Title of the study (one request per article):
Network characteristics of the developing infant brain

Contact person for the proposed study: (please note that this should be level postdoc or higher)	
Name:	Bauke van der Velde
Institution:	Utrecht University
Department:	Developmental and Experimental Psychology
Address:	Heidelberglaan 1, H037
Email:	b.vandervelde@uu.nl

Contact person in YOUth Data Management Committee:	
Name:	Prof. Chantal Kemner
Institution:	Utrecht University
Department:	Developmental and Experimental Psychology
Address:	Heidelberglaan 1, H068
Email:	C.Kemner@uu.nl

Wave (more options are possible):

- ☐ Rondom zw – 20 weeks
- ☐ Rondom zw – 30 weeks
- ☒ Rondom 0 – 5 mo
- ☒ Rondom 0 – 10 mo
- ☐ Rondom 3 (not available yet)
- ☐ Rondom 6 (not available yet)
- ☐ Rondom 9
- ☐ Rondom 12 (not available yet)
- ☐ Rondom 15 (not available yet)

We ask you to provide us with a clear background, methods section and data-analysis plan. These parts of the proposal will be publicly displayed for reference.

Background of the project (max. 500 words): Please provide a short background including the rational of your study as you would do in an introduction of the paper
The infant's brain develops strikingly during the first year of life, with the brain tripling in size (Gilmore, Knickmeyer, & Gao, 2018). This structural development coincides with the largest behavioral development in the human lifespan. During the first year of life, infants start exhibiting complex behaviors, like increasing social competence and behavioral control. An area of research which closely relates brain with behavior is the study of the human connectome, the collection of functionally distinct brain regions and its interconnectivity. Heterogeneous and complex behaviors like social interactions need input from a wide variety of functionally specialized brain areas.

Understanding how these brain areas interact is of critical importance to understand how complex behaviors arise from underlying brain structure.

This study aims to increase our understanding of the relationship between the connectome and complex behaviors, by relating individual differences in the development of both during infancy. Research into the development of the connectome has found that from a structural point of view, most of the macroscopic structures are already present at birth (Ball et al., 2014; van den Heuvel et al., 2014). After birth, the connectome mostly optimizes through selective pruning of unused connections and strengthening of oft-used connections (Keunen, Counsell, & Benders, 2017). The optimization of the connectome during childhood combines increased specialization, task-dependent activity patterns become more focal during development (Durston & Casey, 2006), with increased integration, through a larger dependence of long-range versus short-range connectivity (Fair et al., 2009). It is expected that this optimization process starts from birth and is, therefore, present throughout infancy. Research on this effect, however, is limited.

The integration and segregation of networks can be studied with several graph characteristics (Bullmore & Sporns, 2009): characteristic path length, the average shortest paths between all the nodes in the network; clustering coefficient, representing the amount of highly connected clusters in the network; modularity, the detection of clearly defined modules in the network; and small-worldness, an organizational trait of the network which both combines high clustering with a relatively low characteristic path length. Several cross-sectional studies have used these graph characteristics to determine whether integration and segregation does in fact increase during infancy. A decrease of average path length and an increase of average clustering coefficient during the first year of life it has been reported multiple times (Gao, Alcauter, Smith, Gilmore, & Lin, 2015; Xie, Mallin, & Richards, 2018). Therefore, it is likely that the connectome, under the influence of live experience, optimizes during the first year of life.

Previous studies mainly used cross-sectional data and a relatively limited number of subject pools to reach these conclusions. This limits the possibilities of using this data to connect individual differences in the development of the whole brain network to individual differences in the development of complex behaviors. Network characteristics can prove useful in calculating these individual differences in network development, since they show variability between subjects during a session, but remain relatively stable between sessions (Finn et al., 2015; Mueller et al., 2013). Also, network characteristics have been shown to be reliable over sessions in infants (van der Velde, 2019). A multisession study with a large sample size researching the individual development of whole brain networks can, therefore, prove vital in explaining differences in patterns in the development of complex behavior.

