Title of the study (one request per article):
How do environmental factors impact the development of cognitive control and its neural circuitry?

Contact person for the proposed study: (please note that this should be level postdoc or higher)	
Name:	Prof. Dr. Sarah Durston
Institution:	UMC Utrecht
Department:	Psychiatry, Neuroscience (Hersenen)
Address:	Heidelberglaan 100 – HP A 01.126
Email:	s.durston@umcutrecht.nl

Contact person in YOUth Data Management Committee:	
Name:	
Institution:	
Department:	
Address:	
Email:	
Phone:	
Address:	

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
Cognitive control is the ability to plan and adapt behavior flexibly in the face of changing circumstances. This ability develops from childhood to late adolescence, as does the underlying neural circuitry. There are large individual differences in cognitive control and the rate at which it develops. At one extreme end of this spectrum, 'atypical' development of cognitive control is related to (symptoms of) various developmental psychiatric disorders, such as Attention Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorders (ASD) (Durston and Casey, 2006). Given the great impact that cognitive control has on mental health and behavior (Moffit et al., 2011), more insight into the factors that drive this development is necessary.

While an extensive literature has already investigated the relation between neurobiology, genetics and cognitive control (Casey, 2015), it is relatively unknown how environmental factors affect the rate of development of this ability. However, evidence is emerging that socio-economic status (SES) also affects neurocognition - in particular language and cognitive control - and its associated neural circuitry. Yet, SES is a complex construct, including many different factors, such as prenatal influences (e.g. prenatal drug use, maternal stress), parental care (e.g. parent-child interactions, parental life events, stress) and cognitive stimulation in the home environment that may each contribute uniquely to individual differences in the development of cognitive control. Furthermore, it has been suggested that experience-driven changes in neurodevelopment as a result of SES ('social causation') and (prenatal) changes in neurodevelopment that predispose an individual to a certain SES ('social selection') are not mutually exclusive and may each have their greatest impact during different stages in development (Hackman *et al*, 2010).

Therefore, we would like to use the large cohort of typically developing children from the YOUth-cohort, who have been profiled on a broad, multimodal array of characteristics neurocognition, to investigate which environmental factors predispose a child to perform at the extremes of the cognitive control distribution, and how this is mediated by brain activity.

Research question

To determine whether, and which, environmental factors affect cognitive control, we will first perform an exploratory factor analysis, to find interpretable and informative factors underlying environmental influences on cognitive control. We expect that strengths or weaknesses in cognitive control will manifest through different profiles which, in turn, are mediated by different environmental factors identified through the factor analysis. In terms of neural correlates of cognitive control, we hypothesize that children with poor cognitive control will show decreased activity in underlying neural circuitry compared to children with better cognitive control. In addition, we expect that different environmental factors will have distinct effects on brain activation underlying cognitive control, in children with poor- compared to strong cognitive control.

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)

We plan to perform an Exploratory Factor Analysis on a selection of child environmental questionnaires to establish reliable factors that may affect the development of cognitive control.

Specifically, we aim to include a total of approx. 240 items from the following 'child questionnaires' relating to the child characteristics like personality, competence skills, parental lifestyle (from child perception), and physical development:

- Interpersonal Reactivity Index

- EATQ-R
- BIS
- CBSA
- PDS
- PDS (Parental Control)
- CRBPI
- NRI-SF

In addition, we would like to add a total of approx. 400 items from the following 'parent about child questionnaires', exploring the child's behavior, (mental) health, and parental behavior/lifestyle (from parent perception):

- CBCL
- SDQ
- SWAN
- EATQ-R
- QBF
- NOSI
- NRI-SF – parent report
- Medical questionnaire on child's health

Furthermore, we plan to correlate these factors with the following 'parent questionnaires', relating to SES, general health and life events of the parents:

- Demographics (Mother/Father)
- General Health (Mother/Father)
- List of longterm stressful life events (Mother/Father)

As a cognitive measure of cognitive control, we would like to include:

- the Discount Delay task
- the behavioral data from the fMRI stop-signal anticipation task (SSAT)

Subsequently, we will use the output of the exploratory factor analysis to inform our analysis of brain activity during the fMRI-task investigating the contribution of child characteristics and environmental factors on the neural correlates of cognitive control, as is described in more detail below.

