


RESEARCH ARTICLE

WILEY

Network and state specificity in connectivity-based predictions of individual behavior

Nevena Kraljević^{1,2}  | Robert Langner^{1,2} | Vincent Küppers^{2,3} |
 Federico Raimondo^{1,2} | Kaustubh R. Patil^{1,2} | Simon B. Eickhoff^{1,2} |
 Veronika I. Müller^{1,2}

¹Institute of Systems Neuroscience, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

²Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Research Centre Jülich, Jülich, Germany

³Department of Nuclear Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

Correspondence

Nevena Kraljević and Veronika I. Müller,
 Institute of Systems Neuroscience, Medical
 Faculty and University Hospital Düsseldorf,
 Heinrich Heine University Düsseldorf,
 Düsseldorf, Germany.

Email: nevena.kraljevic@hhu.de and v.mueller@fz-juelich.de

Funding information

Horizon 2020 Framework Programme,
 Grant/Award Number: 945539 (HBP SGA3);
 Helmholtz Portfolio Theme “Supercomputing
 and Modeling for the Human Brain”

Abstract

Predicting individual behavior from brain functional connectivity (FC) patterns can contribute to our understanding of human brain functioning. This may apply in particular if predictions are based on features derived from circumscribed, a priori defined functional networks, which improves interpretability. Furthermore, some evidence suggests that task-based FC data may yield more successful predictions of behavior than resting-state FC data. Here, we comprehensively examined to what extent the correspondence of functional network priors and task states with behavioral target domains influences the predictability of individual performance in cognitive, social, and affective tasks. To this end, we used data from the Human Connectome Project for large-scale out-of-sample predictions of individual abilities in working memory (WM), theory-of-mind cognition (SOCIAL), and emotion processing (EMO) from FC of corresponding and non-corresponding states (WM/SOCIAL/EMO/resting-state) and networks (WM/SOCIAL/EMO/whole-brain connectome). Using root mean squared error and coefficient of determination to evaluate model fit revealed that predictive performance was rather poor overall. Predictions from whole-brain FC were slightly better than those from FC in task-specific networks, and a slight benefit of predictions based on FC from task versus resting state was observed for performance in the WM domain. Beyond that, we did not find any significant effects of a correspondence of network, task state, and performance domains. Together, these results suggest that multivariate FC patterns during both task and resting states contain rather little information on individual performance levels, calling for a reconsideration of how the brain mediates individual differences in mental abilities.

KEYWORDS

brain-based prediction, brain–behavior relationships, fMRI, functional connectivity, interindividual differences, machine learning

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Human Brain Mapping* published by Wiley Periodicals LLC.

Practitioner Points

- Better prediction of behavior from task versus resting-state functional connectivity (FC) only in a cognitive domain.
- Little evidence for specificity of state, network, or task similarity.
- Predicting complex behavior based on FC remains a significant challenge.
- We extend research on brain-based behavior prediction beyond the cognitive domain.

1 | INTRODUCTION

No two individuals are alike in perception, affect, thought, and behavior, but also brain structure and function. A major goal of neuroscience is uncovering the relationships between these dimensions by investigating individual differences. An approach that has recently become popular is predicting individual behavior, affective characteristics or cognitive abilities from brain data (Gao et al., 2019; Greene et al., 2018; Kong et al., 2019; Larabi et al., 2021; Nostro et al., 2018; Ooi et al., 2022; Rosenberg et al., 2020; Sasse et al., 2023; Shen et al., 2017). Such predictive modeling is thought to yield important insights about generalizable brain-behavior relationships and is considered a crucial step toward personalized medicine (Mueller et al., 2013; D. Wang et al., 2015).

A number of studies in this area have shown that, for example, regional gray-matter volume and structural connectivity significantly predict age (Cole et al., 2017; Franke et al., 2012; More et al., 2023), reading comprehension (Cui et al., 2018), inhibitory control (N. He et al., 2020), or fear of pain (X. Wang et al., 2019). Similarly, functional neuroimaging data has also been reported to predict different behaviors or traits, ranging from personality (Dubois, Galdi, Han, et al., 2018; Nostro et al., 2018) or life satisfaction (Itahashi et al., 2021) to cognitive abilities such as creative thinking (Zhuang et al., 2021), cognitive flexibility (Chén et al., 2019), or working memory (WM) capacity (Stark et al., 2021).

While a variety of different brain characteristics have been employed as features to predict behavior, one of the most widely used measures (Yeung et al., 2022) is resting-state (in the following also “rest” or “REST”) functional connectivity (FC) obtained from functional magnetic resonance imaging (fMRI). Some recent studies, however, suggest that behavior prediction may benefit from the use of *task-based* FC, as compared to resting-state data (Avery et al., 2020; Greene et al., 2018; Jiang et al., 2020; Rosenberg et al., 2016, 2016; Stark et al., 2021). For example, Sripada and colleagues found that the correlation between predicted and observed scores of a general cognitive ability factor improved when using FC from the 2-back WM-task state ($r = .50$), as compared to using resting-state FC ($r = .26$; Sripada et al., 2019, 2020). A similar pattern has been reported for the prediction of different measures of attention (Yoo et al., 2018) as well as for the prediction of intelligence based on FC from tasks taxing executive functioning (L. He et al., 2021) or attention (Rosenberg et al., 2016).

Importantly, all the studies mentioned above showed an improvement in prediction performance for task-fMRI data derived from the

same domain as the predicted measure (i.e., prediction of stop-signal task performance based on FC derived from stop-signal task-fMRI data). However, there is not only resting versus task states but rather different task states depending on which task is performed during fMRI data acquisition. That is, every task performed in the scanner can be thought of as eliciting a specific state. Interestingly, it has been shown that in predicting intelligence, using almost any other task state (i.e., fMRI acquired during a WM task as well as an emotion task) or task-rest combinations outperforms using resting-state FC only (Gao et al., 2019; Greene et al., 2018, 2020; Sripada et al., 2020).

Based on the concept of convergent and discriminant validity (Campbell & Fiske, 1959; Schumann et al., 2022), it would be expected, however, that connectivity patterns observed during the same or a similar task, hence coming from the same domain as the predicted target behavior (i.e., representing the same state; convergent validity), lead to better prediction performance than do patterns observed during a task state from an unrelated domain (discriminant validity). In line with this idea, recent studies reported better accuracies for predicting general cognitive ability (Sripada et al., 2020) and fluid intelligence (Gao et al., 2019) from FC during task states involving executive functions (“same-domain”), as compared to prediction from unrelated task or resting states (“other-domain”). This improvement was particularly pronounced when FC data of the cognitively demanding WM task was used, as compared to task states from other domains (although the authors did not test for the statistical significance of the observed numerical differences between prediction accuracies). These examples suggest the possibility of state specificity when predicting behavior from corresponding FC patterns.

