make it applicable for infants, may not be easily met under all circumstances. Moreover, applying this novel approach to the analysis of the data previously published in Hoehl, Reid, Mooney, and Striano (2008), it could be shown that current assumptions about the quality of trials from the earliest stages of a test-session do not hold. Following Luck's (2005) reasoning, it should not have been possible to derive clean enough averages (i.e., with sufficient signal-to-noise ratio) from the first three artefact-free trials from a group of four-month-olds to draw meaningful implications from them (see also Picton et al., 2000). However, significant differences between the conditions could be found and conclusions about the cortical responses involved in processing these stimuli could be drawn (see Analysis 2 in

Chapter 2). Therefore, this approach to data-analysis may be a valuable tool when researchers, who seem to be naturally limited in the overall number of trials to be obtained from an infant, want to make more efficient use of the volume of data that they can collect. Moreover, the results of Analysis 3 (Stets, & Reid, 2011) do not only lend further support to the idea that an average derived from fewer artefact-free trials from the beginning of a test-session can be as meaningful as an average which is based on a larger volume of trials. Analysis 3 also shows that there does not seem to be a qualitative difference between these earlier analysable trials compared to those collected later in a test-session. This means that the re-analysis did not only show effects on a methodological level. Also on a theoretical level, new insights could be gained into the attentional processes in this young group of participants. As has been mentioned earlier, applying this approach to data-analysis to other ERP-components could help to uncover further details about underlying processes that have so far been unknown and unexpected.

An Alternative Approach to Improving Data-Analysis

Another novel approach to preserve and include trials from larger numbers of participants is described in Stahl, Parise, Hoehl, and Striano (2010). These authors analyzed EEG-data recorded from 6-month-old infants in response to emotional facial expressions including different eye-gaze directions; i.e., a neutral face with infant-directed eye-gaze, a neutral face with an averted gaze, an angry face with an infant-directed gaze-direction, and an angry face with an averted gaze. In order to be able to maximize the number of usable infants, Stahl et al. (2010) weighted the participants' data based on the number of artefact-free trials they contributed for the conditions. The

mean-number of trials that could be obtained for each of the four conditions was 6.5 (*SD* = 4.3, ranging between 1 and 23; Stahl et al., 2010, p. 309). As was practised in the Analysis Two of the study presented Chapter 2 and later on in the study presented in Chapter 5 as well, the traditional strategies of setting a minimal criterion which has to be met by all infants to be included in a study's final analysis and to average over as many trials as possible in order to obtain a cleaner ERP-average are directly violated. However, despite this violation meaningful results could still be obtained and reported. Therefore, this alternative approach to improving current strategies for data-analysis can be seen as a valuable addition to the toolset. However, as is the case with the approach outlined in this thesis, further studies investigating different ERP-components and populations are needed to validate its usefulness. Moreover, it is less likely that such the strategy outlined in Stahl et al. (2010) will be able to uncover previously unknown changes in information-processing as have been found using the approach presented in this thesis.

Attrition as an Indicator for Issues of Study-Design?

The first study showed that the analysis of infant ERP-data can be customised to the needs of developmental populations and that currently, data containing meaningful information is often removed prior to final analysis. Therefore, the question arose what the rejection of such large volumes of trials and, ultimately, participants may be based upon. An important factor with respect to participant-exclusion is a study's minimal criterion for the number of trials contributed to the average for each condition. The decision on how high or low to set the minimal criterion for participant inclusion in a specific study often seems to be made in an

arbitrary fashion and can vary greatly between different labs and researchers. For instance, in the majority of infant ERP-studies using a standard paradigm (i.e., the stimuli for the individual conditions are presented with equal frequency), the number of artefact-free trials that are required per condition for the infants to be included in the study's final analysis ranges between 10 and 15 (e.g., 12 in Männel, & Friederici, 2009; 15 in Snyder, & Keil, 2008; 10 in von Koss Torkildsen Sannerud, Syversen, Thormodsen, Simonsen, Moen et al., 2006). However, there are a number of studies using standard paradigms that set their minimal criterion higher – at 20 or 30 analysable trials per condition (e.g., 30 in Csibra, Davis, Spratling, & Johnson, 2000; 20 in Simos, & Molfese, 1997). Additionally, there are also standard-paradigm studies using a lower minimal criterion (e.g., seven in de Haan, & Nelson, 1997; five in Marshall, & Shipley, 2009). Fundamental questions remain with respect to the basis of the decision for a particular criteria, and what impact it may have on the study's attrition rate and the quality of obtained data. If the minimal criterion is set comparatively high, many participants may have to be excluded from the study's analysis and, consequently, the attrition rate for this respective study may increase substantially. In the meta-analysis presented in Chapter 3, I attempted to provide some answers to this question. Among others, such as the number of experimental conditions presented, the duration of the inter-stimulus interval or the nature of the presented stimuli, studies' minimum criteria were investigated with respect to how they would affect the attrition rate of an infant ERP-study.