The above-mentioned questionnaires, including 8 child-, 8 parent about child- and 6 parent questionnaires, are indicated in the attached Excel form (data selection template v1.xls).

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)

We plan to include the maximum number of datasets that are currently available from the rondom-9 cohort, as a large sample will allow us to perform the most comprehensive factor analysis.

More specifically, strict rules regarding sample size for Exploratory Factor Analysis (EFA) have mostly disappeared. Studies have shown that adequate sample size is partly determined by the nature of the data (Fabrigar et al., 1999; MacCallum, Widaman, Zhang, & Hong, 1999). In general, the 'stronger' the data, the smaller the sample can be for an accurate analysis. What data is 'strong' depends on: 1) The communalities of the items; 2) Number of factors/number of variables; and 3) Size of loading.

Although strong data allows the use of relatively smaller samples, research has proven that for an Exploratory Factor Analysis it is highly desirable to use large samples. For instance, seminal work by Costello and Osborne (2005) showed that the number of misclassified items was significantly affected by sample size. On average, almost two of 13 items were misclassified as belonging to the wrong factor in the smallest samples, whereas just over one item in every two analyses were misclassified in the largest samples. Further, two problem indicators were exclusively observed in the smaller samples: the presence of Heywood cases (factor loadings greater than 1.0, an impossible outcome) and failure to converge. Finally, larger samples tended to produce solutions that were more accurate. Only 10% of samples in the smallest sample produced correct solutions, while 70% in the largest produced correct solutions.

Taken together, the most replicable results that will generalize beyond a particular sample or to our population of interest are obtained by using as large a sample as possible. Therefore, we would like to include data of the maximum number of children that has currently been tested in the Rondom 9 cohort.

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

Questionnaire data has already been processed to the level that it can be distributed to researchers. As we will use raw item scores, recoding of data intro T-scores or subscales is not necessary. For the fMRI data, we will follow the standardized pipeline by Matthijs Vink to harmonize our analyses with other studies in the consortium.

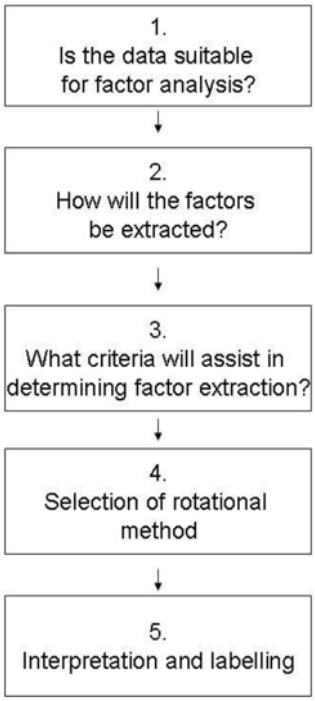
Handling missing data (describe how you will detect and handle missingness in the data)

Results reported by McNeish (2016) show that, when handling missing data, removing of rows with missing values (list-wise deletion) do not extract the proper number of factors and estimate the factor loadings with severe bias, even when data are missing completely at random. For this reason, we will not remove rows with missing values and will keep incomplete cases for data analysis.

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.)

Exploratory Factor Analysis: Factor analysis of the questionnaires will be performed using the R 3.5.1 software package. All items from all questionnaires will be included in the factor analysis, but we will make a distinction between questionnaires about the child and about the parent.