While most studies predicted task performance from states of the same domain (i.e., prediction of intelligence from FC of a WM task state), others predicted task performance from FC observed during the exact same task. Avery et al. (2020), for example, predicted individual performance accuracy in an n-back WM task based on FC derived from fMRI data obtained while the n-back task was performed, which showed an increase in accuracy when using this task's fMRI data, as compared to rest data (Avery et al., 2020). Building on this, Stark and colleagues investigated the difference in prediction accuracy between predictions of performance in different working and episodic memory tasks from FC obtained while performing an n-back WM task. Importantly, the highest prediction accuracy ($r = .36$) was achieved when n-back task performance was predicted using FC during the very same task (i.e., n-back performance measured in the MR scanner), followed by the prediction of performance in a different WM task (list sorting; $r = .24$), followed by predictions

of episodic memory performance scores ($r = .05-.11$) (Stark et al., 2021). These results suggest a specificity benefit of the state used for calculating FC, which goes beyond the prediction of ability in a given broadly defined cognitive domain (i.e., WM) and narrows it down to specific tasks (i.e., n-back versus list sorting). That is, beyond the effect of state specificity (i.e., domain congruence benefits for prediction accuracy), a state–target similarity effect (i.e., task congruence benefits for prediction accuracy) should manifest in even better prediction accuracies for the performance in tasks during which FC data were acquired, as compared to the performance in other tasks of the same domain.

The literature to date is inconclusive regarding the effects of state specificity and state–target similarity on FC-based predictions of mental abilities and psychological traits. In particular, the majority of studies investigating such prediction models have focused on cognitive targets such as intelligence or attention (Yeung et al., 2022). Thus, clear evidence for state specificity and state–target similarity is still lacking, especially in domains like emotion processing and social cognition.

The studies above mainly used whole-brain FC for behavior prediction. Sometimes, *post hoc* examination of the most predictive features from the whole-brain feature space are used for better interpretability (e.g., Chén et al., 2019; Dubois, Galdi, Paul, & Adolphs, 2018; Itahashi et al., 2021; Jiang et al., 2020; Pläschke et al., 2020). However, such *post hoc* analyses come with their own limitations, as feature weights are context-dependent, their reliability is rather low, and the results can be highly specific to the given dataset (Tian & Zalesky, 2021). Besides predicting behavior from whole-brain FC, several studies reported on predictions using particular functional networks as priors (J. Chen et al., 2021; Heckner et al., 2023; Nostro et al., 2018). That is, prediction models in these investigations exclusively rely on FC between regions that show activation during a given task. It is argued that this aides and constrains the functional interpretability of any observed associations (e.g., most predictive features), since such models are based on brain regions for which the association between brain and mental function has already been established independently.

Therefore, network-based prediction has the advantage of better interpretability of the results due to the a priori knowledge about the mental function a given network subserves (Nostro et al., 2018; Pläschke et al., 2017). Similar to state specificity, FC within networks associated with functions that are more closely related to the target behavior (e.g., predicting WM performance from WM network features) should also be more informative than networks that are associated with very different functions (e.g., predicting WM performance from pain network features). Few studies have investigated this network specificity, with some suggesting some network specificity with regard to personality (Nostro et al., 2018), but others showing a lack of specificity (Heckner et al., 2023; Pläschke et al., 2020).

The current project, therefore, aimed to investigate the influence of brain state (same- vs. other-domain), similarity of target behavior to the features within one domain (same vs. similar task from same domain), and functional network priors (same- vs. other-domain

network) on the predictability of individual behavior. This included the specific question of whether FC from same-domain states and in same-domain networks can predict individual behavior better than FC from other-domain states or networks. Hence, we tested the following three hypotheses: (1) State specificity: behavior should be better predicted based on FC patterns observed in the same domain, hence during the state corresponding to the behavior to be predicted, as compared to FC patterns observed in other (non-corresponding) domains. (2) State–target similarity: task performance should be better predicted based on FC patterns observed during the exact same task, as compared to another similar task from the same domain. (3) Network specificity: behavior should be better predicted based on FC patterns observed in the networks corresponding to the predicted behavior, as compared to FC patterns in other (non-corresponding) networks.

2 | METHODS

To investigate whether there is state specificity, state–target similarity and/or network specificity in brain–behavior prediction, we used the Human Connectome Project (HCP) Young Adult dataset. We divided it into two samples: in the first sample, we defined networks, and in the second sample, we computed FC in predefined networks from the first sample during different task states. Using FC within each network as features, we predicted six different target variables, matching the selected states and networks. We included the following three phenotypic domains: WM, theory of mind/social cognition (SOCIAL), and emotion processing (EMO).

2.1 | Samples

Data were obtained from the Young Adult S1200 release of the publicly available database provided by the HCP (Van Essen et al., 2013), which comprised data from 1206 healthy individuals. We only included participants for whom all the data required for our analyses were available. That is, (a) all four resting-state fMRI scans; (b) fMRI data of the WM, SOCIAL, and EMO tasks; and (c) the performance measures (accuracy and reaction time) of these three tasks performed in the scanner, as well as (d) all the performance measures we aimed to predict for tasks that were performed outside the scanner for each domain. Hence, every subject was required to have both (in-scanner and out-of-scanner) tasks per domain (three domains, six tasks in total). Of the 1206 individuals, 180 participants were excluded due to missing imaging data and 71 due to data quality issues. We further excluded subjects with accuracy below 50% in the six tasks of interest ($n = 77$). Performance accuracy was measured as the percentage of correct trials. We chose to include only subjects producing more than 50% correct trials, to ensure that only participants were included who were attentive during the task and hence present the given states we aimed to investigate. From the remaining sample of 878 subjects,

