A cross-cohort replicable and heritable latent dimension linking behaviour to multi-featured brain structure

Eliana Nicolaisen-Sobesky[1], Shahrzad Kharabian-Masouleh[1,2], Agoston Mihalik[3,4,5], Fabio Ferreira[3,4], Felix Hoffstaedter[1,2], Holger Schwender[6], Somayeh Maleki Balajoo[1,2], Sofie L. Valk[1,2,7], Simon B. Eickhoff[1,2], B.T. Thomas Yeo[8,9], Janaina Mourao-Miranda[3,4], 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; [3]Centre for Medical Image Computing, Department of Computer Science, University College London, London, United Kingdom; [4]Max Planck University College London Centre for Computational Psychiatry and Ageing Research, University College London, London, United Kingdom; [5]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom, 6Mathematical Institute, Heinrich Heine University Düsseldorf, Düsseldorf, Germany; [7]Otto Hahn Research Group "Cognitive Neurogenetics", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; [8]Department of Electrical and Computer Engineering, ASTAR-NUS Clinical Imaging Research Centre, Singapore Institute for Neurotechnology and Memory Networks Program, National University of Singapore, Singapore; [9]Centre for Cognitive Neuroscience, Duke-NUS Medical School, Singapore, Singapore.

e.nicolaisen@fz-juelich.de

Introduction

- Brain-behaviour associations are often not replicable[1,2,3].
- Our aim is to identify robust and replicable associations between interindividual variability in **brain structure and behaviour**.
- We use a multivariate method to link brain and behaviour: **Regularized Canonical Correlation Analysis** (RCCA)[4,5].
- To test the generalisability of the model, we use a novel machine learning framework that embeds RCCA in **multiple holdouts** of the data[4,6].
- Crucially, we tested the **cross-cohort replicability** of the brain-beahviour associations.
- In addition, the nature vs nurture influence on associations was studied with **heritability**.

Methods

1) Datasets:

- **HCP Young Adult (HCP-YA):** (S1200), n=1047 (560 females), age = 28.8, 22-37 years (mean, range)
- **HCP-aging (HCP-A):** n=601 (353 females), age= 58.5, 36-100 years (mean, range)

Brain data (639 measures):
- *Cortical Thickness* (CT)
- *Surface Area* (SA)
- *Grey Matter Volume* (GMV)

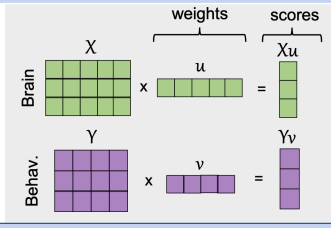
Behavioural data (32 measures):
- *Alertness*
- *Cognition*
- *Emotion*

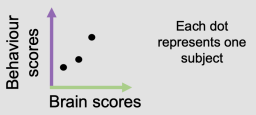
Age and gender were regressed out from brain and behaviour.

Processing of brain data:
- CT and SA: FreeSurfer v5.3.0-HCP in HCP-YA; v6.0 in HCP-A
- GMV: CAT12.5 (both cohorts)
- Averaged by regions (200 cortical[7], 32 subcortical[8], 7 cerebellar[9])
- Normalised within subjects by brain size

2) Regularized Canonical Correlation Analysis:

- X and Y: subjects are in rows, measures are in columns
- RCCA finds the weights which maximize the canonical correlation
- Regularization: L2-norm constraints on weights
- RCCA was applied independently in each cohort.

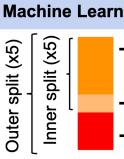




Brain loadings: corr(X, Xu)

Behaviour loadings: corr(Y,Yv)

3) Machine Learning Framework: Multiple holdouts[4,6]:



- 1) Train set (64%): Train several RCCA models with different regularization parameters
- 2) Test set (16%): Test models and select model with best generalizability
- 4) Hold-out set (20%): Test best model

3) Optimisation set (80%): Train best RCCA model

Code available at: https://github.com/anaston/cca_pls_toolkit

4) Statistical analyses:

- 1000 permutations, shuffling rows of Y, respecting family structure
- Omnibus hypothesis[6]: states that there is no association in any of the outer splits. Hence, if a significant association is found in at least one split, the null hypothesis can be rejected.

5) Replicability:

Cross-cohort similarity of loadings:
- Pearson's correlation for behaviour
- Spin test for CT and SA

6) Heritability:

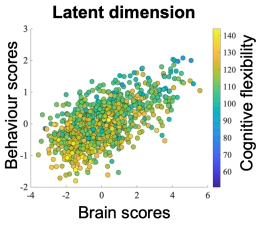
Heritability (h^2) and genetic correlation (ρ_g) of brain and behavioural scores were studied using a twin-based analysis in SOLAR-Eclipse[10] (HCP-YA).

Results

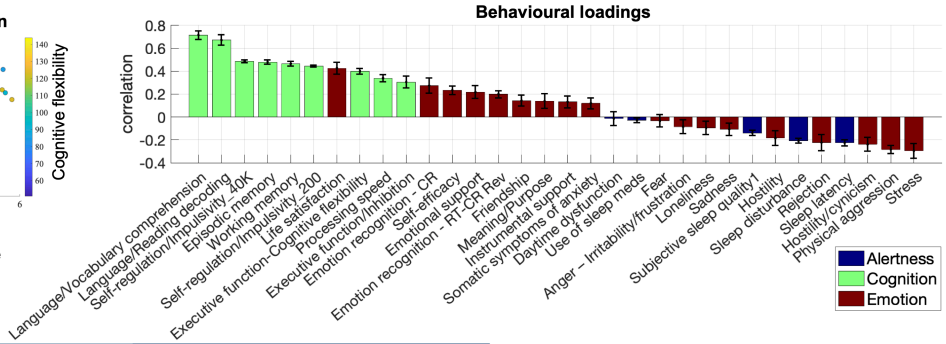
The RCCA model in the HCP-YA yielded 3 significant latent dimensions. The RCCA model in the HCP-A yielded 2 significant latent dimensions. Of these, only **one latent dimension replicated across cohorts**, with significant correlations on the behavioural (r=0.73, p<0.001), CT (r=0.81, p<0.001) and SA (r=0.56, p<0.001) loadings.