The meta-analysis of 181 experimental groups revealed large variations between study-features such as the duration of a single trial (ranging from 800 ms to 6150 ms), the total number of infants tested for a study or experimental group (ranging from 9 to 175), or the number of trials included in the averaging processes per

condition (ranging from 8 to 1186; see Table 2 in Chapter 3; Stets, Stahl, & Reid, 2012, p. 234). Moreover, as can be seen in Table 1 in Chapter 3 (Stets, Stahl, & Reid, 2012, p. 230), there seems to be a slight asymmetry in the age-groups researchers tend to use for infant ERP-studies. Fifty-seven out of 314 experimental groups investigated six-month-old infants. Another 31 and 36 groups consist of three- and four-month-olds, respectively. One issue is whether this frequent usage of these age-groups is mainly due to theoretical reasons (i.e., the developmental changes that may have taken place at these stages) or due to practical reasons. A commonly held belief among infant ERP-researchers is that older infants are harder to test than younger age-groups (e.g., Hoehl, & Wahl, 2012). This would entail that attrition rates should be much higher in studies using an older age-group compared to studies using younger infants. However, Stets, Stahl, and Reid (2012) had collapsed their analyses across age-groups (ranging from newborns to 24-month-olds) as no differences were found between the attrition rates of the different ages in this respect. In fact, the study which experienced the highest attrition rate of all assessed studies, with a rate of 83.8% (data used from 12 infants out of 74 who had been tested) reported in Chapter 3 (Stets, Stahl, & Reid, 2012) had been conducted with six-month-old infants (Nelson, & Collins, 1991) – an age-group which is commonly thought to be easier to test. Moreover, also the study with the second highest attrition rate found among the 314 experimental groups under investigation (80.6%; data used from 34 infants out of 175 who had been tested; de Haan, Pascalis, & Johnson, 2002) had been conducted with the same age-group. The authors of these papers had explained these high attrition rates as being a consequence of the strict inclusion criteria that had been applied. In the case of de Haan, Pascalis, and Johnson's (2002) study, infants had to provide minimally 15 artefact-free trials out of 80 presented trials for each of their four conditions, equalling fewer than 20% of

trials. The mean-number of analysable trials which had been included per condition from the 34 participants was 25.5 equalling a rate of 31.9%. As a comparison, in Csibra, Tucker, and Johnson (1998), a study which also tested six-month-olds, the attrition rate was 68.4% with the minimum criterion set to 10 out of 60 trials for each of the four conditions. Again, fewer than 20% of the trials were required to contribute to the final analysis with the mean of included trials per condition equalling 17.1 (a ratio of 28.5%). This indicates that the researchers' expectations about how many analysable trials to obtain from an infant per condition are comparatively low. This is also true for other age-groups as illustrated by the eight artefact-free trials for each of the two experimental conditions that Csibra, Tucker, Volein, and Johnson (2000) required from their group of 12-month-olds (attrition rate = 61.5%) or the 10 out of 60 trials required from a group of 13- to 17-month-olds in Mills, Coffey-Corina, and Neville (1997; attrition rate = 28.2%). Considering such comparatively low inclusion criteria, it seems apparent that none of these or other infant ERP-studies will meet the high standards for signal-to-noise ratio that are recommended in methodological literature (e.g., Luck, 2005; Picton et al., 2000).

Among the 34 study-features investigated in Stets, Stahl, and Reid (2012), many would traditionally be expected to cause infants to be removed from a study's final analysis. For instance, as has already been stated in the introduction, it is frequently recommended to keep the cognitive load low for infants (e.g., DeBoer, Scott, & Nelson, 2007). However, the number of conditions presented or the duration of the stimuli, the inter-stimulus intervals, or the trials in total did not show to have an impact on the reported attrition rates (Stets, Stahl, & Reid, 2012; see Chapter 3). Counter to expectation, there were only three factors that did show to affect attrition: (a) the age of the research article itself (in years), (b) how many infants had originally

been tested for the study or experimental group, and (c) the nature of the stimuli. Additionally, as the number of infants who had been tested in total is inherently related to a study's attrition rate, the meta-regression had to be run a second time without including this particular factor in the model. In the first meta-regression which included the total number of infants tested, attrition was affected by the additional factor of whether the stimuli were of an animated nature. However, in the second meta-regression without the total number of infants included, this was not the case. Stets, Stahl, and Reid (2012; see Chapter 3) suggested that the length of the animated stimuli may be the reason why attrition was higher in those studies compared to studies using static stimuli. This would indicate that the cost of the animation in terms of the time taken to display it actually offsets the increase in interest due to the animated nature of the stimuli. Even though it might be true that habituation to the stimuli takes longer, the length of each trial counteracts this benefit, with the result that attrition rates are higher for this class of stimuli.

In general, it may seem surprising that only so few study-features were shown to show an effect on attrition rates – especially so, as they are not the ones that would typically be expected (e.g., DeBoer, Scott, & Nelson, 2007; Hoehl, & Wahl, 2012). The finding that older publications featured higher attrition rates than newer ones (Stets, Stahl, & Reid, 2012), seems to indicate that EEG-systems have become more "participant friendly" and/or that researchers have learned from experience or been trained by expert trainers, and have refined their strategies for collecting EEG-/ERP-data from infant participants. However, conducting a test-session still remains a highly individual process as an infant's behaviour is often hard to predict – sometimes even for the caregivers themselves – with respect to feeding- and sleeping-times, for instance. Therefore, as has been mentioned earlier, it is essential that an experimenter

is well-trained for the work with infants and can draw on an extensive set of tools and strategies in order to be prepared for any eventuality (see also Hoehl, & Wahl, 2012).