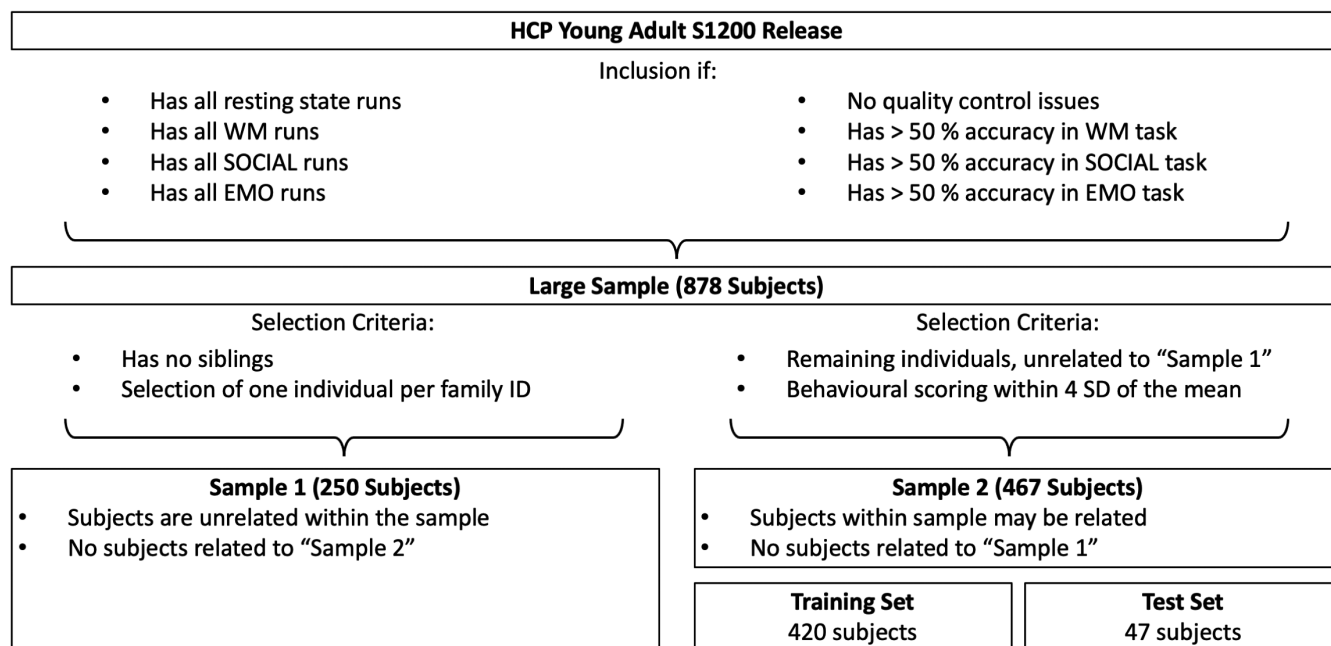


FIGURE 1 Overview of the sampling procedure.

two nonoverlapping subsamples were randomly generated: one for independently delineating task-based networks via general linear modeling (GLM; sample 1) and one for brain-behavior prediction within those (and other) networks (sample 2). Thus, the first sample can be thought of as the sample for “network extraction” and the second sample for “feature extraction and prediction.” We carefully accounted for the family structure and resulting dependencies in this dataset by ensuring that (i) sample 1 contained only one individual per family, (ii) there was no kinship between the two samples, and (iii) a leave-n-family-out cross-validation (CV) scheme for prediction analyses within sample 2 was used (Poldrack et al., 2020). For an overview of the sample selection, see Figure 1.

Sample 1, used for delineating task networks, consisted of 250 unrelated subjects (138 females; age mean = 28.6 years, standard deviation [SD] = 3.8, range = 22–36 years). Sample 2, used for prediction, consisted of all the remaining individuals with no siblings in the first sample. Further, we removed individuals that scored higher/lower than 4 SDs from the mean in any of the six target scores of interest, leaving us with 467 participants (252 females, mean age = 28.8 years, SD = 3.7, range = 22–36 years) for sample 2. Note that this sample contains siblings, which was accounted for in the prediction pipeline through family subsampling in the CV (Poldrack et al., 2020). From sample 2 we further randomly selected a holdout sample (47 subjects), which was not used in any of the CVs. Therefore, sample 2 consisted of 420 individuals that were used for CV and final training, while $n = 47$ participants were held back for subsequently testing generalizability.

The analyses of the HCP data were approved by the ethics committee of the Medical Faculty at the Heinrich Heine University Düsseldorf.

2.2 | Network delineation

Two different approaches were employed for delineating task-specific networks: (i) networks reflecting brain activation in a large sample of participants during the tasks of interest using the task fMRI data of sample 1, and (ii) activation likelihood estimation (ALE) meta-analyses across previously published neuroimaging results of the same tasks. For brevity, we here only report the methods and results of the first approach to network delineation. Further details on the results of the second, meta-analytic, approach can be found in the supplementary material.

2.2.1 | Delineation of task-networks in sample 1

Ultimately, our network extraction approach aimed to delineate networks that were as closely as possible related to the states we aimed to predict in the second sample. For this, we included strictly only the task of interest for network delineation in both approaches—the single study and the meta-analyses. To cover a variety of domains, we chose three very different tasks performed in the scanner: n-back for the WM domain, emotion recognition/face processing for the EMO domain, and social cognition/theory of mind for the SOCIAL domain. For details on the tasks, see Barch et al. (2013). Briefly, an n-back task was used for WM, presenting a sequence of different stimuli with the instruction to either decide whether the current stimulus is the same as the one used two trials ago (2-back) or to recognize a specific target (0-back). EMO was a face-matching task in which angry or fearful faces had to be matched (EMO condition), in contrast to matching shapes (neutral condition). In the SOCIAL task, animated moving

shapes were shown in either interacting or random manner and had to be labeled subsequently as interacting or randomly moving.

For the delineation of the task-specific networks revealing nodes that are activated in our task of interest, we used the minimally preprocessed volumetric task fMRI data of participants of sample 1. The preprocessing included artifact removal, motion correction, and registration to the MNI standard volume space. More details regarding the preprocessing pipeline can be found in Glasser et al., 2013. The minimally preprocessed data were input for the GLM, performed using FSL (Version 5.0.9) (Jenkinson et al., 2012; Smith et al., 2004; Woolrich et al., 2009). For the subject-level GLM we modified the scripts provided by the HCP (Barch et al., 2013; <https://github.com/Washington-University/HCPpipelines>), which are based on the FSL FEAT module (Woolrich et al., 2001, 2004), for use of volumetric data.

The subject-level GLM included for either run (two different phase encoding directions) temporal high-pass filtering (200 s cutoff), spatial smoothing (8 mm FWHM Gaussian kernel), and the GLM fitting. The respective stimuli in each task were modeled as blocked predictors, temporal derivatives of each predictor, 6 movement parameters and their derivatives as regressors of no interest. For each task, linear contrasts between conditions were computed: 2-back > 0-back for WM; interaction > random for SOCIAL; and faces > shapes for EMO, respectively. Data across both phase encoding directions were then combined with a fixed-effects GLM analysis.

For the group-level GLM, we modified an FSL workflow developed by (Esteban et al., 2019), which estimated the group effects using FSL FMRIB's Local Analysis of Mixed Effects by performing a one-sample *t* test across subjects. Group-level activation maps were thresholded at cluster-level $p < .05$ (FWE-corrected for multiple comparisons) with a cluster-forming threshold of $p < .001$. The resulting activation maps can be seen in the Supplemental Figures S1–S3. From these clusters, we extracted only the peak coordinates from gray matter with a minimum distance of 15 mm. This resulted in three networks: WM-NW, SOCIAL-NW, and EMO-NW. For an overview of the workflow, please see Figure 2. For comparison with the task-specific networks, we used FC between the Power nodes (Power et al., 2011) as a functionally defined, spatially distributed, whole-brain representation of the connectome. The Power nodes represent a combination of resting-state FC ROIs and task-based meta-analytic ROIs, yielding 264 nonoverlapping independent ROIs.

