*Preregistration*

**Functional brain representations of neuroticism**

**Content**

[**1 Introduction 1**](#_gjdgxs)

[1.1 Background 1](#_30j0zll)

[1.2 The present study 4](#_1fob9te)

[1.3 Hypotheses 4](#_3znysh7)

[**2 Methods & Data Analytic Strategy 4**](#_2et92p0)

[2.1 Dataset 4](#_tyjcwt)

[2.2 Psychological measures & structural equation modelling 5](#_3dy6vkm)

[*2.2.1 Outcome modelling of neuroticism 5*](#_1t3h5sf)

[*2.2.2 Other psychological constructs 5*](#_4d34og8)

[2.3 Neuroticism Pattern 6](#_2s8eyo1)

[*2.3.1 Data cleaning 6*](#_17dp8vu)

[*2.3.2 Training & Evaluation 6*](#_3rdcrjn)

[*2.3.3 Spatial pattern illustration 7*](#_26in1rg)

[*2.3.4 Pattern psychometrics 7*](#_lnxbz9)

[2.4 Comparison to other approaches 8](#_35nkun2)

[*2.4.1 Affective State Patterns 8*](#_1ksv4uv)

[*2.4.2 Network approach 8*](#_44sinio)

[*2.4.3 Region approach 9*](#_2jxsxqh)

[2.5 Additional Analyses 9](#_z337ya)

[**3 References 10**](#_3j2qqm3)

# Introduction

## Background

Around 30% of the world’s population are diagnosed with a mental disorder during their lifetime1. This high prevalence not only leads to substantial strain on healthcare systems2, but more importantly reflects enormous psychological suffering and life years lost due to disability3. While clinical research has expanded in recent decades, improving health care and reducing stigma, overall therapy efficacy appears to be mainly driven by intervention-unspecific (i.e., common) factors4 and remains unsatisfactory5. This issue has been increasingly attributed to shortcomings of current taxonomic systems for mental disorders6.

Taxonomies for mental disorders are essential to organize research on etiology, mechanisms, and treatments in a way that is practically useful for healthcare providers and practitioners6. Unfortunately, the high comorbidity and shared etiology between mental disorders as well as heterogeneity within disorder categories pose fundamental challenges to clinical research and question the utility of current taxonomies, namely the ICD-10 and DSM-56,7. This problem is also evident in research on the biology of mental disorders, as most genetic and brain-based markers cut across disorder boundaries8,9.

Recent research initiatives aim to develop novel diagnostic systems through basic dimensions of individual differences. These attempts include the reformulation of mental disorders in terms of the big five traits10 and constructing a hierarchical organization of specifically clinical traits (HiTOP)7. Similarly, both the ICD-11 and DSM-5 moved towards dimensional trait-based assessment of personality disorders. All these approaches heavily rely on dimensionality reduction techniques applied to self-/other-report measures (e.g. factor analysis). In contrast, the Research Domain Criteria (RDoC) initiative of the NIMH follows a long-term basic science approach where continuously varying core dimensions of human functioning are investigated on all possible levels of analysis, from molecules over brain networks to self-reports and behaviors11. This additional focus on biological levels provides a potential link to pharmacological and animal research. Moreover, identifying the physical systems underlying the RDoC dimensions would alleviate the criticism of self-report-based approaches as language-dependent tautological redescriptions of behavior. A long-term interdisciplinary approach might yield more universal and stable nosological dimensions and classes of mental disorders12.

Individual differences in the tendency to experience negative emotions—which we further relate to as “negative affectivity”—are a core feature of psychopathology, represented in all approaches above. The big five trait *neuroticism*—also termed emotional instability or vulnerability—is a fundamental dimension of personality with particular relevance for both mental and physical health13–15. Individuals high in neuroticism have an increased tendency to experience negative emotions like anger, anxiety, or depression16. It has been estimated that the global costs of high neuroticism exceed those of all common mental disorders combined17. Virtually all mental disorders coincide with elevated neuroticism, albeit to a different degree14 and likely due to different functional roles18 depending on the specific disorder. Moreover, neuroticism shares genetic antecedents with most mental disorders8. Most importantly, neuroticism is highly correlated with negative valence system measures of the RDoC as well as the affective instability trait in the DSM-5 personality disorder system (*r* ≥ .66, not corrected for (un)reliability)19 and maps on the internalizing spectrum of HiTOP 7. Hence, all the proposed systems include a similar major domain which represents the propensity to experience negative emotions, highlighting the transferability of research findings between systems concerning this feature.