Apart from the animated nature of the stimuli, the choice of modality was also found to influence a study's eventual attrition rate (Stets, Stahl, & Reid, 2012). In Chapter 3 it is argued that the significantly lower attrition rates of studies presenting infants with purely auditory stimuli compared to those showing visual stimuli only, were due to the much lower likelihood for movement-artefacts in the data as many of the experimental groups (especially newborns) were allowed to sleep during the test-session (e.g., deRegnier, Wewerka, Georgieff, Mattia, & Nelson, 2002; Weber, Hahne, Friedrich, & Friederici, 2004). Alternatively, in the case of slightly older infants, the participants were often entertained by a silent puppeteer while listening to the target-stimuli (e.g., Brannon, Libertus, Meck, & Woldorff, 2008; Friederici, Friedrich, & Christophe, 2007). Therefore, in such a case, the alertness and the focus on the stimuli required from the infants is different from a study in which the participants must follow a visual stimulus-presentation. Consequently, the effort that needs to be taken to attend to the visual stimuli will likely be higher for an extended period of time. Consequently, the likelihood for boredom and/or habituation to occur will increase the longer the presentation lasts (see also Stets, & Reid, 2011; Chapter 2).

Related to this is also Stets, Stahl, and Reid's (2012) claim that infant ERP-studies may suffer from a selection bias with respect to which are or are not published. Figure 2 in Chapter 3 illustrates that the likelihood for a study with a large standard error (i.e., much variation in the data) and a high attrition rate to be published is practically non-existent. This is indicated by the upper right corner of the illustration being empty. Instead, the majority of experimental groups and studies which are published, tend to have a smaller standard error and an attrition rate which lies in a

range roughly between 35 and 75% of the expected norm (Stets, Stahl, & Reid, 2012). The issue that arises here can occur on either of two levels. On the one hand, it may occur at the point when a researcher decides if a study should be prepared for publication or not based on the question if a study's results are meaningful enough or not, or if they are of high enough quality. On the other hand, a large part of rejected manuscripts may be it may attributable to reviewers and editors. However, in that case, the additional question arises what factors exactly editors and/or reviewers base their decisions upon. Fanelli (2010) argues that researchers may be biased in their perception of what are meaningful and publishable results as they are subject to increasing pressure to publish from grant-agencies and potential employers. In this instance, 'meaningful results' are those that are positive and no null results. Furthermore, Ioannidis (2006) remarks that it is highly unlikely that any given research facility will only be producing positive results and hardly any null findings However, as Fanelli (2010, p. 1) states further, "[…] all results are equally relevant to science, as long as they have been produced by sound logic and methods […]". This seems to indicate that the question raised earlier – if the cause for a publication bias may originate from researchers or from publishers – is not easy to answer. Statements such as the above indicate that researchers may attempt to adjust their manuscripts in such a way that they better fit the assumed preferences of editors and publishers. Consequently, it is likely that there is an interaction between what is offered by the researchers and what is demanded by the publishers and that this interaction is negatively impacting on researchers' freedom to publish their results, i.e., a publication bias. Moreover, as null findings will most likely not be reported, there is the potential for studies on a certain research question to be unwittingly attempted by several labs (see also Ioannidis, 2006). This ultimately means that the progress of research in a

specific field may be prevented instead of promoted and that efforts in terms of money and time may be repeatedly spent on unsuccessfully addressing the same research question. It seems apparent that such a trend may not be beneficial for the wider research community in the long run and that measures should be taken to prevent the situation from further increasing in intensity. It is interesting to note that journals such as *PLOS One* now explicitly state that reviewers cannot reject papers based on the novelty or importance of results. They are given strict instructions to comment only on the rigour of the methods and the quality of results reportage, and to point out issues that may not be addressed in the introduction and the discussion section of these manuscripts.

Another major finding of the meta-analysis reported in Chapter 3 refers to how infant ERP-studies are reported with respect to details provided in the respective method-sections. As seen in Stets, Stahl, and Reid (2012), with a mean-attrition rate of 47.3%, on average, only every second infant tested for a study entered the study's final analysis. Moreover, almost the same ratio was true for the inclusion of the studies or experimental groups in the meta-analysis. The reason for excluding almost 50% of the experimental groups was that the respective research articles did not provide information on attrition. One of the recommendations mentioned in Picton et al. (2000) was for such important information to be mentioned and explained in a detailed fashion. In order for other researchers to be able to evaluate the reported data it is crucial to know how many infants had to be excluded for what reason. The same holds true for study-features such as the number of stimuli presented per condition or the number of different stimuli used per condition. However, during the preparations for the meta-analysis, it became apparent that such basic information about the paradigm had often either been omitted or mentioned in insufficient depth of detail in many

studies. Therefore, one of the fundamental features of a research study, its replicability, is limited which could potentially restrict the wider research community from gaining new insights about earlier findings. Consequently, it may be necessary for editors and publishers to monitor the description of all necessary information with respect to the paradigm more closely and encourage authors to make provide information to the readership. One possibility would be the more extensive use of online-resources. It may be helpful to offer researchers the possibility to make such detailed descriptions available in the form of online-supplements if page-space is restricted in the printed journal. Alternatively, it may be useful to strengthen the sector of online journals. This way the issue of page-space would be resolved automatically. Additionally, the possibilities for presenting study-methodologies and results are much more varied as, for instance, short animations may be uploaded to illustrate the paradigm (see Angell, & Smith, 1998). However, this requires the research community to become more open to (peer-reviewed) online journals and to accept their credibility more than is currently the case.