2.3 | Prediction in sample 2

2.3.1 | Targets: Behavioral measures

To assess state specificity and state–target similarity, we selected behavioral performance during different tasks: First, we used performance collected in the scanner for our three domains of interest (WM, SOCIAL, EMO → “same task”/in-scanner task). Second, we selected scores of tasks/questionnaires that measured behavior not exactly in the same state but still in the same behavioral domains

(“similar task”/out-of-scanner task). The two levels of tasks (“same” and “similar”) from the same domain enable us to advance insights beyond state specificity, into state–target similarity.

For “same task” scores (in-scanner task), reaction time and accuracy of task performance were used. These two scores were combined by calculating the Inverse Efficiency Score (IES; Townsend & Ashby, 1983), which is defined as the mean response time across correct trials of the condition of interest divided by its accuracy. This was employed to address the issue of ceiling effects in the accuracy scores. Hence, for WM (subsequently called “n-back”), IES was calculated using mean response time and accuracy of the 2-back blocks. For EMO, we used response time and accuracy in the face-block of the emotional face-matching task (subsequently called “matching” or “EMO matching”). For SOCIAL, since the accuracy in both interaction and random trials involved theory-of-mind cognition (Castelli et al., 2000), we averaged response time and accuracy of both interaction and random trials before creating the IES (subsequently called “Social Cognition”).

For “similar task” scores (out-of-scanner task or questionnaire scores) in the WM domain, we selected the unadjusted list sorting score from the NIH Toolbox List Sorting Working Memory Test (subsequently called “List Sorting”). For SOCIAL, we computed a social satisfaction compound score (Babakhanyan et al., 2018) across five different scales (friendship, loneliness, emotional support, instrumental support, and perceived rejection) of the self-report Emotion Battery of the NIH Toolbox (Salsman et al., 2013) (subsequently called “Social Satisfaction” or “Satisfaction”). For EMO, we computed the IES using reaction time and accuracy of the Penn Emotion Recognition Test (Gur et al., 2001, 2010) (subsequently called “Emotion Recognition” or “Recognition”). See supplementary Table S1 for an overview of all targets included.

2.3.2 | Features: FC

Resting-state fMRI and the three sets of task fMRI data (WM, SOCIAL, EMO) from sample 2 were used for calculating FC within each network of interest (WM, SOCIAL, EMO, and Power). The network extraction is explained in detail in the section “Delineation of task-networks in Sample 1”; for an illustration of the methods applied, please see Figure 2. For all four states we used all runs available (four runs for REST and two runs each for the tasks) and their full duration per run. MRI protocols of HCP were previously described in detail (Glasser et al., 2013; Van Essen et al., 2013). For the task-fMRI data, we used the minimally preprocessed version provided by the HCP, which includes removal of spatial distortions, volume realignment, registration to the anatomical image, bias field reduction, normalization to the global mean, and masking the data with the final brain mask (Glasser et al., 2013). The approach to treat task fMRI comparable to resting state fMRI data has been suggested by (Greene et al., 2020). For the resting-state fMRI data, we used the ICA-FIX denoised data provided by the HCP, which uses the minimally preprocessed fMRI data (processed in the same way as task fMRI data) as input and

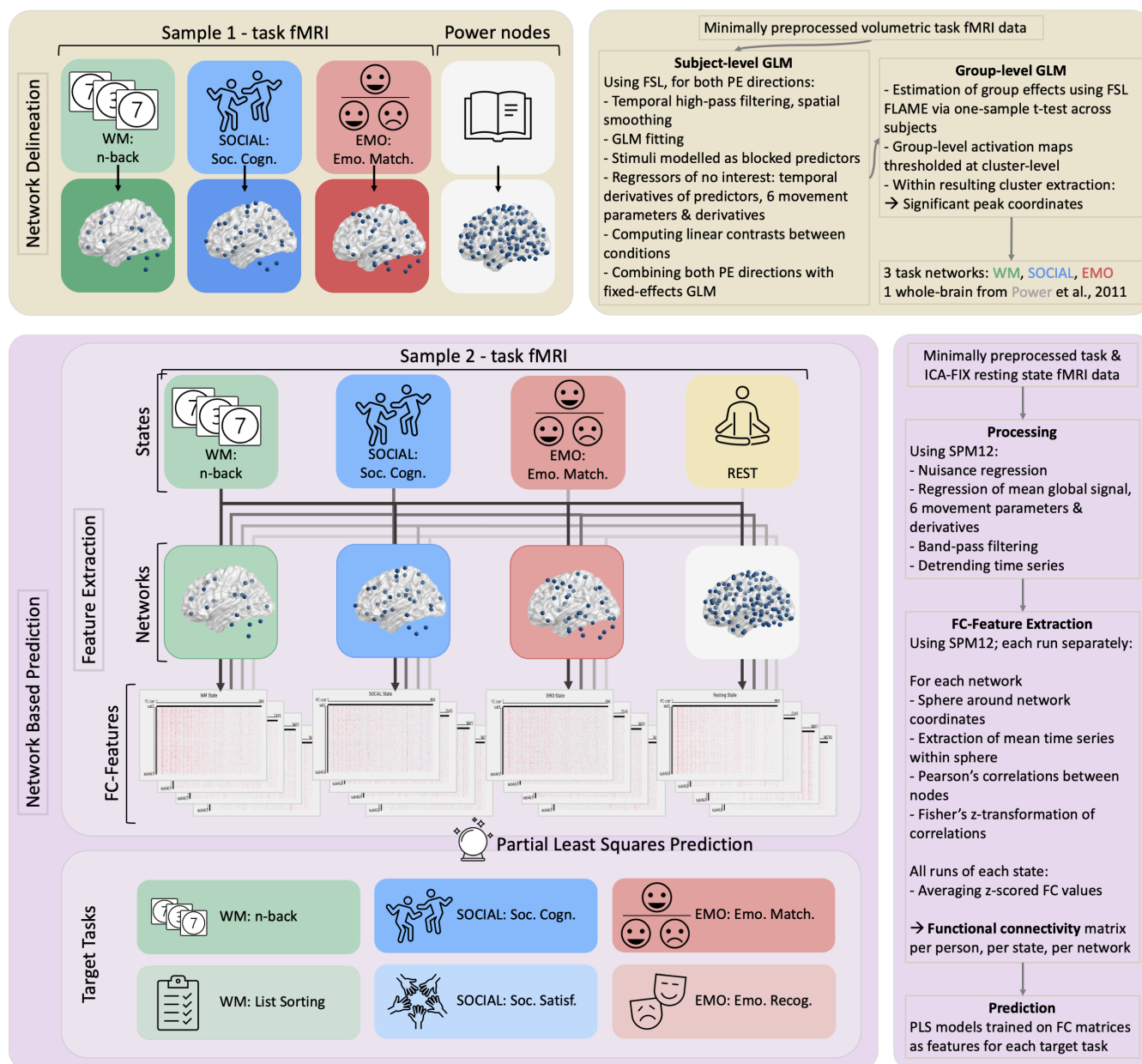


FIGURE 2 Overview of the applied methods. Yellow blocks depict the network extraction from sample 1. Violet blocks depict the network-based prediction in sample 2, together with the feature extraction (functional connectivities) from the networks delineated in the first step and sample. The upper heatmap under “FC-Features” shows the FC from the different states in the WM-network. GLM: General linear modeling, PE: Phase encoding; FC: Functional connectivity; Soc. Cogn., Social cognition task (in-scanner task); Soc. Satisf., Social Satisfaction Questionnaire (out-of-scanner score); Emo. Match., Emotional Face-Matching task (in-scanner task); Emo. Recog., Emotional Face Recognition task (out-of-scanner score).

denoises it through classification of ICA components. This classifier identifies “good” and “bad” components and automatically removes artifactual or “bad” components. For further details, see Griffanti et al. (2014), Salimi-Khorshidi et al. (2014), and Smith et al. (2013).

