Searching for replicable associations between cortical thickness and psychometric variables in healthy adults: empirical facts.

Shahrzad Kharabian Masouleh[1], Simon B. Eickhoff[1,2], Eliana Nicolaisen-Sobesky[2], Sarah Genon[1,2]

[1]Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Research Centre Jülich, Jülich, Germany
[2]Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany;
s.kharabian@fz-juelich.de@fz-juelich.de

Introduction

- Search for underlying neuronal structural correlates of variability in behavior (SBB) has a long history and its impact goes beyond the research community.

- Previously, we demonstrated that correlations between inter-individual variations in gray matter volume (GMV) and normal variation in behavior are predominantly spurious and could not be robustly evidenced within moderately sized healthy samples[1].

- However, GMV is shown to be an impure marker of gray matter structural changes, which may explain the spurious correlations.

- Here we use estimates of cortical thickness (CT) as marker of structural variations an assess empirical replication rates of behavioral correlates of CT.

Methods

Participants:
HCP[2]: 420 healthy young adults (50% female, age: 28 ± 3.7).

Behavioral data:
- Age for validation.
- 34 standard tests in the categories of Personality, emotion and cognition (cognitive flexibility, episodic and working memory. Composite scores, assessing crystallised and fluid intelligence, early childhood and general measure of cognitive functioning). In-scanner tasks : two-back working memory, language task as well as relational processing task.

Structural data:
CT estimates derived using FreeSurfer v5.3-HCP pipeline, registered to fsaverage surface and smoothed using a gaussian kernel of 15 mm FWHM.

Statistical analysis:
Replicability of whole brain exploratory SBB:
Association between each behavioral score and CT is assessed by fitting a linear model at each vertex, controlling for confounding effects of age, sex, education, using general linear model in PALM[3] with 1000 permutations. Inference is made at cluster-level, using TFCE ($p < 0.05$).

- This procedure is applied on 100 randomly generated subsamples, of same size (e.g. 50% of the original cohort) and binary maps of significant clusters are aggregated to identify rate of spatial overlap of significant findings for each behavioral score.

Replicability of ROI-based SBB:
For every subsample, an independent matched-sample is generated from the main cohort. Partial-correlation of behavioral score and average CT in the significant clusters are compared between the original and matched-subsamples. Replicated effects are defined based on three criteria:

- Same **direction** in the original and replication sample.
- Same **direction** + **Significant** ($p<0.05$)
- **Bayes factor**[4] (BF_{10}) ≥ 3
(H0: absence of SBB; H1: presence of SBB in the same direction as original effect.)