Study-Design as a Key to Attrition and Infant Attention

Based on the findings of the meta-analysis and the meta-regressions (Stets, Stahl, & Reid, 2012), the third study presented in this thesis (see Chapter 4; Stets, Burt, & Reid, under review) attempted to address the issue of paradigm-design mentioned in the Introduction. As the number of presented conditions did not show to affect attrition in infant ERP-studies, and had also been recommended in Hoehl, and Wahl (2012), it was hypothesized that a stimulus-presentation with more variable stimuli may be an option to keep infants interested for a longer period. One option

would have been to design a study with the typical two experimental conditions and a large variety of different stimuli for each of the conditions. Jeschonek, Marinovic, Hoehl, Elsner, and Pauen (2010) conducted an ERP-study with such a design with a group of eight-month-olds. They presented them with 59 different images of animals and with another 59 images depicting different furniture-items with none of the animals or objects being presented more than once. Despite the variability of the stimuli, the authors reported to have used the data from 14 out of 55 infants yielding an attrition rate of 74.6%. Of these infants, 51.2% had to be excluded for reasons of fussiness or inattentiveness and the remaining 48.8% had to be excluded because of too many movement-artefacts in the data (Jeschonek et al., 2010). Therefore, the variability of the stimuli does not seem to have affected the way the study progressed. It is entirely plausible that despite having different images, the stimuli sets themselves, being only two categories (animal or furniture), served to habituate the infants, thereby inducing the high attrition rate.

Another option is to follow the typical setup of having a more restricted number of stimuli per condition but including a larger variety of experimental conditions. The ERP-study presented in Chapter 4 (Stets, Burt, & Reid, under review) opted for this second possibility and presented a total of 18 infants aged between 12 and 13 months with eight experimental conditions which came from three distinct research articles (Csibra, Tucker, Volein, & Johnson, 2000; Halit, de Haan, & Johnson, 2003; Hoehl, Reid, Mooney, & Striano, 2008). As has been discussed in Chapter 4, there were both similarities and differences between the results of the original studies and those found reported in Stets, Burt, and Reid (under review). Some of these variations between the results could be explained with the setup of the stimuli as six of the eight conditions included faces in the stimuli (i.e., upright and inverted human and

monkey faces based on Halit, de Haan, & Johnson [2003] and females looking toward or away from a toy based on Hoehl, Reid, Mooney, & Striano [2008]). However, the social contexts of the faces were different from each other. The upright and inverted faces did not convey any information other than the species and the orientation of the faces. The eye-gaze direction of the faces in the directed- and averted-gaze conditions, however, created situations in which the depicted female directs the infants' attention to a particular side of the screen. In order to prevent such interferences between the conditions from arising, Stets, Burt, and Reid (under review) recommend taking care in the stimulus-selection process. Despite addressing different cognitive questions, the stimuli of the conditions used in study 3 may well have been too similar to each other on a perceptual level (see Chapter 4). Therefore, it may be helpful to also include conditions involving auditory stimuli, for instance. This would further increase the variability of the stimulus-presentation and may add to the positive effect this paradigm had on the infants' willingness to produce adequate data during an EEG test-session.

However, the most important findings in terms of methodology were the increased amount of data obtained from the infants and the decrease in attrition. As mentioned earlier, many infant ERP-studies have a minimal criterion for inclusion of 10 artefact-free trials per experimental condition. The inclusion criteria used in the three original studies were different from each other and ranged from eight in Csibra, Tucker, Volein, and Johnson (2000) to 25 in Halit, de Haan, and Johnson (2003). Due to the high number of experimental conditions, the number of possible stimulus-presentations for one of the conditions was naturally limited in Stets, Burt, and Reid (under review). Additionally, based on the general assumption that infant attention is very limited (e.g., Hoehl, & Wahl, 2012; Thierry, 2005), the expectation would be that

the infants would not provide enough artefact-free trials per condition for a traditional data-analysis. However, despite the challenge that this paradigm posed for the infants, minimally 10 usable trials could be collected from the majority of the 18 originally tested one-year-olds for all of the analyses and conditions (see Stets, Burt, & Reid, under review, in Chapter 4). This means that, instead of saturating the infants with too much variation, the large variety in the stimuli resulting from the high number of experimental conditions prevented the participants from becoming easily habituated to, and bored by, the stimulus-presentation. Based on this finding, it seems indicative that infant attention is more malleable than generally expected. On the one hand, attention can shift between conditions when stimuli from two conditions are repeated again and again as seen in Chapter 2 (Stets, & Reid, 2011). On the other hand, the study presented in Chapter 4 shows that infant attention can be prolonged when multiple different conditions are presented in an intermixed fashion in the same test-session.

Additionally, while all of the original studies had attrition rates of 61.5% (Csibra, Tucker, Volein, & Johnson, 2000), 69.4% (Halit, de Haan, & Johnson, 2003), and 73.4% (Hoehl, Reid, Mooney, & Striano, 2008), Stets, Burt, and Reid (under review) had to exclude only three and four infants from the analyses of the respective conditions. Such low attrition rates of 16.7% and 22.2% are challenging for currently held assumptions (e.g., DeBoer, Scott, & Nelson, 2007). Moreover, as all of the stimuli were of a purely visual nature, these low attrition rates are also in contrast with the findings from the meta-analysis presented in Chapter 3. According to Stets, Stahl, and Reid (2012), studies with purely visual stimulation had significantly higher attrition rates than those with purely auditory stimulation. However, most of these studies and experimental groups using visual stimuli did not present more than two or four conditions. The only two experimental groups among the 314 originally collected ones

