

Developmental Psychology

The Development of Scene Semantics: First ERP Indications for the Processing of Semantic Object-Scene Inconsistencies in 24-Month-Olds

Laura Maffongelli^{1 a}, Sabine Öhlschläger², Melissa Lê-Hoa Võ²

¹ Department of Psychology, Scene Grammar Lab, Johann Wolfgang-Goethe-University, Frankfurt, Germany; Department of Psychology, Johannes-Gutenberg-University Mainz, Germany, ² Department of Psychology, Scene Grammar Lab, Johann Wolfgang-Goethe-University, Frankfurt, Germany

Keywords: object-scene inconsistency effect, n400, development, scene processing, scene knowledge, erps

<https://doi.org/10.1525/collabra.17707>

Collabra: Psychology

Vol. 6, Issue 1, 2020

Finding a bottle of milk in the bathroom would probably be quite surprising to most of us. Such a surprised reaction is driven by our strong expectations, learned through experience, that a bottle of milk belongs in the kitchen. Our environment is not randomly organized but governed by regularities that allow us to predict *what* objects can be found in *which types of scene*. These *scene semantics* are thought to play an important role in the recognition of objects. But when during development are the semantic predictions so far implemented that such scene-object inconsistencies would lead to semantic processing difficulties? Here we investigated how toddlers perceive their environments, and what expectations govern their attention and perception. To this aim, we used a purely visual paradigm in an ERP experiment and presented 24-month-olds with familiar scenes in which either a semantically consistent or an inconsistent object would appear. The scene-inconsistency effect has been previously studied in adults by means of the N400, a neural marker responding to semantic inconsistencies across many types of stimuli. Our results show that semantic object-scene inconsistencies indeed elicited an enhanced N400 over the left anterior brain region between 750 and 1150 ms post stimulus onset. This modulation of the N400 marker provides first indications that by the age of two toddlers have already established their scene semantics allowing them to detect a purely visual, semantic object-scene inconsistency. Our data suggest the presence of specific semantic knowledge regarding *what* objects occur in a certain scene category.

1. Introduction

Humans have an amazing capability to efficiently perceive and interact with their visual world. In order to deal with the complexity of the world, we have learned to use rules to boost our perception and navigation skills. For instance, we would search for a toothbrush in the bathroom, not the kitchen. However, we are not born with this knowledge, but instead have to learn these distinctions throughout our lives by continuous interaction with our environment. How do toddlers perceive their environment and what expectations govern their attention, perception, and actions?

Much like words in a sentence, objects in a scene seem constrained by a kind of grammar (Boettcher et al., 2018; Võ et al., 2019; Võ & Wolfe, 2013). Roughly speaking, different types of rules have been categorized as *semantic* (*what* objects in which scene) and *syntactic* (*where* are these objects located within the scene) (Võ & Wolfe, 2013). Besides helping us to find things, the rules underlying this grammar are thought to play an important role in the recognition of objects as well as in reducing the computational load of perceptual processes (Biedermann et al., 1982; Davenport & Potter, 2004). That is, when scene grammar is intact, objects that are semantically related to the context of the scene are recognized better and faster than objects that do not show relatedness with the context (Davenport & Potter, 2004; Võ

& Wolfe, 2013). In addition to evidence gained from behavioral studies, the object-scene inconsistency effect has been measured via the N400 event-related-potential (ERP). This component historically described in the language domain is nowadays known to mark semantic processing difficulties occurring along different stimulus modalities and various dimensions (Kutas & Federmeier, 2011). It is hereby assumed that when the processed information fits the previous context, the processing of upcoming information is facilitated. When this information does not fit with prior semantic predictions a mismatch occurs, resulting in an N400 component, which in language is typically elicited when a word in a sentence is semantically inconsistent. Similarly, objects that are semantically incongruent with the global meaning of the scene violate the semantic expectations activated by the scene in which they appear. Presenting a bar of soap next to a laptop elicits an N400 when compared to the presentation of a mouse next to a laptop (Võ & Wolfe, 2013), i.e. a semantically inconsistent object elicits a stronger negativity compared to a consistent one (Mudrik et al., 2010). As adults, we know what objects tend to be in which context, but very little is known about how such scene knowledge develops in children. Taking a developmental perspective on the semantic processing of objects in scenes could provide important new insights into scene understanding.