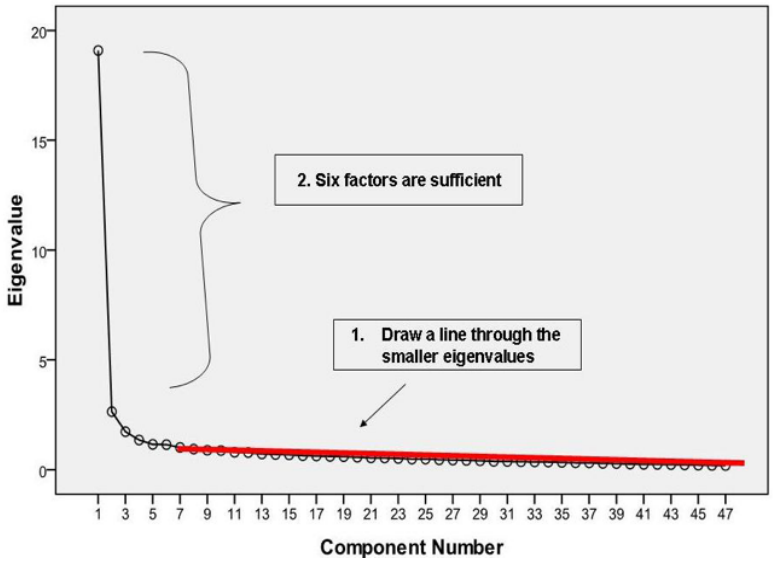
The protocol of analysis will consist of 5 steps, highlighted below:



First, we will check if the data is suitable for Exploratory Factor Analysis, by creating a correlation matrix, displaying the relationships between all individual variables. The correlation matrix will be inspected for correlation coefficients over .30, and categorize them as +0.30 = minimal, +0.40 = important, and +0.50 = practically significant. Further, prior to factor extraction we will run several additional tests, including the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy, and Bartlett's Test of Sphericity. The KMO index ranges from 0 to 1, with a minimum of 0.50 considered suitable for factor analysis. The Bartlett's Test of Sphericity should be significant ($p<.05$) for factor analysis to be suitable.

In Step 2, we determine how factors will be extracted. We will use Principal Component Analysis (PCA), the method most commonly reported in the literature (Henson and Roberts, 2003; Tabachnick and Fidell, 2007; Thompson, 2004).

In Step 3, we define our criteria for factor extraction and examine how many factors we will retain. This process is based on the following 'rule of thumb' criteria: we will select a factor only, if; 1) they have an eigenvalue larger than 1 (Guttman-Kaiser rule); 2) they account for about 70-80% of the total variance; 3) they have their data point before the breaking point or elbow in a scree-plot ----------→

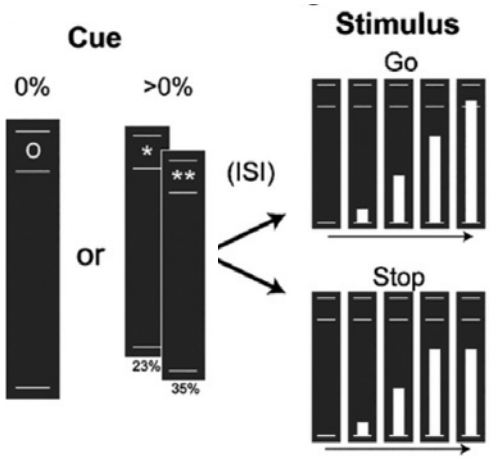


After factor extraction, we will check the produced communalities of the items, and possibly remove items if values are low (below 0.40), or retain more factors in order to provide a better account of the variance.

In Step 4, we will select the type of rotation. Rotation methods are either orthogonal or oblique. Simply put, orthogonal rotation methods assume that the factors in the analysis are uncorrelated, whereas oblique rotation produce factors that are correlated. First, we will try one oblique rotation method (direct oblimin and/or promax). If factors in the produced component correlation matrix have values over ±0.32 we will choose oblique over orthogonal, if not, we will use the orthogonal rotation method (the varimax rotation).