(Stets, Stahl, & Reid, 2012) which had been exposed to six experimental conditions, listened to very short auditory stimuli (25 different stimuli presented once for each of the six conditions with a mean-duration of 320 ms; Ferguson, & Molfese, 2007). Moreover, this study had to be excluded from the more detailed meta-analysis as no information on attrition had been provided. Therefore, it is unfortunately not possible to make statements about what impact this paradigm had on the infants' likelihood to be excluded from the study's final analysis. However, in order to further explore the methodological possibilities that a multi-condition study such as this could help to create, it is logical for follow-up work to also include two or four conditions presenting auditory stimuli. It could be argued that this would make the stimulus-presentation even more variable, "interesting" and engaging for the participants. According to Bahrick and Lickliter's (2000) intersensory redundancy hypothesis this should be especially true when multi-modal stimuli are included (i.e., stimuli which have both an auditory and a visual component) as this will maximise variation in the presented stimuli.

The question remains if the same effects of multiplying the obtained data and, simultaneously, decreasing attrition rates, would also appear in a different paradigm in terms of stimulus-presentation. Instead of showing the eight conditions in the intermixed fashion used in Stets, Burt, and Reid (under review), these stimuli could also have been presented in a block-design within one test-session. However, it is not clear what the impact of such a dramatic change in the paradigm would have on the infants' attention. Furthermore, changing the paradigm in such a way may also have consequences for the potential interference that was observed between the conditions seen in Study 3. They may either be as likely to occur in a blocked stimulus-presentation or the likelihood for interference to occur may be slightly decreased.

Therefore, presenting infants with stimuli for more than one study in a blocked design may or may not be as effective in terms of maintaining and, potentially, prolonging infant attention (see Chapter 4). It would be helpful to examine such dynamics through conducting another study using the same set of stimuli and the same age-group but presenting the stimuli in a blocked design and comparing the results with respect to both methodological and theoretical issues.

Further Insights into Infant Cognition through Novel Methodological Approaches

The final study presented in Chapter 5 (Stets, Burt, & Reid, in preparation) attempted to combine the findings gained in Studies 1 and 3. Therefore, the approach to data-analysis presented in the first study (Stets, & Reid, 2011; Chapter 2) was applied to the data collected in the third study (Stets, Burt, & Reid, under review; Chapter 4). As mentioned earlier, Stets, and Reid (2011) did not answer whether the reported changes in the cortical responses would only be found in attention-related components such as the Nc or if they would apply to a wider range of ERP-components. From the results presented in Chapter 5 it seems apparent that changes in the cortical responses towards conditions over the course of an experimental session indeed seem to be restricted to attentional processes (Stets, Burt, & Reid, in preparation). Components related to face-processing and -recognition such as the N290 and the P400 as well as saccade-planning seem to remain largely unaffected by the duration of a test-session as indicated by the comparatively stable directions of the amplitudes for the individual conditions illustrated in the respective graphs. For instance, there is a drift from all amplitudes being negative for the N290 in response to the upright and inverted human and monkey-faces from F1 to F3 to three in four

conditions eliciting positive N290-responses when all available artefact-free trials were included in the average. However, the directionality of which condition yields the more negative responses only changes between human inverted faces and upright monkey-faces whereas it stays the same for upright human and inverted monkey faces (see Figure 2a in Chapter 5). This shows that the amplitudes of infant ERP-components are generally very likely to undergo changes of a certain nature when the number of included trials is varied. However, what kind of change will occur – i.e., a shift in polarity as seen in the N290 or a reversal in directionality as seen in the Nc – depends on the cognitive process/es required by the respective experimental stimuli. Given that this approach to data-analysis was applied to further data-sets involving other infant components which have not been investigated here, it may be possible to establish categories of ERP-components based on the dynamics seen in the detailed analysis. For instance, a component which has previously not been related to attention processes may show a similar pattern in its amplitudes as seen for the Nc and may, therefore, show to involve attentional processes as well. Additionally, finding common underlying structures for components with similar functions would also be helpful when potential infant precursors for adult components are investigated (e.g., Halit, de Haan, & Johnson, 2003).

Interestingly, while the amplitudes of the N290 and the P400 were unchanged with respect to directionality, the latencies of these two components were shown to be affected as a function of the number of trials involved in the creation of the average (see Figures 2b and 3b in Stets, Burt, & Reid, in preparation). In case of the N290, the latency oscillated between human faces showing longer latencies and monkey-faces having the later N290 peak-amplitude. For the P400, both the species and the orientation of the depicted faces impacted on how long it took for the positive peak to

appear in the different datasets. Whereas no differences had been found for the P400-latencies in Halit, de Haan, and Johnson (2003), Stets, Burt, and Reid (in preparation) could report species*orientation-interactions in several of the F-datasets with the latencies cycling from human upright faces having a later P400-latency to inverted monkey-faces showing a later peak and back to the P400 peaking later for the upright human faces again. The difference between inverted human faces and upright monkey-faces remained relatively stable. It was argued in Chapter 5 that such previously unexpected changes in latency-directionality may result from the increased amount of different information with which the infants were confronted during the test-session. Causing the one-year-olds to retain an alert state of mind, the variability of the stimuli may also have contributed to the infants' depth of processing of different aspects of the images that were presented. Parallel to Stets, and Reid's (2011) argument, the infants may initially pay increased attention to the stimuli depicting something that is more familiar to them. This would explain the amplitude- and latency-directionalities in the datasets including the earlier artefact-free trials (e.g., upright human faces peaking later than inverted monkey-faces or the Nc-amplitude being significantly more negative for the Toward-condition compared to Away in F1). Later on in the experimental session, when infants have experienced all of the conditions multiple times and have started to habituate to them, their focus may change to other conditions that were less attended towards in the earlier stages of the test-session (see also Chapter 2). Depending on how much meaningful information can be extracted from the previously less attended conditions, the infants may maintain increased interest in these stimuli. Alternatively, they may switch back to attending more to those stimuli that depict the information that is more meaningful and familiar to them (Stets, Burt, & Reid, in preparation).