In the language domain, the N400 as marker of semantic

integration difficulties is already present during early language acquisition. For example, starting after the child's first birthday an N400 has been shown in a picture-word paradigm (Friedrich & Friederici, 2004). From a language production perspective, around this age children go through the so called "two-word" stage of language acquisition where they start using two-word sentences and to name objects and combine words in a sentence (Bates et al., 2003). This age is associated with the organization of the mental lexicon by semantic categories (Rämä et al., 2013) as well as with a vocabulary spurt (Nazzi & Bertoncini, 2003). It is therefore suggested that substantial developmental changes are taking place around the children's second birthday and that they are driven by different processes, such as cognitive change, a change in children's object conceptualizations, word segmentation abilities or advances in pragmatics/social cognition (for a review see Ganger & Brent, 2004). Focusing on children around their second birthday, the current study aimed to investigate whether known adult mechanisms of scene semantic integration in visual scenes are already at work early in development, and to gain insights about their functional meaning. Indeed, finding a similar component in response to scene violations as reported in language processing studies might suggest that both scene knowledge and language understanding rely on similar neurocognitive processes.

To this aim, in this study we presented 24-month-olds with a visual scene (i.e. a dishwasher in a kitchen in which either a consistent (i.e. a mug) or inconsistent object (i.e. a roll of toilet paper) would appear. We hypothesized that if children this age already have established their scene grammar to an extent that allows for the detection of non-verbal, purely visually based, semantic object-scene inconsistencies, an N400 response should be elicited for semantically inconsistent objects. If, however, children at the age of 24 months have not yet developed strong visual predictions on what objects fit a semantic scene context, no modulations of the N400 response would be expected. However, we anticipate that this study represents a first attempt to examine the processing of the semantic relationship between visually presented objects and a scene from a developmental perspective and should be taken with caution.

2. Methods

2.1 Participants

Thirty toddlers (15 female) aged 24 months ($M = 864.90$ days, range = 688-1061 days) who provided a minimum of 10 trials per condition were included in the final sample. In total, an additional 17 toddlers were tested but excluded due to technical problems ($n = 2$), because of unwillingness to wear the net on their head ($n = 5$), or because they contribute too few artifact-free trials ($n = 10$). All children were born full-term (week of gestation ≥ 37), were German monolingual and were recruited from local kindergartens. Parents gave written informed consent. The research protocol was approved by the local ethics commission of the Faculty of Psychology and Sport Sciences at Goethe University Frankfurt. For their attendance in the study, the toddlers received a small age-appropriate gift of a value of about 5 Euro.

2.2. Stimuli and Procedure

The stimuli were selected from the SCEGRAM image database (Öhlschläger & Vö, 2017) and consisted of pictures of real-world indoor scenes. We presented toddlers with

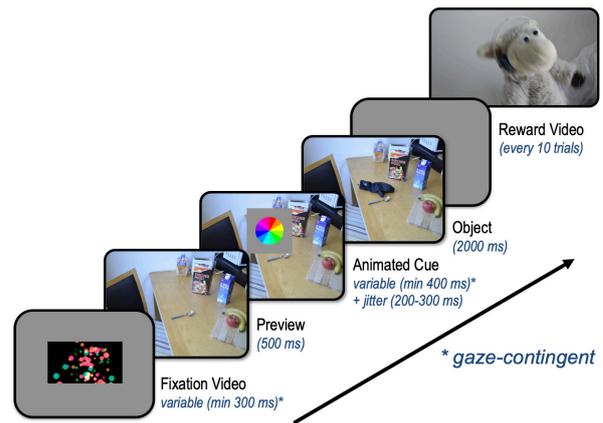


Figure 1. Example of a trial sequence in the EEG experiment.

36 sequences per condition (36 consistent, 36 inconsistent) leading to a total of 72 trials. Stimuli were presented approximately 80 cm from a 24-inch monitor with a resolution of 1920 x 1080 pixels and a refresh rate of 60 Hz. The experiment took place in a sound attenuated and electrically shielded chamber. During stimuli presentation toddlers were seated on a car seat or their parent's lap.