Additional processing as well as the FC analysis for both resting-state and task-fMRI data were performed using SPM12 (www.fil.ion.ucl.ac.uk/spm/software/spm12/) and MATLAB 2020a (The MathWorks, Natick, MA). Nuisance regression was done to control for mean white matter and cerebrospinal-fluid signals, mean global signal,

within-scanner motion using the 6 movement parameters and their derivatives stored in the Movement_Regressors_dt.txt file provided by the HCP. Further, we applied band-pass filtering [0.01–0.1 Hz] and detrended the time series. We opted for the band-pass filter, as this has been shown to be successful in filtering out movement and physiological artifacts, without leading to information loss (Ciric et al., 2017; Satterthwaite et al., 2013). Using the network coordinates obtained from sample 1 (depicted in Figure 2 and in the supplemental material in Figures S7–S13 and Tables S2–S7), for each network, we

modeled a 5-mm sphere around each node's coordinate. The sphere size was the same for all coordinates to ensure the same number of voxels within each node. However, as we extracted multiple peak coordinates from larger clusters of task-activation in sample 1, larger activated regions are represented by multiple spheres. From each sphere, we extracted the mean time series. We then calculated the Pearson's correlations between all pairs of nodes of each respective network, before applying Fisher's Z-transformation. These steps were done for each run separately (four runs for REST and two runs each for the tasks) and for each state (REST, WM, SOCIAL, and EMO). The z-scored FC values were finally averaged across all runs of each state. This was done for all four networks as well as for each of the four states. Connectivity matrices for the WM network can be seen in Figure 2, all other networks averaged across all subjects can be found in the supplemental material (Figures S14–S20).

To ensure that any effects were not due to the different lengths of the tasks performed in the scanner, we trimmed all time series to the length of the shortest scan duration (EMO: 2:16 min) for a control analysis. These results are reported in the supplemental material (Figure S22).

2.3.3 | Network-based prediction of individual behavior

We predicted the task performance/characteristics for each domain from resting- and task-state FC (four states: REST, WM, EMO, SOCIAL; to investigate state specificity) and task of interest (same and similar tasks in WM, EMO, SOCIAL; to further investigate state–target similarity) of the delineated networks (four networks: whole-brain Power nodes and three task networks [WM, SOCIAL, EMO from sample 1; to investigate network specificity]).

For the main analysis, we used partial least squares (PLS) as the prediction model. PLS is a form of supervised learning which uses linear regression fitting, but it can handle violation of the assumption of no multicollinearity by reducing the dimensionality of correlated variables. However, to confirm our results and to cover models that have been used in the past for behavioral prediction, we additionally performed analyses using other algorithms: Kernel ridge regression, which, like PLS, is a linear parametric model. As well as support vector regression (with both linear and nonlinear RBF kernel) and random forest as nonparametric models, where both can capture nonlinear relationships. Finally, we used PLS and kernel ridge also with connectivity-based prediction modeling (CBPM; Finn et al., 2015; Shen et al., 2017) as a popular feature reduction. CBPM correlates the features to the target variable, retaining only the features showing a significant relation to the target for the model to learn. Note that all models were trained using the same set of FC features and target variables. All results from the additional analyses are presented in the supplemental material (Figures S21–S28).

Separate prediction analyses were conducted for each combination of network, state, and behavioral score, resulting in a total of 4 states \times 4 networks \times 6 targets = 96 predictions. The FC pattern

of the respective network and state constituted the given feature space, and the respective behavioral scores were the targets. All algorithms were used as implemented in JuLearn (Hamdan et al., 2023), which is a toolbox based on scikit-learn (Pedregosa et al., 2011). It includes hyperparameter tuning, nested-CV, and feature reduction methods making sure that data leakage is avoided. For PLS, we tuned the hyperparameters in an inner fivefold CV, with the number of latent components increasing in steps from 1 to 10. As having a sibling in the training set could lead to a better prediction of the related participant's score in the validation set, we applied a 100 \times leave-30%-families-out CV scheme on 420 subjects from sample 2 to account for the family structure of the sample (i.e., individuals from the same family were not split into training or validation sample but kept in either one of them). This is done to counter potential nonindependence induced by the family structure in the HCP dataset (Poldrack et al., 2020). We deconfounded the features by regressing out age and sex as well as normalizing them by z-transformation. For comparability of prediction performance between the different behavioral scores, we additionally normalized the targets. To avoid data leakage, confound regression and normalization were done within the CV. That is, the confound regression models and parameters for z-transformation were computed in the CV on the training set only (70% of the families) and then applied to the test set (30% of the families) (Poldrack et al., 2020). Prediction performance was evaluated by the root mean squared error (RMSE) as well as the coefficient of determination (COD), as a measure of goodness of model fit, averaged across all CV runs. Additionally, the mean Pearson correlation across all CV runs between predicted and observed scores was calculated. After hyperparameter tuning and CV, we finally applied the model, that has been trained on all the data provided and with the hyperparameter tuning performed on it, to the randomly drawn holdout sample (47 subjects from sample 2) to evaluate the model's generalizability.

The RMSE was used for testing for significant differences of prediction performance between states, networks and, tasks using machine-learning (ML)-adjusted *t* tests (Nadeau & Bengio, 1999). These modified *t* tests are evaluated and adjusted for comparing ML algorithms (Bouckaert & Frank, 2004) to account for violating the independence assumption in a paired Student's *t* test. This is done by correcting the variance estimate through considering the training and sample size. In our case, due to the leave-30%-families-out CV scheme, the number of data points changed in each fold. Therefore, we used the mean training sample size across the 100 folds for the adjustment.

Within each phenotypic domain, we first tested effects of state and network by averaging prediction performance of the respective other factors (i.e., averaging across networks and task when testing for state effects, and across state and task when testing for network effects). As state–target similarity is an extension of state specificity, we here only averaged across networks for same and similar tasks, respectively. Significant effects (Bonferroni corrected for multiple comparisons) were then further assessed by comparing the respective individual prediction scores between each other.