General Conclusions

The results reported in the meta-analysis in Chapter 3 (Stets, Stahl, & Reid, 2012) clearly indicate that new strategies for both data-collection and data-analysis are needed in infant ERP-studies using standard paradigms. This became especially apparent when attrition rates and the numbers of usable trials were compared for standard and odd-ball paradigms in Chapter 4 (Stets, Burt, & Reid, under review). In this more detailed analysis of the experimental groups reported in the meta-analysis, it was shown that odd-ball paradigms often experienced lower attrition rates and obtained more data-points per condition from their participants when contrasted with other paradigms (e.g., He, Hotson, & Trainor, 2007; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005). However, the main-difference between these and the majority of the standard paradigm studies in which conditions were presented with equal frequency and which were included in Stets, Stahl, and Reid (2012) is that the odd-ball studies mainly used very short auditory stimuli as opposed to the generally slightly longer visual stimuli used in the standard paradigms. Together with the finding that studies with auditory stimuli tend to experience lower attrition rates compared to studies using purely visual stimuli – presumably for the reasons mentioned earlier –, this indicates that odd-ball studies may be too different from standard-paradigm studies to be comparable when examining these issues. Apart from this, the meta-analysis presented in Stets, Stahl, and Reid (2012) shows that there are additional factors lying outside the scope of this consideration, which account for higher or lower attrition rates in the different studies. Examples may be the temperament of the infant in general (e.g., Marshall, Reeb, & Fox, 2009), the experimenter's experience with running a test-session (see Hoehl, & Wahl, 2012), or, as mentioned in Study 2 (see Chapter 3),

infants may be more irritable during certain periods when changes in their living situation occur such as the end of maternity or paternity leave. Alternatively, subtle differences between laboratories such as where the infant will be situated relative to their caregiver during the test-session may have an impact as well. Further investigation into these and other factors will be helpful for developmental researchers to plan their studies accordingly.

Based on the results reported in Chapters 2, 4, and 5, however, it does seem possible to improve the situation in terms of attrition and the volume of data to be obtained for visual standard studies. Using new approaches to data-collection and -analysis, all of these studies could show that infant ERP-studies can be tailored to better fit the needs of these young participants. Moreover, what has been known among developmental researchers previously (e.g., Reid, & Geangu, 2008), that data-analyses based on strategies used for adult ERP-data cannot be easily applied to infant data, is supported by the findings presented in this thesis as well. It may be necessary to collect a large number of trials per condition from adult participants because the amplitudes are generally smaller and harder to detect in this population. For instance, as illustrated in Figure 3.1 in de Haan, Johnson, and Halit (2007, p. 83), the amplitude-difference for the N290 in response to upright and inverted human faces is much larger and the morphology of the component is more pronounced over a larger time-period in a group of 12-month-olds. However, the likely adult counterpart for this component, i.e., the N170, shows a smaller difference between the conditions and the time-period is shorter in the group of adults. Therefore, the same difficulty of having to record large numbers of trials to find ERP-components does not apply to infants. As mentioned in Chapter 1, the physiological differences between infants' and adults' brains and scalps are too large to be ignored. Based on this, it may not be that

surprising that Stets, and Reid (2011; see Chapter 2) found a tendency towards a significant difference between conditions when only the first three artefact-free trials were included in the average (see Analysis 1). Moreover, if fewer trials are required from a participant per condition, a study's attrition rate should naturally decrease as more of the originally tested infants would be able to contribute meaningful data to the analysis (see Analysis 2 in Chapter 1). In Stets, Burt, and Reid (in preparation), a significant difference between the Nc-amplitudes in response to the Toward- and Away-stimuli could even be detected in F1 – the dataset in which only the very first analysable trial was included in the analysis.

Together with the substantial decrease in attrition rates seen in Stets, Burt, and Reid (under review), the increase in analysable data is encouraging – both methodologically and in terms of theories about infant cognition. By examining infant ERP-data in more detail (see Studies 1 and 4), it could be shown that fewer trials from the early stages of a test-session are as meaningful as a larger volume of artefact-free trials. Additionally, unanticipated underlying changes in both the components' amplitudes and latencies were discovered. On the one hand, establishing the presence of these changes adds to our knowledge about cognitive processes in general. On the other hand, the nature of the changes provides new insights into the ways in which infants process certain stimuli. Therefore, in order to both explore the potential of the methodological approaches proposed in the current thesis and to re-evaluate psychologists' present assumptions about infant cognition, further studies are needed.