By means of an eye tracker (EyeLink 1000, sampling rate 500Hz) each trial started with the presentation of a gaze-contingent fixation video of colored dots at the center of the screen for a variable amount of time (min 300 ms; see Fig. 1). This was solely to ensure that children paid attention to the screen. As soon as the child fixated on the screen, a scene without the critical object was shown for 500 ms (Preview). After that a gaze-contingent cue (animated cue, dynamic colored dot) at the location where the object would appear was presented for a variable amount of time (min. 400 ms + 200-300 ms random jitter upon cue fixation). This again was to ensure fixation of the location where the critical object would appear. Also, the jitter was chosen to ensure that no entrainment by the offset of the cue would be created, while minimizing EEG artefacts that could arise if the time window between cue offset and scene onset were kept constant.

After cue offset, the object appeared at the cued location and remained on the screen for 2000 ms. Additionally, after the presentation of the scene a blank screen was presented for 400 ms (inter-trial interval). Every 10 trials a reward video (about 10sec) was presented to keep the toddlers' engaged and their attention focused on the screen. Experimental presentation of stimuli was controlled with Matlab (The MathWorks Inc., USA) using the Psychophysics Toolbox (Brainard, 1997).

2.3 EEG recordings and pre-processing

The EEG was continuously acquired by means of 128-channel Geodesic Sensor Net and recorded by the acquisition software Netstation 4.4.2 (EGI, Eugene, OR, USA). During recording the EEG signal was amplified by an EGI Net Amps 300 with a sampling rate of 500 Hz and a 0.1-100 Hz online-high-pass filter. EEG signal was referenced to the vertex electrode (Cz) during acquisition and impedances were kept below 50 k Ω . EEG data pre-processing was performed using EEGLAB (Delorme & Makeig, 2004). An offline 50 Hz notch filter and a digital 0.3-30 Hz band-pass filter

were used. The outermost ring of electrodes was removed because of high noise due to poor contact with the scalp (Maffongelli et al., 2018; Nyström et al., 2011). Artifacts (blinks, eye-movements, muscle) were removed through visual inspection by Independent Component Analysis (ICA) taking into account topographic, time, and spectral distribution of the component (Delorme & Makeig, 2004; see also Lauer et al., 2018; Maffongelli et al., 2018). After artifacts removal, missing channels were interpolated using spherical interpolation and data were re-referenced offline to the average reference.

2.4 Analysis

Thanks to the gaze-contingent paradigm used in this EEG experiment, we were able to track whether the children were looking at the screen or not. In addition, the presentation of the object in the scene was contingent on the child looking at the cue (dynamic dot) for at least 400 ms, increasing the likelihood that the child was indeed attending to the critical part of the scene. This procedure made it possible to exclude all trials from the recorded signal in which the children were not attending to the screen. Therefore, in the analysis we only considered trials in which infants did not move and were attending to the presented stimuli. The EEG recordings were segmented into epochs from -200 ms (Maffongelli et al., 2018; Reid et al., 2009) to 1200 ms relative to the onset of the target (scene/object) using a common baseline (-200 ms to 0 ms; (Hoehl & Wahl, 2012), during which all children looked at the same cue (dynamic dot). Epochs were averaged separately for each participant and experimental condition (consistent/inconsistent). The minimum criterion for inclusion in the final sample was 10 artifact-free trials per condition (e.g. de Haan, 2013; Maffongelli et al., 2018). On average, we obtained 27 trials (range 13-36, $SD = 6.07$) for the consistent condition and 26 trials (range 13-35, $SD = 5.33$) for the inconsistent condition.

ERPs were obtained for each experimental condition by averaging corresponding epochs and were then compared in EEGLAB with a repeated-measures analysis of variance (ANOVA).

Since so far no other developmental research focused on the investigation of purely visual semantic violations in scene processing, we followed standard procedures to analyze our data. To evaluate the magnitudes of ERP components, developmental research often uses fixed time windows of 100-200 ms. Here, following similar procedures (Friedrich & Friederici, 2005; Helo, Azaiez, et al., 2017; Hoehl & Wahl, 2012; Rämä et al., 2013) and based on visual inspection of the resulted ERPs we used pre-defined consecutive time windows of 200 ms, starting from 150 ms. Following this approach, mean amplitudes of ERPs were calculated during five time windows: from 150 to 350 ms, 350 ms to 550 ms, 550 ms to 750 ms, 750 ms to 950 ms, 950 ms to 1150 ms.