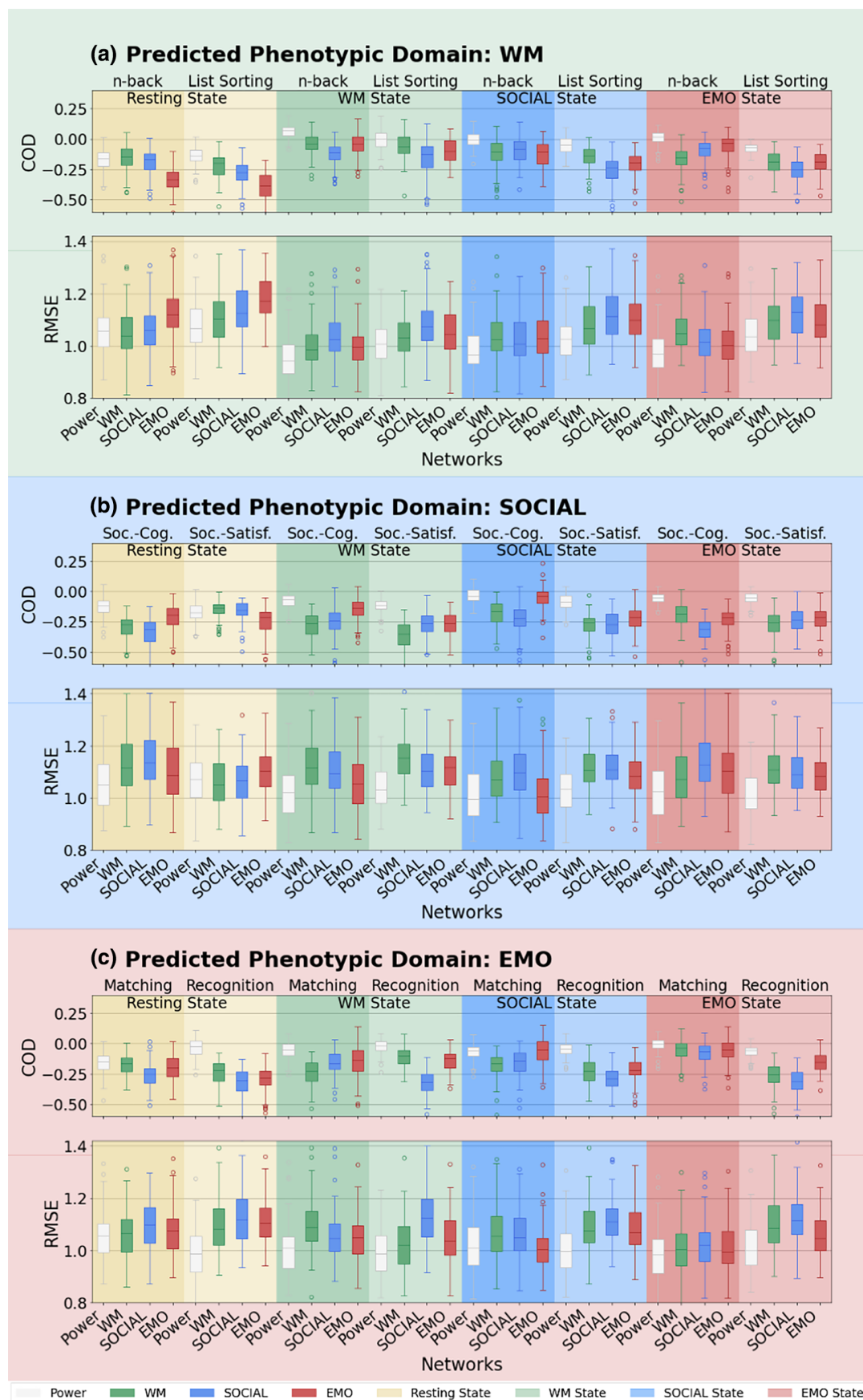


FIGURE 3 Legend on next page.

3 | RESULTS

In a first step (upper/yellow part of Figure 2), we compared the resulting task-networks from both sample 1 and the meta-analyses to ensure they covered and showed overlap with frequently observed regions in previously reported large-scale analyses (WM: Daamen et al., 2015; Fuentes-Claramonte et al., 2019; Kennedy et al., 2017; Rottschy et al., 2012; Snoek et al., 2021; <https://identifiers.org/neurovault.collection:7103>; SOCIAL: T. Chen et al., 2023; Hennion et al., 2016; Mossad et al., 2022; Patil et al., 2017; EMO: Chaudhary et al., 2023; Herrmann et al., 2020; Nord et al., 2017; Snoek et al., 2021; <https://identifiers.org/neurovault.collection:7103>). The resulting WM-NW had 49 nodes, the SOCIAL-NW had 66 nodes, and EMO-NW had 84 nodes. For an overview of the networks, please see Figure 2. Please refer to the supplemental material for a comprehensive depiction of the original activation maps resulting from the group-level (Figures S1–S3) and the activation likelihood maps of the meta-analyses (Figures S4–S6), extracted networks (Figures S7–S13) and further details on the exact coordinates, including the anatomical labels (supplemental Tables S2–S7).

Following this, we examined the FC of sample 2 within the networks derived from sample 1 for each individual state. Averaged across all subjects, the pattern of FC of all networks looks similar between different states. However, we see a tighter network coupling within the WM-NW in the WM state, compared to the other states (i.e., SOCIAL, EMO, and resting state; see Supplemental Figures S14 and S15). A tendency of higher FC within the congruent network was also visible in the EMO domain (Supplemental Figures S18 and S19). However, the picture was more fuzzy within the SOCIAL domain (Supplemental Figures S16 and S17), where no apparent pattern was discernable. This further motivated the next step—the application of ML and its assessment—to investigate whether the algorithms would be able to pick up complex and subtle signal that we were unable to observe within the averaged FC matrices.

The main focus of this study, therefore, lies on the prediction analyses (bottom/violet part of Figure 2). We mainly report on the outcomes of the prediction analyses using PLS. Predictions using different algorithms and approaches showed highly similar patterns and their details can be found in the supplementary material. Furthermore, regarding outcome measures, we here focus on the RMSE from the CV as well as the COD as a measure of goodness of fit. In the supplementary material, additional results of Pearson's r of the predicted and observed score from the CV can be found.

Averaged across all CV-folds per prediction, the COD and RMSE (Figure 3) revealed that the models show a poor fit and prediction accuracies are rather low. Note, that COD values below zero indicate that

prediction of individual scores were worse than predicting the mean of the target. The mean COD showed a positive mean value only for 2 out of 96 predictions, while all others showed a mean COD of zero or a negative value. Other models (e.g., kernel ridge regression or SVR) yielded some more COD values above zero, but no model achieved a mean COD higher than 0.07. Similarly, the RMSE was quite high for all predictions. Because these scores indicated a generally poor fit to the data, we refrained from applying the best model to the holdout sample. The correlations (for details, see the supplementary material) between predicted and observed values ranged from -0.11 to 0.32 with a mean prediction accuracy of 0.08 (SD: 0.09) and only one mean correlation from the 96 predictions reaching a medium effect size.