Trait negative affectivity largely manifests in increased reactivity to stressful or threatening environments20. Naturally, neuroimaging research on negative affectivity *traits* has focused on individual differences in the responsiveness of neural systems related to negative affective *states*, most commonly induced with pictures of negative scenes or facial expressions. The proposed neural systems have differed in their level of complexity and spatial scale, including (1) single regions, (2) large-scale networks, and (3) unconstrained whole-brain patterns. All these approaches come with a set of limitations and their relative utility for understanding negative affectivity has not been systematically explored.

Several brain regions have been reported to show heightened responsiveness to negative stimuli in individuals with high negative affectivity21. Of these brain regions, the amygdala has received the most attention, likely due to its prominent role in fear conditioning22. Amygdala reactivity to threat-related pictures is increased in anxiety and affective disorders23–25 as well as individuals with exposure to childhood adversity26,27, a major risk factor of trait negative affectivity18. Therefore, amygdala reactivity is often interpreted as a direct indicator of negative affectivity. This is likely an oversimplification, as the amygdala plays a broader role in the processing of relevant stimuli independent of valence28–32. Moreover, amygdala reactivity neither predicts self-reported affective states following negative pictures above chance33 nor was its association with trait neuroticism meta-analytically confirmed34. Therefore, the reverse inference from observed amygdala reactivity to affective states or traits is not warranted35. Similar criticisms apply to other candidate regions, like the anterior insula and the dorsal anterior cingulate cortex (dACC)33,36.

An alternative perspective is that specific mental processes are performed by distributed networks instead of single regions. Resting state fMRI revealed that subsets of voxels over the whole brain are synchronously active as orchestrated networks37. These networks represent useful summaries of communicating neuron populations which might jointly enact a single process. Most research in this area has focused on connectivity measures between voxels, regions, or networks during rest and showed aberrations in mental disorders9,38. Still, the existence of these networks can also be observed in voxel coactivation patterns during specific tasks, which were used to characterize their function39. The salience network has received most attention as a neural system relevant to affective processing40, including affective disorders41. It comprises all the single regions above as central network hubs (anterior insula, dACC, amygdala,) and has been shown to be a viable target for neurofeedback therapy in major depression42. Nevertheless, a one-to-one mapping of single networks to complex mental phenomena is again likely not warranted as single networks appear to be “domain general”40. The salience network specifically is not an indicator of affective responses33 or discrete emotions43 and its network hub regions do not show abberant coactivation in affective disorders or high neuroticism34. Hence, as for single regions, activity of the salience network and its component regions during threat-processing should not easily be taken to reflect individual differences in negative affectivity.

Notably, two more complex network approaches emerged, which either view a mental process as the sum of activations across all networks or as the emergent consequence of complex interactions between several38 or all networks44. For example, a recent study suggests that both salience and default mode network form an interoceptive network, which is implicated in the experience of affect by constructivist emotion theories45.

Multi-voxel pattern analysis (MVPA) was developed to solve the problems of limited accuracy and specificity of neural measures. Seminal papers have shown that information is stored as complex activity patterns of neuron populations, instead of a one-to-one mapping of regions or neurons to complex features46,47. Therefore, mental events can be “decoded” by regressing their occurrence on the activity of voxels across the brain. The regression weights represent the magnitude and direction of effects in contributing voxels. While this approach can also be applied to single regions, a decoding model of subjective pain has shown that accuracy increased when models comprise voxels from the whole-brain, outperforming single networks47.

The occurrence of discrete emotions can be decoded from whole-brain patterns with relatively high accuracy43,48. The frequency of these negative emotional pattern responses during resting state were also associated with the respective negative emotional traits49. Interestingly, these patterns do not correspond to canonical resting-state parcellations, revealing potential shortcomings of approaches focusing on canonical resting-state networks. Similarly, affective responses to negative pictures can be decoded with high accuracy, while the performance of concrete regions and networks remained insufficient33. Hence, affective-emotional MVPA patterns might be viable building blocks for new insights into dimensions for mental disorders and emerging classification systems12. Still, a major shortcoming of this approach is that patterns can be tuned to stimulus features of the task which are not primarily of interest or simply capture confounding processes47. For example, a pattern which accurately predicts affective self-reports following negative pictures33 might capture affective *picture* *content*, instead of affective *experiences*. Moreover, even if a pattern can decode affective experiences across multiple diverse experimental settings, it is unclear whether people with higher pattern expressions in response to negative experimental stimuli will also be more reactive to threats or tonically experience more negative emotions outside the scanner in everyday life. This could represent a lack of ecological validity of typical fMRI tasks, rather than preclude the existence of hyperactive neural systems in people scoring high on negative affectivity traits. Still, it might be possible to probe hyperactive neural systems related to negative affectivity, even if they are not the most predictive systems for affective states induced in laboratory tasks. A pattern which directly predicts negative affectivity traits from task-based fMRI data could set a benchmark for how much variance can be explained using this approach, reveal commonalities and differences with patterns optimized for negative affective states as well as region or network approaches, and could serve as a fundamental building block to understand the biology of mental disorders.