Therefore, considering each time window, ANOVA included the within-subject factor CONDITION (consistent, inconsistent), HEMISPHERE (left, right), ANTPOST (anterior, posterior) and average potential in μV as dependent variable. The factors hemisphere and antpost were defined based on the subdivision of the scalp into (ROIs) regions of interest (Helo, Azaiez, et al., 2017). The anterior part of the scalp was formed by: left anterior (E12, E18, E19, E20, E22, E23, E24, E26, E27, E28, E33) and right anterior (E2, E3, E4, E5, E9, E10, E117, E118, E122, E123, E124); the posterior part of the scalp was obtained by considering: left posterior (E36, E37, E41, E42, E47, E51, E52, E53, E54, E60, E61) and

right posterior (E78, E79, E85, E86, E87, E92, E93, E97, E98, E103, E104). Mean amplitudes of all factors were statistically compared with the R statistical package (R Development Core Team, 2011). Post-hoc analyses were performed by means of paired-sample t-tests, where we compared the two conditions within each ROI and hemisphere. In the statistical analysis only significant interactions or main effects ($p < 0.05$) are reported. In the analysis, semantically consistent objects (i.e. a mug in a dishwasher) were compared with objects appearing in the inconsistent scenes (i.e. a roll of toilet paper in a dishwasher; see [Figure 1](#)).

3. Results

Looking at each time window specifically, the ANOVA for the first time window from 150 ms to 350 ms only revealed a main effect of HEMISPHERE, $F(1,29) = 8.49$; $p = 0.006$.

In the second time window from 350 ms to 550 ms, a main effect of HEMISPHERE ($F(1,29) = 15.5$; $p = 0.04$) as well as a two-way interaction CONDITION*ANTPOST were reported. Post-hoc tests run on this interaction showed a significant difference for the consistent condition (CON) compared to the inconsistent condition (INCON) in the anterior region ($t(59) = 2.05$, $p = 0.02$), with the INCON condition inducing more negative amplitude ($M = -0.45 \mu V$, $SD = 1.68 \mu V$) than the CON condition ($M = 0.49 \mu V$, $SD = 1.49 \mu V$).

In the time window from 550 ms to 750 ms, a main effect of HEMISPHERE, $F(1,29) = 19.8$, $p = 0.0001$, as well as a 3-way interaction CONDITION*ANTPOST*HEMISPHERE, $F(1,29) = 4.22$, $p = 0.04$, was found. Post-hoc tests run on this interaction showed a trend towards a significant difference for the consistent condition (CON) compared to the inconsistent condition (INCON) in the left anterior region ($t(29) = 1.54$, $p = 0.06$) with the INCON condition inducing more negative amplitude ($M = -1.81 \mu V$, $SD = 0.99 \mu V$) than the CON condition ($M = 0.49 \mu V$, $SD = 1.56 \mu V$).

In the later time windows between 750 ms and 950 ms the ANOVA revealed a main effect of HEMISPHERE, $F(1,29) = 1.42$, $p = 0.0007$, as well as a 3-way interaction CONDITION*ANTPOST*HEMISPHERE, $F(1,29) = 6.34$, $p = 0.01$. Post-hoc analysis run on this interaction showed a significant difference for the consistent condition (CON) compared to the inconsistent condition (INCON) in the left anterior region, $t(29) = 1.71$, $p = 0.04$, with the INCON condition inducing more negative amplitude ($M = -2.24 \mu V$, $SD = 1.27 \mu V$) than the CON condition ($M = -0.82 \mu V$, $SD = 1.46 \mu V$) ([Fig. 2](#)).

Within the last time window from 950 ms to 1150 ms, the ANOVA revealed a main effect of HEMISPHERE, $F(1,29) = 8.77$, $p = 0.006$, as well as a two-way interaction CONDITION*ANTPOST, $F(1,29) = 4.63$, $p = 0.03$. Post-hoc analysis run on this interaction revealed a significant difference for the consistent condition (CON) compared to the inconsistent condition (INCON) in the anterior region, $t(59) = 1.82$, $p = 0.03$ with the INCON condition inducing more negative amplitude ($M = -0.97 \mu V$, $SD = 1.83 \mu V$) than the CON condition ($M = -0.19 \mu V$, $SD = 1.55 \mu V$). Moreover, a three-way interaction CONDITION*ANTPOST*HEMISPHERE, $F(1,29) = 6.93$, $p = 0.01$, was also found. Post-hoc analysis run on this interaction showed a significant difference for the consistent condition (CON) compared to the inconsistent condition (INCON) in the left anterior region, $t(29) = 2.27$, $p = 0.01$, with the INCON condition inducing more negative amplitude ($M = -2.41 \mu V$, $SD = 1.40 \mu V$) than the CON condition ($M = -0.81 \mu V$, $SD = 1.52 \mu V$), ([Fig. 2](#)).