3.1 | State specificity in network-based prediction

To answer the question if the correspondence between state and target (e.g., WM score predicted using FC during WM state) improves predictions, and whether there even is state specificity (e.g., WM scores predicted significantly better from FC during WM state than from FC during other states: REST, SOCIAL, or EMO), we examined the differences in prediction accuracy between states.

No significant differences were found for the SOCIAL and EMO domain (Figure 4b,c). For the WM domain, the ML-adjusted t test (Bonferroni corrected for multiple comparisons) showed a significant benefit for all task states compared to the resting state (see Figure 4a; see Table 1 for mean RMSE and significant t test statistics). However, non-WM domain states (i.e., SOCIAL and EMO) only showed a significant difference to the WM state at an uncorrected threshold (not shown in Table 1). This difference was also significant when using other algorithms and feature selection approaches (PLS with CBPM, random forest, and SVR with the RBF kernel). At an uncorrected threshold, the difference was also significant for all other models (kernel ridge regression, SVR with linear kernel, and CBPM with ridge regression), as well as when using only the trimmed time-series.

To assess which effect was driving the significant differences, we compared prediction performance between states using *post hoc* ML-adjusted t tests, while keeping network and task constant. That is, we only compared predictions between same-domain networks and tasks (e.g., comparing the prediction performance of “same” WM task score based on FC between Power nodes in *resting state* to the prediction performance of “same” WM task score based on FC between Power nodes in *WM state*). Comparing prediction performance between rest and different states for each network and WM score revealed that the difference between REST and WM state was driven by the difference in prediction performance of the n-back task using the Power

FIGURE 3 Prediction performance: Boxplots of the distribution of COD and RMSE from the $100\times$ CV for each phenotypic domain (a—WM, b—Social, c—EMO domain), state and network. Boxes represent the model fit/COD and RMSE of prediction of a specific score (WM, SOCIAL, EMO; performed in [darker background] or outside [lighter background] the scanner) based on functional connectivity within a given network (POWER, WM, SOCIAL, EMO) in a given state (REST, WM, SOCIAL, EMO). Green: WM, blue: SOCIAL, red: EMOTION, yellow: resting state, white: Power nodes. Darker background: target is the task performed in the scanner; lighter background: target is the task performed outside the scanner. Soc. Cog., Social cognition task (in-scanner task); Soc. Satisf., Social Satisfaction Questionnaire (out-of-scanner score); Matching: Emotional Face-Matching task (in-scanner task); Recognition: Emotional Face Recognition task (out-of-scanner score).

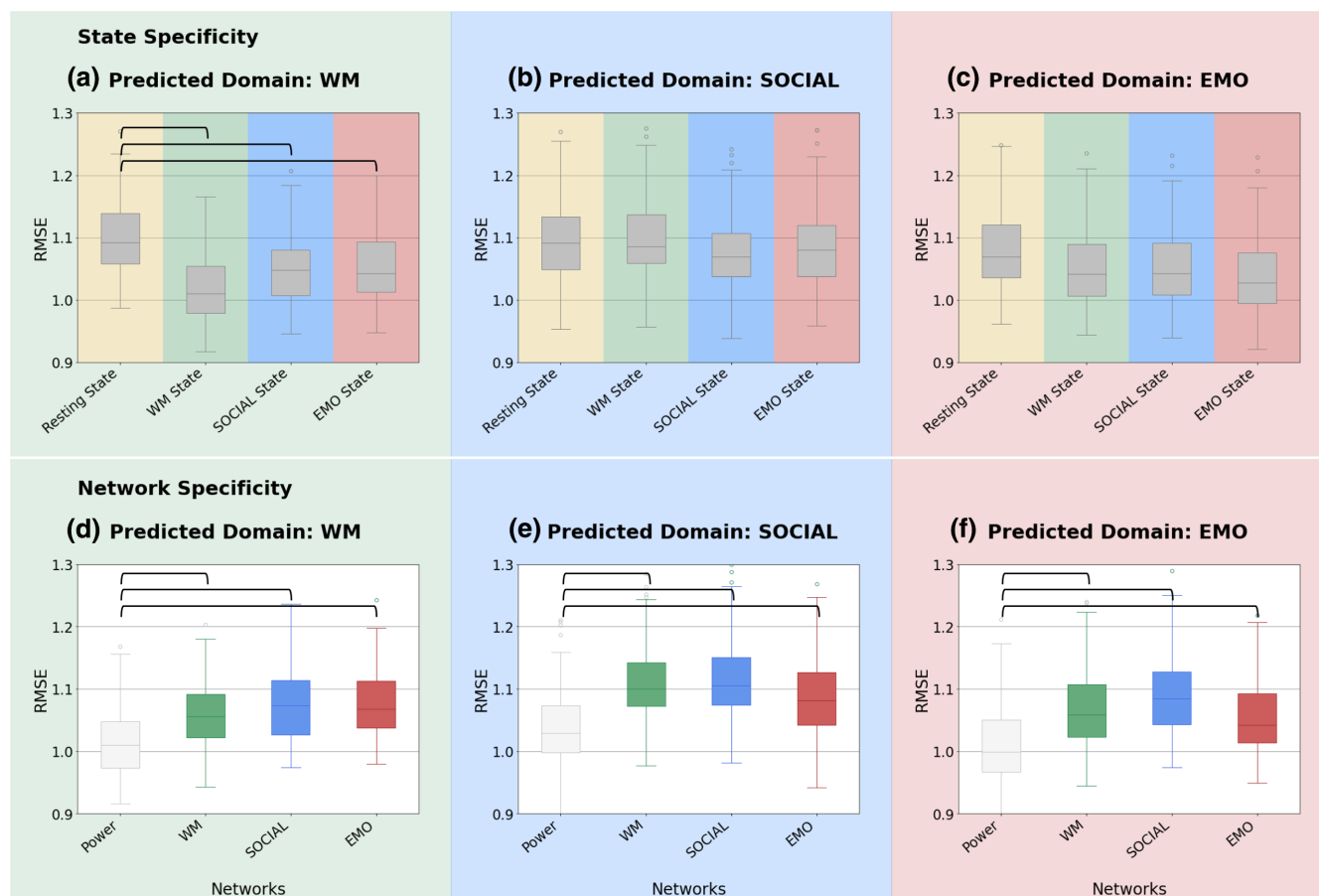


FIGURE 4 State and network specificity: State (a–c) and network (d–f) specificity for each phenotypic domain (a and d—prediction of WM scores, b and e—prediction of SOCIAL scores, c and f—prediction of EMO scores). For state specificity, (a–c) RMSE of all networks (POWER, WM, SOCIAL, EMO) and the two tasks of the respective domain in a given state (REST, WM, SOCIAL, EMO) are averaged. For network specificity, all states (REST, WM, SOCIAL, EMO) and the two tasks of the respective domain are averaged in a given network (Power nodes, WM, SOCIAL, EMO). Green: WM, blue: SOCIAL, red: EMOTION, yellow: resting state. Horizontal bars indicate significance $p_{\text{corr}} < .05$.

nodes and the EMO network (for mean RMSE and significant t test statistics, see Table 1). The EMO network additionally showed a difference between WM and REST state when predicting list sorting. The difference between REST and EMO state was also driven by Power nodes and the EMO network when predicting n-back performance, and by the EMO network when predicting list sorting. The significant difference in REST versus SOCIAL was driven by the Power nodes in predicting n-back performance.