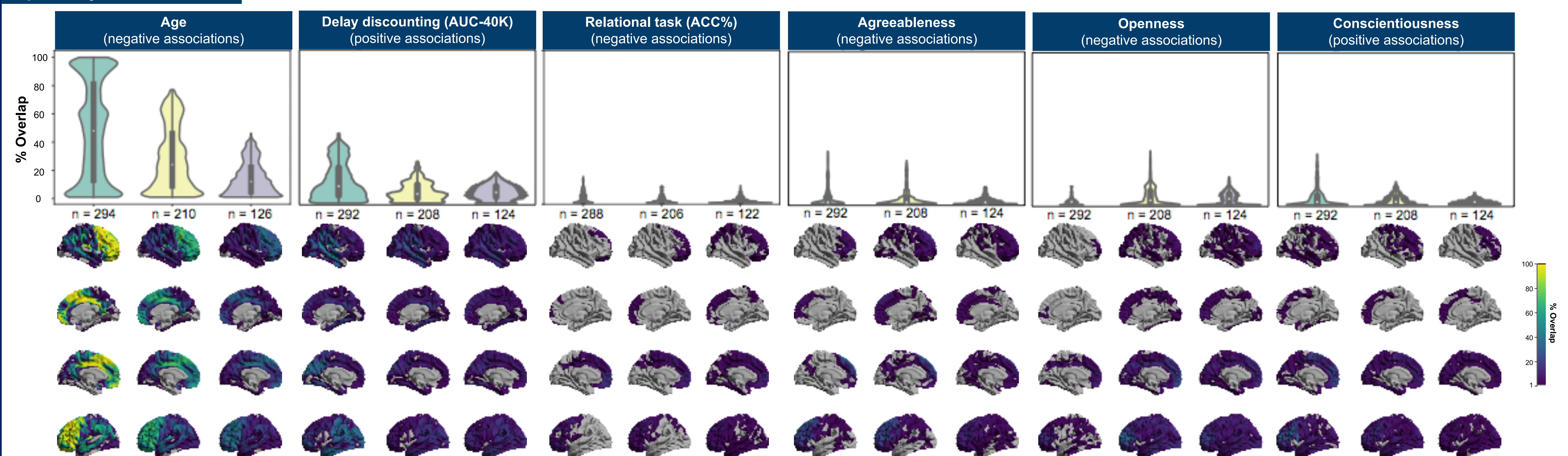## Results

Exploratory whole-brain SBB

Fig1. Frequency of spatial overlap (density plots and aggregate maps) of significant findings from exploratory analysis over 100 random subsamples, calculated for three different sample sizes (X-axis). Here the top five behavioral scores with the highest frequency of overlapping results for all tested sample sizes are demonstrated.
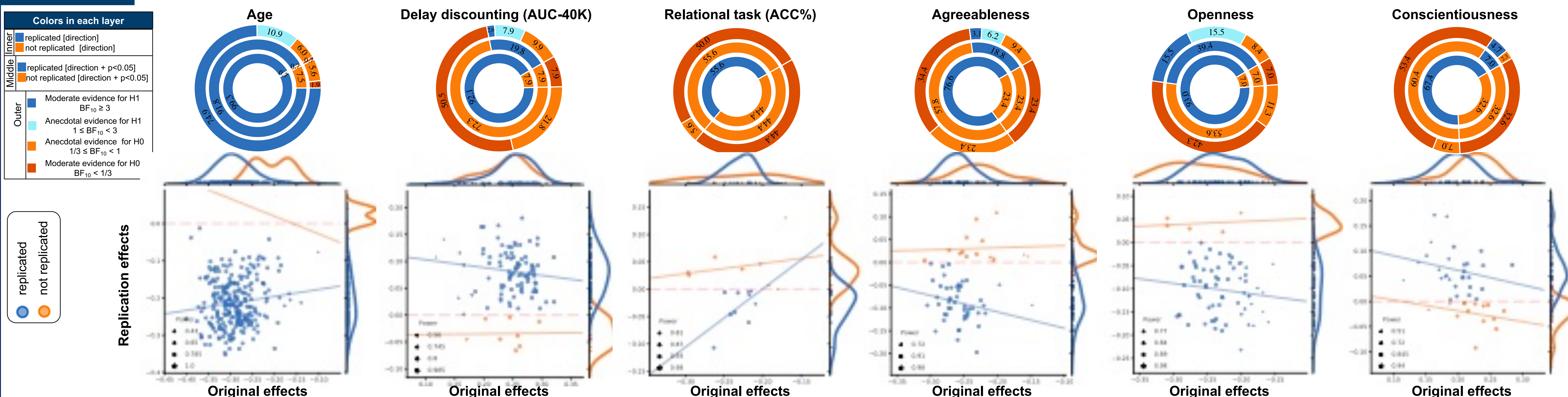
ROI-based SBB

Fig2. Upper row: Summary of ROI-based SBB-replication (% of ROIs) *using three different criteria: Inner layer: "direction". Middle layer: direction + Significant. Outer layer: BF_{10}.* Lower row: Original versus replication effects sizes for all ROIs from 100 splits*; (replication defined using "direction" only) and size of each point is proportional to the power of replication. Original and replication samples have equal size($n \approx 208$).*

Conclusions

- Similar to our findings with GMV, for most of the tested behavioral measures, we did not find any significant association with CT in more than 90% of exploratory analyses.

- ROI-based analysis: Low rate of "significant" replication; Lack of clear association between original and replication effect sizes; Over-estimation of effects size derived from *exploratory analysis in small samples*. This questions usefulness of small pilot studies with to define the expected effect size and create and estimation of sample size[5].

- Our empirical assessments show that irrespective of the chosen morphological marker, robust associations between brain structure and normal variation in behavior are very unlikely.

- Importantly, composite scores of cognitive functioning, as a measure of general intelligence, did not show robust associations with CT.

- Considering the publication bias[6], these results are alarming and demonstrates that harking[7] (hypothesizing after the results are known) in such context can result in serious misleading conclusions about the true neurobiological associations

- Thus reported SBB associations within small to medium samples of healthy individuals should be interpreted with caution, before their replication is demonstrated in independent datasets.

References: [1] Kharabian Masouleh, 2019. [2] http://www.humanconnectome.org/ [3] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM. [4] Boekel, et, al., 2015. [5] Button et. al., 2013. [6] Dwan et. al., 2008. [7] Kerr et. al., 1998.