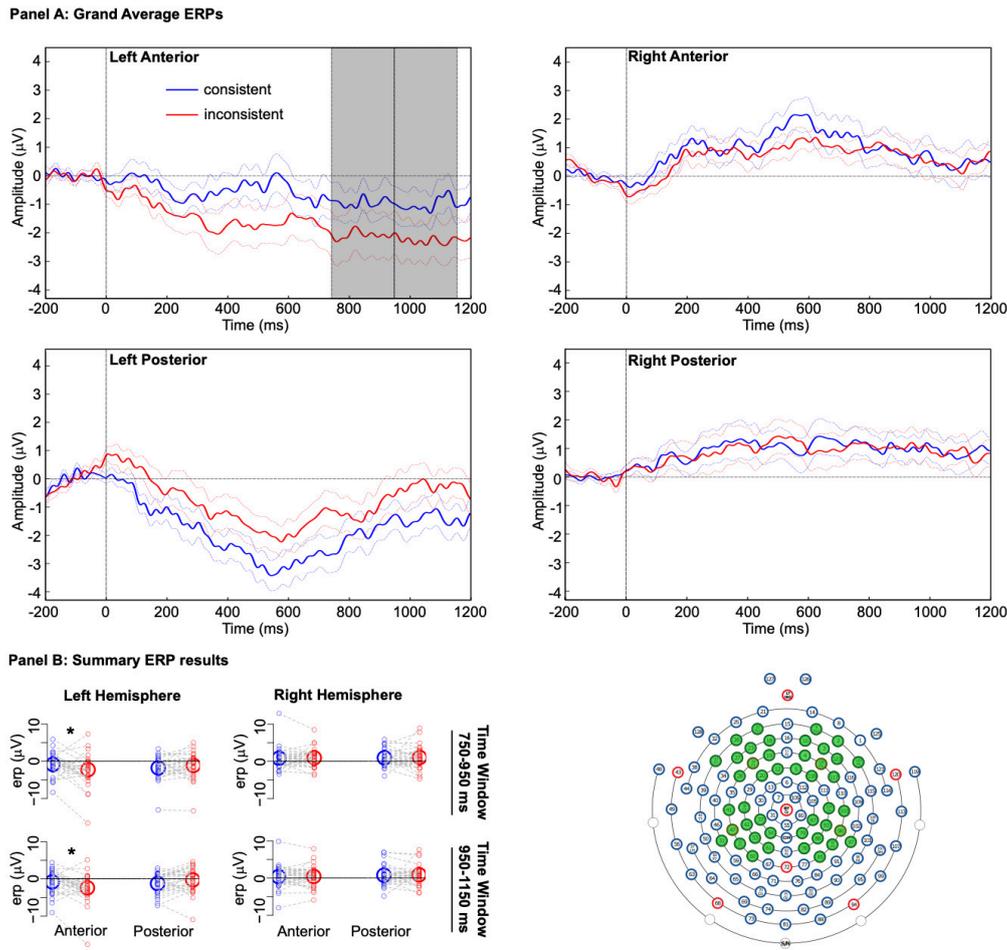


Figure 2. Event-related potentials relative to all experimental conditions/Summary ERPs results. **Panel A:** Grand average event-related potentials (ERPs) of the Consistent (blue) and Inconsistent condition (red) for the selected regions of interest (ROIs) are reported. Dashed lines represent the standard error of the mean for each condition. A significant negative deflection was found in the anterior ROIs within the 750-950 ms as well as 950-1150 ms time windows (highlighted in grey). **Panel B:** Paired average ERP amplitudes for the 750-950 ms time window (upper row) and for the 950-1150 ms time window (lower row) for each participant for the Consistent (blue) and Inconsistent (red) conditions are depicted. Big circles represent the mean value of the conditions. Small circles represent the amplitude value of the subjects. Grey segments show the modulation of the mean value of the subjects related to the different conditions. Asterisks represent significant effects. On the right side the scalp channel configuration is shown. Channels selected for each ROI are highlighted in green.

4. Discussion

The main goal of the current study was to test whether children at the age of two years already show similar semantic processing of purely visual stimuli as adults do and to characterize the neurophysiological activity associated with the processing of scene grammar violations by toddlers. In doing so, we also wanted to test whether the N400 could prove to be a useful marker for semantic processing abilities in children.