However, the amplitude- and latency- changes seen in the above analyses may also be due to noise which remained in the data due to the smaller numbers of trials especially in the averages including very few trials. As seen in Stahl et al. (2010), the assumptions about signal-to-noise ratio for instance (e.g., Luck, 2005) are violated in

the analyses presented here. Especially the comparatively large standard deviations and standard errors seen, for instance, in the F-datasets in Chapter 5 may indicate that the reported effects are not based on a true difference between conditions or on interactions between certain factors, but are merely due to random noise and outliers in the datasets. However, as reported above, in the cases when significant differences were found between datasets in the post-hoc one-way ANOVAs reported above, they always appeared between the dataset including the fewest trials (see Study 1 in Chapter 2) or only the very first artefact-free trial (see Study 4 in Chapter 5) and the other datasets representing the same condition and channel-group. Datasets never differed from each other when larger volumes of trials formed the average which is in line with the assumptions made for the signal-to-noise ratio. I acknowledge that the argument that effects found for datasets with fewer trials and larger standard deviations may be based on noise rather than a true difference is harder to rule out for these datasets. However, given the relative homogeneity of the datasets as indicated by the results of the post-hoc one-way ANOVAs, it seems apparent that adding single trials to the averages did not result in an alteration the overall pattern. Therefore, it is more likely that the discrepancies between the results reported in Chapters 2, 3, and 4 and the respective original studies are rooted in the way the data have been analysed (see Studies 1 and 4) and what method has been used for stimulus-presentation (see Studies 3 and 4).

Clearly, further work on designing and testing more methodological tools is necessary. The methodological approached outlined here should be seen as a mere starting point as, given their limitations (i.e., mainly the fact that original results could not be replicated in all cases), conclusive statements about their usefulness may be arguable. Without further attempts at implementing such less traditional strategies, the

validity of the outlined approaches cannot be properly established. However, as one of very rare first steps in the direction of improving the efficiency of infant ERP-studies, they could be helpful for developmental researchers to overcome the issue of having to apply tools designed for research with adult populations to developmental or clinical populations. It is nonetheless true that far more remains to be understood about how specific paradigms interact with the quality and nature of obtained data. Clearly, much work is yet to be done before an encompassing understanding of the relationship between ERP-methods and infant populations will be achieved.

References

Angell, B.D., & Smith, G.E. (1998). Print versus electronic: Editors' insights on the costs and benefits of online journals. *The Journal of Technology Studies*, *24*(1), 55-58.

Bahrick, L.E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*, 190-201.

Brannon, E.M., Libertus, M.E., Meck, W.H., & Woldorff, M.G. (2008). Electrophysiological measures of time processing in infant and adult brains: Weber's law holds. *Journal of Cognitive Neuroscience*, *20*(2), 193-203.

Clifford, A., Franklin, A., Davies, I.R.L., & Holmes, A. (2009). Electrophysiological markers of categorical perception of color in 7-month-old infants. *Brain and Cognition*, *71*, 165-172.

Csibra, G., Davis, G., Spratling, M.W., & Johnson, M.H. (2000). Gamma oscillations and object processing in the infant brain. *Science*, *290*, 1582-1585.

Csibra, G., Tucker, L.A., & Johnson, M.H. (1998). Neural correlates of saccade planning in infants: A high-density ERP study. *International Journal of Psychophysiology*, *29*, 201-215.

Csibra, G., Tucker, L.A., & Johnson, M.H. (2001). Differential frontal cortex activation before anticipatory and reactive saccades in infants. *Infancy*, *2*(2), 159-174.

Csibra, G., Tucker, L.A., Volein, A., & Johnson, M.H. (2000). Cortical development and saccade planning: the ontogeny of the spike potential. *NeuroReport*, *11*(5), 1069-1073.

DeBoer, T., Scott, L.S., & Nelson, C.A. (2007). Methods for acquiring and analysing infant event-related potentials. In M. de Haan (ed.), *Infant EEG and Event-Related Potentials* (pp. 5-38), Hove, England: Psychology Press.

de Haan, M., Johnson, M.H., & Halit, H. (2007). Development of face-sensitive event-related potentials during infancy. In M. de Haan (ed.), *Infant EEG and Event-Related Potentials* (pp. 77-99), Hove, England: Psychology Press.

de Haan, M., & Nelson, C.A. (1997). Recognition of the mother's face by six-month-old infants: A neurobehavioral study. *Child Development*, *68*(2), 187-210.

de Haan, M., Pascalis, O., & Johnson, M.H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199-209.

deRegnier, R.A., Wewerka, S., Georgieff, M.K., Mattia, F., & Nelson, C.A. (2002). Influences of postconceptual age and postnatal experience on the development of auditory recognition memory in the newborn infant. *Developmental Psychobiology*, *41*, 216-225.

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States data. *PLOSOne*, *5*(4), 1-7.

Ferguson, M., & Molfese, P.J. (2007). Breast-fed infants process speech differently from bottle-fed infants: Evidence from neuroelectrophysiology. *Developmental Neuropsychology*, *31*(3), 337-347.

Friederici, A.D., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, *17*, 1208-1211.

Halit, H., de Haan, M., & Johnson, M.H. (2003). Cortical specialisation for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*, *19*, 1180-1193.

He, C., Hotson, L., & Trainor, L.J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*, *19*(5), 878-892.

Hoehl, S., Reid, V., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye-gaze. *Developmental Science*, *11*(1), 10-16.

Hoehl, S., & Striano, T. (2008). Neural processing of eye gaze and threat-related emotional facial expressions in infancy. *Child Development*, *79*(6), 1752-1760.

Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, *37*(3), 187-209.

Ioannidis, J.P.A. (2006). Journals should publish all "null" results and should sparingly publish "positive" results. *American Association for Cancer Research*, *15*, 186.

Jeschonek, S., Marinovic, V., Hoehl, S., Elsner, B., & Pauen, S. (2010). Do animals and furniture items elicit different brain responses in human infants?. *Brain & Development*, *32*(10), 863-871.

Luck, S.J. (2005). Ten simple rules for designing ER experiments. In T.C. Handy (ed.), *Event-related potentials: a methods handbook* (pp. 17-32). Cambridge, MA: MIT Press.

Männel, C., & Friederici, A.D. (2009). Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. *Journal of Cognitive Neuroscience*, *21*(10), 1988-2006.

Marshall, P.J., Reeb, B.C., & Fox, N.A. (2009). Electrophysiological responses to auditory novelty in temperamentally different 9-month-old infants. *Developmental Science*, *12*(4), 568-582.

Marshall, P.J., & Shipley, T.F. (2009). Event-related potentials to point-light displays of human actions in five-month-old infants. *Developmental Neuropsychology*, *34*(3), 368-377.

Mason, A.J.S., Braddick, O.J., Wattam-Bell, J., & Atkinson, J. (2001). Directional motion asymmetry in infant VEPs – which direction?. *Vision Research*, *41*, 201-211.

Mills, D.L., Coffey-Corina, S., & Neville, H.J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*, *13*(3), 397-445.

Nelson, C.A., & Collins, P.F. (1991). Event-related potential and looking-time analysis of infants' responses to familiar and novel events: Implications for visual recognition memory. *Developmental Psychology*, *27*(1), 50-58.

Nelson, C.A., Henschel, M., & Collins, P.F. (1993). Neural correlates of cross-modal recognition memory by 8-month-old human infants. *Developmental Psychology*, *29*(3), 411-420.

Nelson, C.A., & Salapatek, P. (1986). Electrophysiological correlates of infant recognition memory. *Child Development*, *57*(6), 1483-1497.

Nikkel, L., & Karrer, R. (1994). Differential effects of experience on the ERP and behavior of 6-month-old infants: trends during repeated stimulus presentation. *Developmental Neuropsychology*, *10*, 1-11.

Pang, E.W., Edmonds, G.E., Desjardins, R., Khan, S.C., Trainor, L.J., & Taylor, M.J. (1998). Mismatch negativity to speech stimuli in 8-month-old infants and adults. *International Journal of Psychophysiology*, *29*, 227-236.

Parise, E., Handl, A., & Striano, T. (2010). Processing faces in dyadic and triadic contexts. *Neuropsychologia*, *48*, 518-528.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, Jr., R., … Taylor, M.J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*, 127-152.

Reid, V., & Geangu, E. (2008). Research methodology and social cognition. In T. Striano, & V.M. Reid (eds.), *Social cognition: Development, Neuroscience and Autism* (pp. 17-34). Oxford: Blackwell.

Richards, J.E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: An ERP study. *Developmental Science*, *6*(3), 312-328.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P.K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, *8*(2), 162-172.

Scott, L.S., & Nelson, C.A. (2006). Featural and configural face processing in adults and infants: A behavioural and electrophysiological investigation. *Perception*, *35*, 1107-1128.

Simos, P.G., & Molfese, D.L. (1997). Electrophysiological responses from a temporal order continuum in the newborn infant. *Neuropsychologia*, *35*(1), 89-98.

Snyder, K.A., & Keil, A. (2008). Repetition suppression of induced gamma activity predicts enhanced orienting toward a novel stimulus in 6-month-old infants. *Journal of Cognitive Neuroscience*, *20*(12), 2137-2152.

Stahl, D., Parise, E., Hoehl, S., & Striano, T. (2010). Eye contact and emotional face processing in 6-month-old infants: Advanced statistical methods applied to event-related potentials. *Brain and Development*, *32*, 305-317.

Stets, M., Burt, M., & Reid, V.M. (under review). Infants need more variety – Increased data acquisition with reduced participant attrition in infant ERP studies. *FRONTIERS in Developmental Psychology*.

Stets, M., Burt, M., & Reid, V.M. (in preparation). The fine print of infant ERPs: The effect of additional trials on components.

Stets, M., & Reid, V.M. (2011). Infant ERP amplitudes change over the course of an experimental session: Implications for cognitive processes and methodology. *Brain and Development*, *33*(7), 558-568.

Stets, M., Stahl, D., & Reid, V.M. (2012). A meta-analysis investigating factors underlying attrition rates in infant ERP-studies. *Developmental Neuropsychology*, *37*(3), 226-252.

Thierry, G. (2005). The use of event-related potentials in the study of early cognitive development. *Infant and Child Development*, *14*(1), 85-94.

von Koss Torkildsen, J., Sannerud, T., Syversen, G., Thormodsen, R., Simonsen, H.G., Moen, I., … Lindgen, M. (2006). Semantic organization of basic-level words in 20-month-olds: An ERP study. *Journal of Neurolinguistics*, *19*, 431-454.

Webb, S.J., Long, J.D., & Nelson, C.A. (2005). A longitudinal investigation of visual event-related potentials in the first year of life. *Developmental Science*, *8*(6), 605-616.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A.D. (2004). Discrimination of word stress in early infant perception: electrophysiological evidence. *Cognitive Brain Research*, *18*, 149-161.