Research question

This study aims to answer the following main and subquestions:

1) **How do network characteristics gathered from infant EEG data develop during the first year of life?**
   a) Are network characteristics stable across sessions?
   b) Are network characteristics continuous across sessions?
   c) Is there variability in the change between sessions?
   d) Is the whole-brain network maturing?
2) **Does the individual maturation of brain networks mediate the development of social competence and behavioural control?**
   a) Does brain network functionality at time point 1 influence social competence / behavioural control at time point 2?
   b) Does social competence / behavioural control at time point 1 influence brain network functionality at time point 2?

Methods Describe the methods as in the paper in which the data will be presented, according to the categories below, with a total **maximum** of 1500 words. For a description of task, methods etc. refer to the website, if possible.

Design of the study (for instance cross-sectional, longitudinal etc.; substantiate your choices)

The EEG-coherence task will be used to create functional whole brain networks for both the 5-month and 10-month waves. Network charactistics, brain network maturation, and their relation to behavioral development will be compared within subject over the two sessions. Therefore, this study will follow a longitudinal design with 2 measures.

Study population and sample-size (entire population or a subset; substantiate your choices e.g. Provide a rationale for the requested sample-size, for instance using a power calculation)

For this particular study we would like to use as many infants as possible. No useful power calculation can be calculated, since the exact effect sizes of what we want to measure are currently unknown. However, this study will be focusing on the individual development of brain networks and child temperament. Individual differences in network characteristics will be relatively small and the larger sample size will be a selling point of this particular study. Lastly, this study can be used as a preliminary study into meaningful effect sizes regarding the development of brain network characteristics during infancy. Later studies in the YOUth project, consisting of the entire population of infants can therefore better understand how many subjects are needed to specifically test certain assumptions. Lastly, this study tries to determine whether children develop along a relatively stable developmental pattern during their first year of life or can be grouped into categories of developmental patterns. This information can later be used to test on the whole population. All this is impossible with a limited sample size.

Data processing and preparation (including necessary recoding of data etc.)

EEG Acquisition

EEG is being recorded using a cap with 32 electrodes (ActiveTwo system, BioSemi) positioned according to the international 10/20 system, at a sampling rate of 2048 Hz. A Common Mode Sense (CMS) and Driven Right Leg (DRL) electrode were used to provide an active ground.

Questionnaires

The Infant Behavior Questionnaire Revised (IBQ-R) (Gartstein & Rothbart, 2003) was used for parent reports of child temperament in infancy. In this study we used factor scores for orienting, surgency/positive affect and negative affect for the infants.

Handling missing data (describe how you will detect and handle missingness in the data)

- For the description of EEG data infants are included only if:
  - Both EEG-sessions are clean and present
- For the combination of EEG-networks and behavior infants are included only if:
  - Both IBQ-R and ASQP:SE questionnaires are adequately filled in for both sessions
  - Both EEG-sessions are clean and present

Data analysis methods (including statistical design and statistical analysis plan. If it is not possible to provide a detailed statistical plan, as this does not fit in with the research questions formulated above, please explain.)

EEG Analysis

EEG data will be analyzed exclusively using Matlab, by means of the FieldTrip toolbox (Oostenveld et al., 2011). The original 2048 Hz data will be down sampled to 512 Hz, using chip interpolation and band-pass filtered at 0.1-70 Hz with a two-way Butterworth filter. Artifacts were removed from the continuous EEG. Artifacts will be defined as absence of signal, clipping, muscle artifacts and excessive noise. Channels will be removed if more than 50 percent of the signal in a channel contained artifacts. If no more than two bad channels are found in a single subject, the two channels will be interpolated by means of weighted averaging neighboring channels. If a subject has more than 2 bad channels, the subject will be removed. The cleaned data was used for further analysis.

Connectivity calculation

The cleaned data for each subject will be bandpass filtered into 6 bands: delta (0.1-3 Hz), theta (3-6 Hz), alpha1 (6 – 9 Hz), alpha2 (9 – 12 Hz), beta (12 – 25 Hz), and gamma (25 – 45 Hz). Since individual theta and alpha peaks are influenced by development, alpha1 and theta bands were chosen to encompass all theta and alpha peaks +/- 1 Hz. The resulting data was cut into 5s. epochs. 20 random epochs were picked per subject per session. For each epoch, connectivity between pairs of electrodes (32*31/2 = 496) was calculated with the phase lag index (PLI) and the debiased weighted PLI, both relying on the same principle of phase locking or phase synchrony (Tass et al., 1998). The phase lag index (PLI), proposed by Stam et al., (Stam et al., 2007), describes the asymmetry of the distribution of phase differences between pairs of signals:

$$\mathrm{PLI} = \left|\langle sign\left[\sin\left(\Delta\varphi(tk)\right)\right]\rangle\right| \ ,$$

where $\Delta\varphi$ is the instantaneous phase difference between signals at time point t for k = 1 … N per epoch (N = 5*512 = 2560), determined using the Hilbert transformation. || stands for absolute values, <> for the mean values and the sign for a signum function (phase difference is either -1, 0, or 1). The resulting PLI can range from 0 to 1. Connectivity matrices will be created, with each cell corresponding to the PLI between two electrodes.

<u>Graph Analysis</u>

Several graph measures were calculated using the acquired individual connectivity matrices. To eliminate the need for arbitrary thresholds, proportional thresholding was performed at 10 different points ranging from 10 percent included until 100 percent included. The following graph measures were calculated for each threshold using the brain connectivity toolbox (Sporns & Rubinov, 2010) (table 1): global connectivity, average clustering coefficient (Cw), characteristic (average shortest) path length (Lw); and small-worldness index (SWI, calculated as the ratio between normalized Cw and normalized Lw). Both the normalized clustering coefficient and normalized path length are used in this study to overcome the problematic influence of total connectivity on graph characteristics (van den Heuvel et al., 2017). An area under the curve (AUC) was calculated for each characteristic in order to simplify this curve to one value.

Table 1. Graph measures references and formulae

Name		Formula	References
Normalized Average clustering coefficient	Cw	$\mathrm{Cw} = \frac{1}{n}\sum_{i \in N} \frac{2t_i}{k_i(k_i - 1)}$	(Onnela et al., 2004)
Normalized Characteristic path length	Lw	$\mathrm{Lw} = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, i \neq j} d_{ij}}{n-1}$	(Watts and Strogatz, 1998)
Small-worldness Index	SWI	$\mathrm{S} = \frac{\mathrm{C}/\mathrm{C}_{\mathrm{rand}}}{\mathrm{L}/\mathrm{L}_{\mathrm{rand}}}$	(Humphries and Gurney, 2008)

<u>Behavior</u>

Behavioral development was assessed with the use of parent questionnaires. To asses behavioral control the effortful control scale of the Infant Behavior Questionnaire Revised (IBQ-R) (Gartstein & Rothbart, 2003) was used. To assess social competence, the ages and stages – social emotional questionnaire (Squires *et al.,* 2002) was used.

<u>Statistics</u>

We aim to answer all questions using structural equation modeling (SEM), which will be calculated using MPlus4. SEM is a confirmatory data analysis technique, using a priori theoretical models to observed data, following which goodness of fit can be ascertained. For the first part of the study, we use the graph characteristics as latent variables to determine the course of development of

these variables throughout infancy. An auto regression analysis will give us an understanding of the stability of these network measures. The variability of these measures within sessions will also be compared over sessions. For the second part of the study, the graph measures (or our neural correlates) will be related to social competence and behavioral control (or our behavioral correlates). A high correlation between our neural correlates is expected. Therefore, a factor analysis will be performed to simplify our neural correlates into one factor explaining most of the variance. This factor will be used in a cross-lagged-panel model to determine the influence of network maturation in session 1 on behavior in session 2 and vice versa.
Planned subgroup analyses (if applicable. Substantiate your choices)
Not applicable
Planned sensitivity analyses (if applicable. Substantiate your choices) Sensitivity analyses are analyses that you plan beforehand to test whether certain factors have a major influence on your results.
Not applicable

2. **Timeline and milestones (including dates of when to analyze/write up):**
   Data analyzed by end of April 2019
   Writing May 2019 (First version on May 10th)

3. **Output (e.g. article, report, etc.):**
1 research article

4. **Proposed authors + affiliations (please note that the YOUth data access committee can request certain authors to be included):**
   Bauke van der Velde
   Chantal Kemner

This form should be sent to: Secretary of Chantal Kemner: i.bleeker@uu.nl