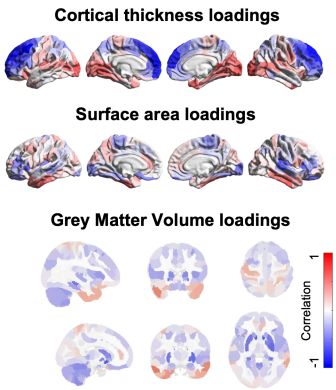
This latent dimension can be understood as an axis capturing:
- Behaviour: one pole of **good cognitive-control/executive-functions and positive affect**, and another pole of **cognitive dysfunction/impulsivity and negative affect**.
- Brain: one pole associated to areas related to **higher cognitive functions**, and another pole associated with **sensorimotor regions**.

Latent dimension in HCP-YA



Canonical correlation in the outer splits:
r_{range}=0.33-0.43,
$p_{omnibus}$=0.001



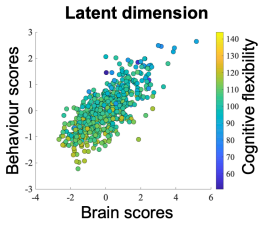


Shown loadings represent the average over the 5 outer splits.
Behaviour: Error bars depict one standard deviation.
CR: correct responses / RT-CT: reaction time to correct responses
GMV: Top row corresponds to MNI coordinates: -43.6, 16, 52.9. Bottom row corresponds to MNI coordinates: -10.3, -3.9, -9.1.

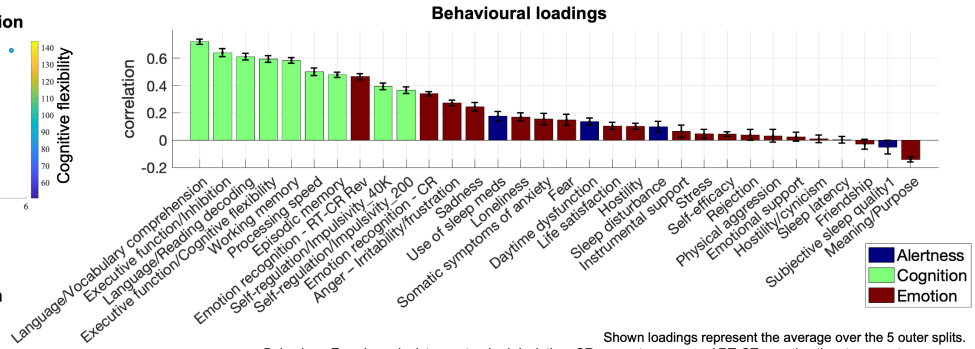
Heritability

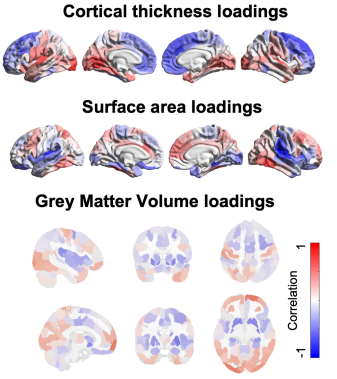
- Brain scores: h^2=0.85; p<0.000; Behavioural scores: h^2=0.823; p<0.000
- Genetic correlation: ρ_g=0.66; p<0.000

Latent dimension in HCP-A



Canonical correlation in the outer splits:
r_{range}=0.29-0.61,
$p_{omnibus}$=0.001





Shown loadings represent the average over the 5 outer splits.
Behaviour: Error bars depict one standard deviation. CR: correct responses / RT-CT: reaction time to correct responses
GMV: Top row corresponds to MNI coordinates: -43.6, 16, 52.9. Bottom row corresponds to MNI coordinates: -10.3, -3.9, -9.1.

Discussion

- Our results show one heritable and replicable latent dimension linking interindividual variability in brain structure with interindividual variability in behaviour.
- The behavioural profile of this latent dimension was previously captured in association with brain function in the HCP-YA cohort[11,12]. We extend such findings by analysing the brain structural profile of this dimension.
- This latent dimension can be interpreted as an axis in which subjects are spread based on their covariance between brain structure and behaviour.
- Crucially, we test the replicability of this latent dimension and extend it to a sample with a wider age range (HCP-A).
- We also show that the brain and behavioural scores of this latent dimension are at least partly influenced by overlapping genetic mechanisms.

References

1. Kharabian-Masouleh 2019. *Elife.*
2. Kharabian-Masouleh 2020. *bioRxiv.*
3. Marek 2022. *Nature.*
4. Mihalik 2020. *Biol Psychiatry*
5. Hardoon 2004. *Neural Computation*
6. Monteiro 2016. *J Neuroscience Methods*
7. Schaefer 2018. *Cereb Cortex*
8. Tian 2020. *Nature Neuroscience*
9. Yeo 2011. *J Neurophysiol*
10. http://www.solar-eclipse-genetics.org
11. Smith 2015. *Nature Neuroscience*
12. Han 2020. *Neuroimage*