## The present study

The goal of the present study is to develop a brain pattern which predicts negative affectivity traits from the two most common task-based fMRI paradigms to induce negative emotions (negative faces/negative scenes), assess its psychometric properties, and compare its performance to three classical approaches: (1) brain patterns for affective states, (2) canonical networks, and (3) single regions.

We use neuroticism to operationalize negative affectivity, as it is one of the most established constructs in this regard, was rediscovered by different researchers throughout history, represents one of the major big five factors of personality, has a vast body of psychometric and neuroimaging research behind it, correlates highly to similar measures in novel dimensional assessment systems for mental disorders, and is rooted in the objective methodology of the psycholexical approach. We computed neuroticism factor scores from structural equation models, which are measurement error-free representations of a latent construct and therefore can lead to larger standardized effect sizes (for a simulation, see: <https://m-clark.github.io/docs/lv_sim.html>).

## Hypotheses

1. A multivariate pattern whole-brain pattern can explain variance in neuroticism
2. A pattern from the IAPS paradigm will explain more variance than the face matching paradigm
3. The pattern will explain more variance in neuroticism and negative affect than in extraversion and positive affect
4. The neuroticism pattern will reflect neuroticism facets to a varying degree. It will explain more variance in withdrawal aspect-related facets, including depression, anxiety, and vulnerability
5. The multivariate pattern will outperform neural signatures of affective states, coarse network approaches, and single regions

# Methods & Data Analytic Strategy

## Dataset

Data from two studies will be used: The Adult Health and Behavior project – Phase 2 (AHAB-2) and the Pittsburgh Imaging Project (PIP). Both contain data for neuroticism (NEO PI-R questionnaire), an fMRI facial expression matching paradigm (FEMP) and an fMRI paradigm showing pictures of scenes with negative content (IAPS).

Only derivative fMRI data will be used, namely the beta images for the contrasts (a) negative stimuli versus rest and (b) neutral stimuli versus rest.

Inclusion and exclusion criteria for the fMRI task will be consistent with prior studies using the same data50, leading to 430 expected participants for the face matching task minus one exclusion due to anatomical abnormalities. Of 339 unique participants who did the IAPS task, one will be excluded due to insufficient brain coverage, leading to an expected sample size of 338 for this task. For fifteen people who participated in both studies, the AHAB-2 data will be used. Participants will be included in the present study if they belonged to either of these two fMRI samples *and* provided data on the NEO PI-R.

Maurizio Sicorello will conduct the statistical analyses and will not have access to the dataset before the preregistration is uploaded and timestamped in an online repository. All other project members have worked with the datasets previously, but will not conduct further analyses related to this project before preregistrations.

## Psychological measures

### Outcome modelling of neuroticism

Participants filled in the NEO PI-R questionnaire which consists of 240 items, while one or two informants also rated the participant using the 60-item abbreviated form (the NEO Five-Factor Inventory, or NEO-FFI). The neuroticism factor and its facets will be scored according to the test manual.

### Other psychological constructs

Extraversion, positive, and negative affect will be used to assess convergent and discriminant validity of the trained neural pattern. Positive and negative affect were assessed with the positive and negative affect schedule and will be scored according to the test manual.

## Neuroticism Pattern

A multi-voxel brain pattern will be trained to predict neuroticism. We follow the guidelines by54

### Data cleaning and preparation

Person-wise mahalanobis distance will be calculated for brain activation during negative stimuli versus rest. Participants will be excluded if their brain activation represented a significant multivariate outlier on a bonferroni-holm corrected alpha = .05.

Voxel-activation data and neuroticism will be z-standardized.

### Model development

Patterns will be trained separately for the two fMRI tasks using identical procedures.