In Step 5, we will examine which variables are attributable to a specific factor, and give the factors a name or a theme. In this decision process, we take objective measures like the number of variables of a factor (minimum of 3, but preferably 5 or more) and the size of the loading (0.32 as minimum, preferably 0.50 or higher) into account. If items do not load or are unable to be assigned to a factor, we will decide whether items should ultimately be discarded.

Behavioral analysis of the fMRI stop-signal anticipation task (SSAT):



All participants will perform the Stop-signal anticipation task, adapted from Zandbelt and Vink (2010). First, response times (for GO trials) and accuracy will be calculated for each stop-signal probability level separately (0, 23, 35%). Cognitive control will be measured as the speed of inhibition, indicated by the stop-signal reaction time (SSRT). We will calculate the SSRT according to the integration method (Verbruggen, 2013), across the 3 stop-signal probability levels. A smaller SSRT reflects a faster speed of inhibition, indicating better cognitive control. Based on task performance we will make three groups: low-, average- versus high- performance, based on +/- 1 SD deviation from the mean, aiming to identify the underlying neural circuitry of cognitive control.

fMRI data analysis:

fMRI data will be analysed using Statistical Parametric Mapping 12 (SPM12) software. Preprocessing steps will follow the standardized pipeline by Matthijs Vink to harmonize our analyses with other studies in the consortium. In short, the pipeline includes slice time correction, realignment, coregistration of the anatomic image to the mean functional image, spatial normalization to a Montreal Neurologic Institute (MNI) template brain and smoothing (using a 6 mm full-width at half maximum (FWHM) Gaussian kernel) (for details: see application from Pascal Pas).

Subsequently, a GLM regression will be performed to estimate task effects on brain activation. Three regressors will be included to model brain activation related to successful stop trials, failed stop trials and go trials with stop-signal probability greater than 0%. Furthermore, response time and stop-signal probability will be included as parametric regressors for go trials. To correct for head motion, the 6 realignment parameters will be included as regressors of no interest.

Five contrasts will be created for each participant (adapted from van Rooij et al., 2014). For all five contrasts, mean activation levels will be extracted from predefined Regions of Interests (ROIs), using Marsbar, implemented in SPM12. The primary ROIs will be defined based on activation maps from an independent sample of healthy volunteers (children) who performed the SSAT in a previous study. Additional relevant ROIs involved in cognitive control, will be defined through a whole brain group analysis using t-tests for each contrast.

Environmental factors versus cognitive control versus brain activation analysis:

To investigate the effects of environmental factors on cognitive control abilities, we will perform a multiple regression analysis comparing the output factors from the Exploratory Factor Analysis, with the SSRT derived from the SSAT, and the performance on the Discount Delay Task.

To investigate whether environmental factors have an effect on brain activation underlying cognitive control, we will run a multiple regression design in SPM12, where we will add the output factors from the Exploratory Factor Analysis as covariances to the model, to correlate with the mean activation levels of the ROIs. All analyses will be corrected for multiple comparisons using the False Discovery Rate (Benjamini & Hochberg, 1994).

Planned subgroup analyses (if applicable. Substantiate your choices)

Additionally, we plan to test whether environmental factors (reflected by the output factors from the EFA) have differential effects on brain activation underlying cognitive control in the low-performance group compared to the high-performance group.

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.

We plan to repeat our analyses to test whether age, gender and intellectual abilities (IQ) influence the relation between environmental factors and cognitive control.

2. **Timeline and milestones (including dates of when to analyze/write up):**

   June-July 2019: organize and prepare data for analysis
   July-Dec 2019: data analysis
   Jan-Feb 2020: preparation of manuscript

3. **Output (e.g. article, report, etc.):**
   Output will be a research article

4. **Proposed authors + affiliations (please note that the YOUth data access committee can request certain authors to be included):**
   1. Bram Gooskens
   2. Dienke Bos
   3. Bob Oranje
   4. Pascal Pas
   5. Matthijs Vink
   6. Sarah Durston

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