To this aim, we presented objects that were either semantically consistent or inconsistent with regard to their scene context. Results show that semantic object-scene inconsistencies indeed elicited an enhanced N400 over the left anterior brain region between 750 and 1150 ms post stimulus onset. The effect seemed to be already developing in the previous time window between 550 and 750 ms, but only barely reached statistical significance. While these time windows are later than the ones known from adult studies (Lauer et al., 2018; Mudrik et al., 2010; Vö & Wolfe,

2013), one needs to consider that ERP components in infants and young children tend to be temporally delayed compared to adults (de Haan, 2013). Therefore, it is possible that the time course of our effects for toddlers might actually be similar to the time course previously reported for scene semantic processing in adults (Lauer et al., 2018; Mudrik et al., 2010; Vö & Wolfe, 2013), just postponed by about 200 ms. In contrast to this previous evidence gained from investigation in adults reporting a mid-central N400 effect (Vö & Wolfe, 2013) we found an effect over the anterior brain region and in the left hemisphere only. Interestingly, the anterior distribution also resembles effects usually reported during action observation. Indeed, it has been suggested that action stimuli call for action-specific mechanism located in anterior brain regions (Aziz-Zadeh et al., 2006; Maffongelli et al., 2018). The frontal shift might therefore be related to the characteristics of the stimuli presented in the current study, which were presented in a sequential fashion, possibly resembling an action to the children. However, it has to be noted that neural activity

is usually widespread in children and only becomes topographically focused following increased experience as well as brain maturation (de Haan, 2013; Johnson, 2011), therefore we refrain from drawing strong conclusions from the scalp topography at this point.

Recent behavioral evidence shows that semantic scene context affects object processing in young children (Bornstein et al., 2011; Duh & Wang, 2014; Helo, van Ommen, et al., 2017). Scene grammar is built through visual experience and top-down control of visual attention is suggested to increase with age (Mandler & Johnson, 1976). Using a free exploration task, it was recently shown that semantically inconsistent objects had a stronger effect (i.e. longer viewing time) in 24 month-olds as compared to semantically consistent objects (Helo, van Ommen, et al., 2017). Further, Duh and Wang (Duh & Wang, 2014) presented 15-month-olds with visual scenes in a habituation paradigm. Presenting the stimulus for 3000 ms, which allows access to scene meaning, children looked longer at the critical location when a change disrupted the meaning of a scene compared to a perceptual salient change of the scene (e.g. replacing a beach umbrella by a table vs. replacing a beach umbrella by a colorful beach umbrella). This result suggests that children at 15 months of age already take into account low and high-level features during scene processing. The question is whether these 15-month-old children really processed the scene on a semantic level or only detected some higher-level oddity in the scene.

In this study, we tried to answer this question by investigating ERPs in toddlers, more specifically the N400 as marker for semantic processing. As far as we know, the current study provides the first EEG evidence of purely visual semantic object-scene inconsistency effects in developmental studies, that is, we opted for a paradigm in which language had been completely excluded as experimental factor. While previous studies using ERPs have addressed a similar question, they had done so in a slightly different way: ERPs were used for the investigation of the semantic violation of object-word recognition. These studies using the object-word paradigm found a larger N400 for inconsistent as compared to inconsistent object-word pairs during early development (Friedrich & Friederici, 2004, 2005; Torkildsen et al., 2006). Helo et al. (Helo, Azaiez, et al., 2017), for instance, investigated how contextual information facilitates word processing, using a scene-word paradigm in which the vision of a consistent scene (kitchen) could be followed by the auditory presentation of a consistent word (knife) or an inconsistent word (bus). Inconsistent scene-word pairs elicited a larger N400 component (measured at the target word) over anterior regions.

In comparison to these previous studies, our data showed a modulation of the N400 response triggered by the presence of a purely visual, semantic object-scene inconsistency. We therefore report first indications that toddlers at the age of two might already have established a form of scene grammar, i.e. visually based semantic knowledge regarding *what* objects occur in a certain scene category. The N400 has been observed across many types of stimuli (linguistic, objects, actions, pictures, sound; (Kutas & Federmeier, 2011), and might prove very useful for the investigation of the development of higher level, non-verbal scene understanding and might well be a universal marker for semantic processing across domains.