2.3.2.1 Hold-out samples

Of participants who contributed data to both tasks, a hold-out sample of N = 100 will be drawn for later unbiased model evaluation. This sample has a power of .90 to detect a true correlation between pattern and neuroticism of r = .29 in a one-tailed test, which is a typical correlation observed between personality traits and behavioral data. Notably, larger hold-out samples would not necessarily lead to higher power, as they would reduce the sample size of the training data, which might in turn lead to lower effect sizes.

2.3.2.2 Algorithm

We will use partial least squares regression (PLS) to predict neuroticism from contrast images. The algorithm will consist of an outer 5-fold cross-validation loop (stratified for neuroticism) to assess model generalization and an inner 5-fold cross-validation loop to determine the number of components to be retained, using the Bayesian hyperparameter optimization procedure (described in the following preprint: <https://www.biorxiv.org/content/10.1101/2020.07.04.182873v2>), restricting the sampling space for the number of components to lay between 1 and R, where R is the rank of the training data matrix. These procedures will be repeated twice to account for randomness in fold slicing during cross-validation. Reported accuracies will be averaged over both repetitions.

2.3.2.3 Accuracy measures

The predictions generated in the validation samples of the outer loop will be concatenated across all folds and correlated with the actual neuroticism scores as a generalizable indicator of accuracy. Squaring the correlation indicates the variance in neuroticism explained by the pattern.

2.3.2.4 Unconditional modelling decisions

To determine the model with the highest accuracy, the following fixed modelling choices will be explored:

1. Contrast choice:
   1. Negative stimuli versus neutral stimuli   
      (neutral scenes for the IAPS task and shapes for the face matching task)
   2. Negative stimuli versus rest (as responses to neutral stimuli might also increase with higher neuroticism; Shackman, …)
2. Within-person standardization choice
   1. Brain-wise z-standardizing activity
   2. Brain-wise centering activity
   3. No brain-wise standardization
3. Effect size standardization choice
   1. beta-values
   2. t-values

The model with the best accuracy will be considered the winning model. Training and evaluation of this model will be repeated within a permutation regimen to assess its statistical significance, based on 1000 iterations and one-tailed alpha = .05.

2.3.2.5 Conditional modelling decisions

If correlation accuracy of the final model is not significantly different from zero, the following options will be explored:

1. Using alternative algorithms
   1. Support Vector Regression
   2. Principle component regression
2. Using only the full data from the AHAB-II study, which contains other-reports for neuroticism, prediction accuracy for different neuroticism scores will be compared, (based only on 60 items which overlap between self- and other-report):
   1. self-report neuroticism as criterion
   2. other-report neuroticism as criterion
   3. latent factor scores for neuroticism as criterion, calculated for both self- and other reports based on the procedure in51
3. Outcomes below the factor level of neuroticism will be used as outcomes:
   1. Neuroticism aspects “withdrawal” and “volatility”, computed from the factor scores in61.
   2. If aspects show no significant accuracy, facet scores will be predicted, scored based on the test manual.

2.3.2.6 Hold-out sample performance

Using the winning model, the neuroticism pattern expressions will be calculated in the hold-out sample as the dot-product between voxel-wise brain activity and regression weights. These expression values will be correlated with the neuroticism scores. Correlations will be accompanied with p-values, confidence intervals, and Bayes factors to assess the evidence for/against the null hypothesis, using the BayesFactor package in R with default noninformative priors52. The result will be interpreted both continuously and according to prespecified decision rules. P-values below .05 will be considered statistically significant; Bayes factors above 3 will be considered substantial evidence, Bayes factors above 10 strong evidence53.

### 2.3.2.7 Pattern psychometrics

Reliability and validity of the patterns will be assessed in the full samples, comprising the training and hold-out samples.

*Reliability***.** Pattern reliability will be assessed using the split half method reported in50. Contrast images were calculated separately for the first and the second half of the respective task. Based on this data, pattern expressions will be calculated by taking the dot product of the trained pattern weight matrix and the contrast image of interest for each participant. Hence, there will be one pattern expression value (i.e. predicted neuroticism value) per participant for each task-half. The spearman-brown corrected correlation between the pattern expressions of both task-halves will be taken as an index of internal consistency.

*Construct validity.*Construct validity reflects the degree to which a measure captures what it is supposed to capture (and not other distinct constructs). We will assess the correlation of the patterns with different psychological constructs, comprising extraversion, negative affect and positive affect. We expect a positive correlation with negative affect and small but slightly negative correlations with extraversion and positive affect55. Moreover, the neuroticism pattern expression will be correlated with scores on the neuroticism facets to test whether it predominantly reflects some of the facets.