Note that the present data should be interpreted with caution since this study only represents a first attempt to

investigate the processing of the semantic relationship between visually presented objects and a scene. In order to further test the universality of the N400 response and to investigate differences and/or similarities with the language domain starting from early development, future studies should also consider the assessment of the language skills of the children. It might also be interesting to track how ERPs in response to semantic manipulations of a scene change as a function of the children's language skills, for example, by comparing children with low and high vocabulary skills or adding language scores into linear mixed model approaches. In addition, to make the interpretation of the collected EEG data stronger one might also want to simultaneously collect behavioral measurements to track the reaction of the toddlers during the presentation of the inconsistency between the presented object and the embedded scene. Since evidence for semantic object processing in 24-month-old toddlers has been mixed (Helo, van Ommen, et al., 2017; Öhlschläger, S & Vö, 2020), collecting behavioral data from the same children whose EEG had been recorded might be able to gain better insights into interindividual differences of such young toddlers and how these come about. Once more established, this could also be done in a free viewing paradigm using fixation-related potentials (Coco et al., 2020; Cornelissen et al., 2019).

To conclude, using a gaze-contingent paradigm and presenting purely visual, semantic object-scene inconsistency, we were able to provide first indications that toddlers at the age of two might be able to already process object-scene inconsistencies to an extent that can be observed in slight modulations of the N400 ERP response. We hope that this line of research will trigger further, larger-scale investigations in the development of scene grammar.

Contributions

M.L.V. and S.Ö. contributed to conception and design of the experiment. S.Ö. collected the data. L.M. conducted analysis of the data. L.M. and M.L.V. interpreted the data and wrote the paper. All authors gave final approval for publication.

Acknowledgements

We thank the parents and their children who contributed to this research as well as KiTa and IDeA (Individual Development and Adaptive Education of Children at Risk) Frankfurt/Main.

Funding

This work was supported by DFG grant VO 1683/2-1 and by SFB/TRR 26 135 project C7 to Melissa L.-H. Vö.

Conflicts Interest

We have no conflict of interest.

Data availability statement

All the stimuli, presentation materials, participant data, and analysis script can be found here [https://osf.io/za3kx/?view_only=69d3d3acf70648a48ce55f9428511dff]

REFERENCES

- Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., & Iacoboni, M. (2006). Lateralization of the human mirror neuron system. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 26(11), 2964–2970. <https://doi.org/10.1523/JNEUROSCI.2921-05.2006>
- Bates, E., Thal, D., Finlay, B., & Clancy, B. (2003). Early language development and its neural correlates. In S. J. Segalowitz & I. Rapin (Eds.), *Handbook of Neuropsychology* (Child Neur). Elsevier.
- Biedermann, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14(2), 143–177.
- Boettcher, S. E. P., Draschkow, D., Dienhart, E., & Vö, L.-H. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects on eye movements during visual search. *Journal of Vision*, 18(13), 11. <https://doi.org/10.1167/18.13.11>
- Bornstein, M. H., Mash, C., & Arterberry, M. E. (2011). Young infants' eye movements over "natural" scenes and "experimental" scenes. *Infant Behavior and Development*, 34(1), 206–210. <https://doi.org/10.1016/j.infbeh.2010.12.010>
- Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision*, (10), 433–436.
- Coco, M. I., Nuthmann, A., & Dimigen, O. (2020). Fixation-related brain potentials during semantic integration of object-scene information. *Journal of Cognitive Neuroscience*, 32(4), 571–589. https://doi.org/10.1162/jocn_a_01504
- Cornelissen, T., Sassenhagen, J., & Vö, L.-H. (2019). Improving free-viewing fixation-related EEG potentials with continuous-time regression. *Journal of Neuroscience Methods*, (313), 77–94. <https://doi.org/10.1016/j.jneumeth.2018.12.010>
- Davenport, J. L., & Potter, M. C. (2004). Scene Consistency in Object and Background Perception. *Psychological Science*, 15(8), 559–564. <https://doi.org/10.1111/j.0956-7976.2004.00719.x>
- de Haan, M. (2013). *Infant EEG and event-related potentials*. Taylor & Francis Group.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neurosci Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Duh, S., & Wang, S. (2014). Infants detect changes in everyday scenes: The role of scene gist. *Cognitive Psychology*, 72, 142–161. <https://doi.org/10.1016/j.cogpsych.2014.03.001>
- 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. <http://doi.org/10.1162/0898929042304705>
- Friedrich, M., & Friederici, A. D. (2005). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *Undefined*. <https://www.semanticscholar.org/paper/Lexical-priming-and-semantic-integration-reflected-Friedrich-Friederici/42a52c43e1ff4426b2cec3376eec0871bf7db269>
- Ganger, J., & Brent, M. R. (2004). Reexamining the Vocabulary Spurt. *Developmental Psychology*, 40(4), 621–632. <https://doi.org/10.1037/0012-1649.40.4.621>
- Helo, A., Azaiez, N., & Rämä, P. (2017). Word Processing in Scene Context: An Event-Related Potential Study in Young Children. *Developmental Neuropsychology*, 42(7–8), 482–494. <https://doi.org/10.1080/87565641.2017.1396604>
- Helo, A., van Ommen, S., Pannasch, S., Danteny-Dordoigne, L., & Rämä, P. (2017). Influence of semantic consistency and perceptual features on visual attention during scene viewing in toddlers. *Infant Behavior & Development*, 49, 248–266. <http://doi.org/10.1016/j.infbeh.2017.09.008>
- Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, 37(3), 187–209. <https://doi.org/10.1080/87565641.2011.627958>
- Johnson, M. H. (2011). Interactive specialization: A domain-general framework for human functional brain development? *Developmental Cognitive Neuroscience*, 1(1), 7–21. <https://doi.org/10.1016/j.dcn.2010.07.003>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Lauer, T., Cornelissen, T., Draschkow, D., Willenbockel, V., & Vö, L.-H. (2018). The role of scene summary statistics in object recognition. *Scientific Reports*, 8(1), 14666. <https://doi.org/10.1038/s41598-018-32991-1>