### Feature-level assessment

2.3.3.1 Bootstrap

Pattern weight maps will be illustrated with bootstrapping-based statistical thresholds. 10,000 bootstrap samples will be drawn with replacement from the datasets used for cross-validation (i.e. not the hold-out dataset[s])54. The model will be retrained on each bootstrap sample and the weight maps saved. The weight-distribution will be z-standardized at each voxel to derive *p*-values for a point null hypothesis, which will be corrected with FDR of q < .05 (whole-brain corrected).

2.3.3.2 Lesion study

Using the procedure described in54 , we will test the predictive utility of the pattern for the hold-out sample after removing (a) single networks of the seven network parcellation and (b) single brain regions, following the parcellations given in section 2.4.3.

**2.3.4 Neurobiological Assessment**

2.3.4.1 Overlap with functional networks

Overlap of the thresholded pattern with seven functional resting state networks will be calculated based on the networks and procedure described in54

2.3.4.2 Functional decoding

The unthresholded regression weight map will be subjected to a decoding analysis using neurosynth.org to determine the psychological terms most strongly related to the pattern.

## Comparison to other approaches

To simplify the comparisons between the trained trait neuroticism pattern and other neural markers, we will use the contrast and standardization of the neuroticism pattern models with the highest accuracy, identified by the procedure above.

### Affective State Patterns

Several patterns have been developed to predict negative emotional experiences, consisting of regression weight maps which can be used for prediction in new data. We will calculate pattern expressions for each participant by taking the dot product of weight map and the participant-specific contrast images. This pattern expression will then be correlated with the neuroticism scores.

We will use:

(1) the unthresholded PINES signature, which is a weight map for the prediction of picture induced negative affect33.

(2) the unthresholded negative emotion patterns from Kragel and LaBar (2015)48, comprising “fear”, “anger”, and “sadness”. A latent factor will be modelled for the pattern expressions (in terms of distance from the classification hyperplane) for the three emotions using three-indicator confirmatory factor analysis with unrestricted path coefficients and uncorrelated errors. Afterwards, we will also inspect the correlations between neuroticism and the three negative emotion pattern expressions separately, correcting p-values for three tests using the bonferroni-holm procedure.

All these analyses will be conducted on the full available samples.

Moreover, our trained weight-maps for neuroticism will be correlated with the affective state patterns to quantify their similarity.

### Network approach

We will create seven binary masks corresponding to the seven resting-state networks identified by Yeo et al. (2011)37 and average the activity within those networks. Several analyses will be conducted to predict neuroticism from these networks, which vary in complexity:

(1) We will correlate the activity in single networks with neuroticism using the seven network parcellation. P-values will be corrected using the Bonferroni-Holm method for 8 comparisons.

(2) We will predict neuroticism from all seven networks within a multiple regression.

(3) We will predict neuroticism from all seven networks using unconditioned random forest regression to accommodate interactions between networks, using the cforest function of the R-package *party* with default unbiased control parameters56. Predictions will be based on out-ouf-bag samples as a more generalizable performance measure. P-values for accuracy will be calculated by randomly permuting the outcome 1000 times and refitting the model.

All these analyses will be conducted on the full samples.

### Region approach

We will predict neuroticism from activity in single regions. These regions include:

(a) Amygdala

(b) The anterior insula

(c) The dorsal anterior cingulate cortex

There will be three analyses branches, which again vary in complexity:

(1) Neuroticism will be predicted from average activity in the defined regions with bivariate correlations, separately for lateralization. P-values of the resulting six tests will be corrected using the bonferroni-holm procedure.

(2) Neuroticism will be predicted from patterns within these a priori regions, one region at a time. We will use the same procedure as for the trait neuroticism pattern. In addition to accuracy, we will report the number of positive weights in the resulting pattern.

(3) We will use a data-driven best-region approach to see how large the correlation between the average activity in the most predictive region and neuroticism is. The sample will be randomly split in half, stratified for neuroticism. The first half will be used for discovery of the region most strongly correlated to neuroticism, the second half will be used to replicate the correlation between this region and neuroticism.

# References

1. Steel, Z. *et al.* The global prevalence of common mental disorders: A systematic review and meta-analysis 1980-2013. *Int. J. Epidemiol.* **43**, 476–493 (2014).

2. Trautmann, S., Rehm, J. & Wittchen, H. U. The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? *EMBO Rep.* **17**, 1245–1249 (2016).