- Maffongelli, L., Antognini, K., & Daum, M. M. (2018). Syntactical regularities of action sequences in the infant brain: When structure matters. *Developmental Science*. In press. <https://doi.org/10.1111/desc.12682>
- Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5), 529–540. <https://doi.org/10.1037/0278-7393.2.5.529>
- Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, 48(2), 507–517. <https://doi.org/10.1016/j.neuropsychologia.2009.10.011>
- Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, 6(2), 136–142. <https://doi.org/10.1111/1467-7687.00263>
- Nyström, P., Ljunghammar, T., Rosander, K., & von Hofsten, C. (2011). Using mu rhythm desynchronization to measure mirror neuron activity in infants. *Developmental Science*, 14(2), 327–335. <https://doi.org/10.1111/j.1467-7687.2010.00979.x>
- Öhlschläger, S. S., & Vö, M. (2020). Development of scene knowledge: Evidence from explicit and implicit scene knowledge measures. *Journal of Experimental Child Psychology*, (194), 104782.
- Öhlschläger, S., & Vö, L.-H. (2017). SCEGRAM: An image database for semantic and syntactic inconsistencies in scenes. *Behavior Research Methods*, 49(5), 1780–1791. <https://doi.org/10.3758/s13428-016-0820-3>
- Rämä, P., Sirri, L., & Serres, J. (2013). Development of lexical–semantic language system: N400 priming effect for spoken words in 18- and 24-month old children. *Brain and Language*, 125(1), 1–10. <https://doi.org/10.1016/j.bandl.2013.01.009>
- 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. <https://doi.org/10.1037/a0015209>
- Torkildsen, J. von K., Sannerud, T., Syversen, G., Thormodsen, R., Simonsen, H. G., Moen, I., Smith, L., & Lindgren, M. (2006). Semantic organization of basic-level words in 20-month-olds: An ERP study. *Journal of Neurolinguistics*, 19(6), 431–454. <https://doi.org/10.1016/j.jneuroling.2006.01.002>
- Vö, L.-H., Boettcher, S., & Draschkow, D. (2019). Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, 29, 205–210. <https://doi.org/10.1016/j.copsyc.2019.03.009>
- Vö, L.-H., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science*, 24(9), 1816–1823. <https://doi.org/10.1177/0956797613476955>

SUPPLEMENTARY MATERIALS

Peer Review Comments

Download: https://collabra.scholasticahq.com/article/17707-the-development-of-scene-semantic-first-erp-indications-for-the-processing-of-semantic-object-scene-inconsistencies-in-24-month-olds/attachment/46114.pdf?auth_token=cTCF_ai8l3Cf7grWfrRX