3. Whiteford, H. A. *et al.* Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575–1586 (2013).

4. Leichsenring, F. & Steinert, C. Is cognitive behavioral therapy the gold standard for psychotherapy? The need for plurality in treatment and research. *JAMA - J. Am. Med. Assoc.* **318**, 1323–1324 (2017).

5. Dragioti, E., Karathanos, V., Gerdle, B. & Evangelou, E. Does psychotherapy work? An umbrella review of meta-analyses of randomized controlled trials. *Acta Psychiatr. Scand.* **136**, 236–246 (2017).

6. Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E. & Reed, G. M. Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health’s Research Domain Criteria (RDoC). *Psychol. Sci. Public Interes.* **18**, 72–145 (2017).

7. Kotov, R. *et al.* The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J. Abnorm. Psychol.* **126**, 454–477 (2017).

8. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science (80-. ).* **360**, (2018).

9. Buckholtz, J. W. & Meyer-Lindenberg, A. Psychopathology and the Human Connectome: Toward a Transdiagnostic Model of Risk For Mental Illness. *Neuron* **74**, 990–1004 (2012).

10. Widiger, T. A. & Trull, T. J. Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *Am. Psychol.* **62**, 71–83 (2007).

11. Kozak, M. J. & Cuthbert, B. N. The NIMH Research Domain Criteria Initiative: Background, Issues, and Pragmatics. *Psychophysiology* **53**, 286–297 (2016).

12. Woo, C. W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: Brain models in translational neuroimaging. *Nat. Neurosci.* **20**, 365–377 (2017).

13. Widiger, T. A. & Oltmanns, J. R. Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry* **16**, 144–145 (2017).

14. Kotov, R., Gamez, W., Schmidt, F. & Watson, D. Linking ‘Big’ personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychol. Bull.* **136**, 768–821 (2010).

15. Caspi, A. *et al.* The p Factor. *Clin. Psychol. Sci.* **2**, 119–137 (2014).

16. McCrae, R. R. & John, O. P. An Introduction to the Five-Factor Model and Its Applications. *J. Pers.* **60**, 175–215 (1992).

17. Cuijpers, P. *et al.* Economic Costs of Neuroticism. *Arch. Gen. Psychiatry* **67**, 1086 (2010).

18. Ormel, J. *et al.* Neuroticism and common mental disorders: Meaning and utility of a complex relationship. *Clin. Psychol. Rev.* **33**, 686–697 (2013).

19. Gore, W. L. & Widiger, T. A. Negative emotionality across diagnostic models: RDoC, DSM-5 section III, and FFM. *Personal. Disord. Theory, Res. Treat.* **9**, 155–164 (2018).

20. Bolger, N. & Schilling, E. A. Personality and the Problems of Everyday Life: The Role of Neuroticism In Exposure and Reactivity to Daily Stressors. *J. Pers.* **59**, 355–386 (1991).

21. Shackman, A. J. *et al.* Dispositional negativity: An integrative psychological and neurobiological perspective. *Psychol. Bull.* **142**, 1275–1314 (2016).

22. LeDoux, J. The amygdala. *Curr. Biol.* **17**, 868–874 (2007).

23. Etkin, A. & Wager, T. D. Functional Neuroimaging of Anxiety: A Meta-Analysis of Emotional Processing in PTSD, Social Anxiety Disorder, and Specific Phobia. *Am. J. Psychiatry* **164**, 1476–1488 (2007).

24. Hamilton, J. P. *et al.* Functional neuroimaging of major depressive disorder: A meta-analysis and new integration of baseline activation and neural response data. *Am. J. Psychiatry* **169**, 693–703 (2012).

25. Schulze, L., Schmahl, C. & Niedtfeld, I. Neural Correlates of Disturbed Emotion Processing in Borderline Personality Disorder: A Multimodal Meta-Analysis. *Biol. Psychiatry* **79**, 97–106 (2016).

26. Heany, S. J. *et al.* The neural correlates of Childhood Trauma Questionnaire scores in adults: A meta-Analysis and review of functional magnetic resonance imaging studies. *Dev. Psychopathol.* **30**, 1475–1485 (2018).

27. Hein, T. C. & Monk, C. S. Research Review: Neural response to threat in children, adolescents, and adults after child maltreatment – a quantitative meta-analysis. *J. Child Psychol. Psychiatry Allied Discip.* **58**, 222–230 (2017).

28. Sander, D., Grafman, J. & Zalla, T. The Human Amygdala: An Evolved System for Relevance Detection. *Rev. Neurosci.* **14**, 303–316 (2003).

29. Ousdal, O. T. *et al.* The human amygdala is involved in general behavioral relevance detection: Evidence from an event-related functional magnetic resonance imaging Go-NoGo task. *Neuroscience* **156**, 450–455 (2008).

30. Adolphs, R. Fear, faces, and the human amygdala. *Curr. Opin. Neurobiol.* **18**, 166–172 (2008).

31. Cunningham, W. A. & Brosch, T. Motivational salience: Amygdala tuning from traits, needs, values, and goals. *Curr. Dir. Psychol. Sci.* **21**, 54–59 (2012).

32. Sergerie, K., Chochol, C. & Armony, J. L. The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* **32**, 811–830 (2008).

33. Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A. & Wager, T. D. A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biol.* **13**, 1–28 (2015).

34. Servaas, M. N. *et al.* Neuroticism and the brain: A quantitative meta-analysis of neuroimaging studies investigating emotion processing. *Neurosci. Biobehav. Rev.* **37**, 1518–1529 (2013).

35. Poldrack, R. A. Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding. *Neuron* **72**, 692–697 (2011).

36. Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J. & Barrett, L. F. The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature. *Cereb. Cortex* **26**, 1910–1922 (2016).

37. Thomas Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

38. Menon, V. The connectome-Feature Review Large-scale brain networks and psychopathology: a unifying triple network model. *Trends Cogn. Sci.* **15**, (2011).

39. Menon, V. *Salience Network*. *Brain Mapping: An Encyclopedic Reference* **2**, (Elsevier Inc., 2015).

40. Barrett, L. F. & Satpute, A. B. Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Curr. Opin. Neurobiol.* **23**, 361–372 (2013).

41. Schulze, L., Schulze, A., Renneberg, B., Schmahl, C. & Niedtfeld, I. Neural Correlates of Affective Disturbances: A Comparative Meta-analysis of Negative Affect Processing in Borderline Personality Disorder, Major Depressive Disorder, and Posttraumatic Stress Disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **4**, 220–232 (2019).

42. Hamilton, J. P. *et al.* Effects of salience-network-node neurofeedback training on affective biases in major depressive disorder. *Psychiatry Res. - Neuroimaging* **249**, 91–96 (2016).

43. Wager, T. *et al.* A Bayesian Model of Category-Specific Emotional Brain Responses. *PLoS Comput Biol* **11**, 1004066 (2015).

44. Lindquist, K. A. & Barrett, L. F. A functional architecture of the human brain: emerging insights from the science of emotion. *Trends Cogn. Sci.* **16**, 533–540 (2012).

45. Kleckner, I. R. *et al.* Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nat. Hum. Behav.* **1**, (2017).

46. Haxby, J. V. Multivariate pattern analysis of fMRI: The early beginnings. *Neuroimage* **62**, 852–855 (2012).

47. Kragel, P. A., Koban, L., Barrett, L. F. & Wager, T. D. Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging. *Neuron* **99**, 257–273 (2018).

48. Kragel, P. A. & Labar, K. S. Multivariate neural biomarkers of emotional states are categorically distinct. doi:10.1093/scan/nsv032

49. Kragel, P. A., Knodt, A. R., Hariri, A. R. & LaBar, K. S. Decoding Spontaneous Emotional States in the Human Brain. *PLoS Biol.* **14**, 1–19 (2016).

50. Gianaros, P. J. *et al.* Affective brain patterns as multivariate neural correlates of cardiovascular disease risk. *Soc. Cogn. Affect. Neurosci.* 1–39 (2020). doi:10.1093/scan/nsaa050

51. Wright, A. G. C., Creswell, K. G., Flory, J. D., Muldoon, M. F. & Manuck, S. B. Neurobiological Functioning and the Personality-Trait Hierarchy: Central Serotonergic Responsivity and the Stability Metatrait. *Psychol. Sci.* **30**, 1413–1423 (2019).

52. Morey, R. D. & Rouder, J. N. BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2. (2018).

53. Jarosz, A. F. & Wiley, J. What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *J. Probl. Solving* **7**, 1037–1040 (2014).

54. Kohoutová, L. *et al.* Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nat. Protoc.* doi:10.1038/s41596-019-0289-5

55. Steel, P., Schmidt, J. & Shultz, J. Refining the Relationship Between Personality and Subjective Well-Being. *Psychol. Bull.* **134**, 138–161 (2008).

56. Strobl, C., Hothorn, T. & Zeileis, A. Party on! *R J.* **1**, 14 (2009).

57. Amunts, K. *et al.* Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. *Anat. Embryol. (Berl).* **210**, 343–352 (2005).

58. Eickhoff, S. B. *et al.* A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* **25**, 1325–1335 (2005).

59. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).

60. Derringer, J. A SIMPLE CORRECTION FOR NON-INDEPENDENT TESTS. (2018). doi:10.31234/osf.io/f2tyw

61. DeYoung, C. G., Quilty, L. C. & Peterson, J. B. Between Facets and Domains: 10 Aspects of the Big Five. *J. Pers. Soc. Psychol.* **93**, 880–896 (2007).

**Notes to self**

**Do I have to incorporate uncertainty in constructing the SEM? maybe via bootstrapping?**

**Preliminary discussion points**

+

Biological substrates for major trait dimensions “stabilize” factor analytic solutions with convergent evidence from a non-language related domain, getting us closer to a “periodic table” if individual differences.

Approach balances variance explained with interpretability

-

All the differences between these approaches are artificial: A pattern can be within a region. A network can be a subset of important regions. A network can be a pattern etc.

Neuroticism represent the p-factor of psychopathology and does not distinguish between mental disorders, but only between people with and without mental disorders.

More research is needed to tell whether a brain measure is a (temporary) consequence, risk factor or feature of psychopathology or shares a common cause (Ormel et al., 2013)

Potentially additional information when we look at the shape of the response, beyond its amplitude.

Negative affect characteristics explained by three pathways: stressor reactivity, tonic increases in negative affect, and increased stressor exposure. We likely target the first one and potentially the second one.

Relationship between this approach and resting state functional connectivity approaches

*Also read:*

*- to parcel or not to parcel.*

*-Poldrack, R. A., Huckins, G., & Varoquaux, G. (2019). Establishment of Best Practices for*

*Evidence for Prediction: A Review. JAMA Psychiatry. doi:10.1001/jamapsychiatry.2019.3671*

*-Scheinost (2019). Ten simple rules for predictive modelling of individual differences in neuroimaging*

*-Kasey Stanton (2020): Increasing Diagnostic Emphasis on Negative Affective Dysfunction: Potentially Negative Consequences for Psychiatric Classification and Diagnosis*

*-Luke Chang wrote: “This regularization shrinks beta parameters to zero. Components with a nonzero beta were selected and then refit using OLS to ensure that they were not unduly affected by the shrinkage (see [85]).”*

*Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction. 2nd ed. New York: Springer; 2009*

-tor meta-analyses of emotions paper

-barrett 2019 interoception paper

-barrett 2017 theory paper

-touroutoglu 2015 barett

-cunningham 2010/deyoung

-Personality and Psychopathology: A Stagnant Field in Need of Development

-Dimensions of Normal Personality as Networks in Search of Equilibrium: You Can't Like Parties if You Don't Like People

-Biomarkers in Psychiatry - A Critique and

-“Toward Robust Anxiety Biomarkers: A Machine Learning Approach in a Large-Scale Sample”: Failed attempt based on resting state fMRI.

Davis, J., Maes, M., Andreazza, A., McGrath, J. J., Tye, S. J., & Berk, M. (2015). Towards a classification of biomarkers of neuropsychiatric dis- ease: From encompass to compass. Molecular Psychiatry

-How can functional neuroimaging inform cognitive theories? Perspect. Psychol. Sci. 8, 98–103

Methods:

-Wang, Y., Fan, Y., Bhatt, P., Davatzikos, C., 2010. High-dimensional pattern regression 2108 using machine learning: from medical images to continuous clinical variables. 2109 NeuroImage 50, 1519–1535. h

* Absatz über neuronale Grundlagen von negativer emotionalität schreiben

Shackman (2016)

One of the most basic questions concerns the nature of the relations between trait-like differences in temperament and more transient emotional experiences and behaviors (Barlow, Sauer-Zavala, Carl, Bullis, & Ellard, 2013; Epstein, 1994). As the pioneering psychologist David Funder noted, traits “describe patterns and consistencies in behavior, but they don’t explain where those patterns and consistencies come from” (Funder, 1994; p. 126)

**Krishnan 2010 partial least squares neuroimaging tutorial**

**Add to intro: The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders Compared With Diagnosis-Specific Protocols for Anxiety Disorders: A Randomized Clinical Trial**

**Add to intro: Identification of a Common Neurobiological Substrate for Mental Illness**