Overall, state had a significant influence on predicting WM scores, with better predictions when using task compared to resting state. This state-specific improvement was, however, not uniformly observed and mainly driven by predictions based on FC within the Power nodes and the EMO network.

3.2 | State–target similarity in network-based prediction

In a next step, we examined differences in predictability between the “same” and “similar” tasks in a given domain. We were interested in whether the behavior would be predicted better in the state where

the predicted behavior was measured (“same task”), and whether the FC-based predictivity could translate to a related task (“similar task”). An example would be the comparison between the predictability of n-back task performance (WM task performed during scanning) and the list sorting task (WM task performed outside of the scanner) based on FC patterns observed during the WM n-back task.

For this comparison of “same task” with “similar task,” we found a slight (numerical) benefit in the performance of the “same task.” However, a direct statistical comparison of RMSE values using ML-adjusted t tests did not show any significant effects after Bonferroni correction for multiple comparisons (see Figure 5).

3.3 | Network specificity in network-based prediction

Finally, we set out to answer the question if predicting task performance does benefit from being based on FC within a network known to be engaged in performing that same task (e.g., n-back task performance predicted from FC within the WM-network), as compared to other task-related networks (e.g., n-back task performance predicted

TABLE 1 Comparison of prediction accuracies between states.

Domain of predicted performance	State A	RMSE mean (SD)	State B	RMSE mean (SD)	t	p-Value
Significant main effect of state						
WM	REST	1.10 (0.06)	WM	1.02 (0.05)	5.10	<.001
	REST	1.10 (0.06)	SOCIAL	1.05 (0.06)	3.49	.013
	REST	1.10 (0.06)	EMOTION	1.05 (0.06)	3.68	.007
Significant <i>post hoc</i> tests for specific state–task–network combinations						
WM	REST (n-back, Power nodes)	1.06 (0.09)	WM (n-back, Power nodes)	0.95 (0.09)	4.83	<.001
	REST (n-back, EMO network)	1.13 (0.10)	WM (n-back, EMO network)	1.00 (0.10)	3.51	.016
	REST (List Sorting, EMO network)	1.19 (0.09)	WM (List Sorting, EMO network)	1.05 (0.09)	4.10	.002
	REST (n-back task, Power nodes)	1.06 (0.09)	SOCIAL (n-back task, Power nodes)	0.98 (0.09)	3.21	.043
	REST (n-back task, Power nodes)	1.06 (0.09)	EMOTION (n-back task, Power nodes)	0.97 (0.09)	3.72	.008
	REST (n-back, EMO network)	1.13 (0.10)	EMOTION (n-back, EMO network)	1.00 (0.10)	3.40	.023
	REST (List Sorting, EMO network)	1.19 (0.09)	EMOTION (List Sorting, EMO network)	1.10 (0.09)	3.32	.030

Note: Machine-learning-adjusted *t* test to assess state specificity using the averaged 100 RMSE values obtained from 100-fold cross-validation within the state listed in column “State A” versus the state listed in column “State B.” *p*-Values are Bonferroni corrected for multiple comparisons. *Post hoc t* tests between individual predictions of the task in the network (both noted in brackets) and the state listed in column “State A” versus the state listed in column “State B.”

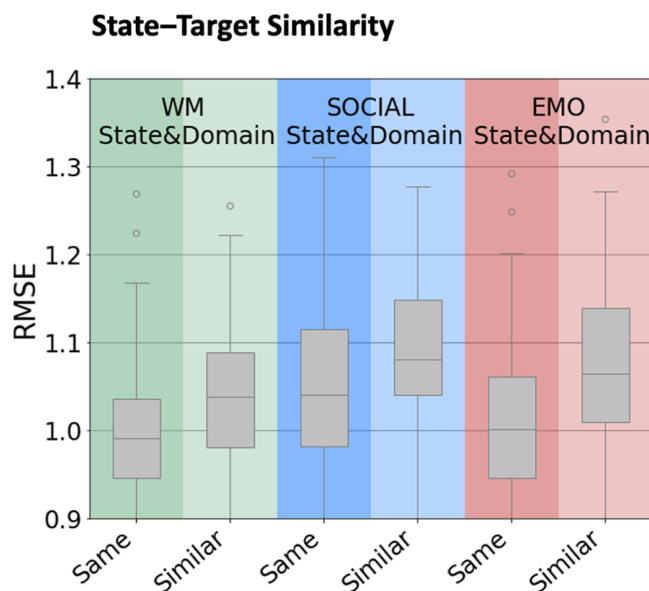


FIGURE 5 State–target similarity: Boxplots of the distribution of RMSE from the 100× CV averaged across all networks (Power nodes, WM, SOCIAL, EMO network) within a given state (WM, SOCIAL, and EMO) and task (SAME or SIMILAR). Green: WM, blue: SOCIAL, red: EMOTION, gray: averaged across networks. Darker background: target is the “same” task performed in the scanner, lighter background: target is the “similar” task performed outside the scanner.

from FC within a SOCIAL task-based network or the whole-brain connectome). ML-adjusted *t* tests showed, for all three domains, a benefit for the whole-brain Power nodes over the task-specific networks (see Figure 4d–f). In the WM domain, FC between the Power nodes predicted WM-related targets better than did FC within the n-back WM-specific network, the SOCIAL-specific network, or the EMO-specific network (see Figure 4d and Table 2 for mean RMSE and *t* test statistics). In the SOCIAL domain, the Power nodes predicted SOCIAL-related targets better than did the WM-specific, SOCIAL-specific, or EMO-specific networks (see Figure 4e). Finally, in the EMO domain, the Power nodes again predicted EMO-related targets better than all three domain-specific networks including the EMO-specific network derived from an emotional face-matching task (see Figure 4f).

Post hoc tests (see Table 2) indicated that the difference between the Power and WM networks was driven by the EMO and WM state when predicting social satisfaction, while there was no specific state or task driving the effects for the EMO and WM domain. The priority of Power over the SOCIAL-specific network was in particular evident when predicting social cognition and social satisfaction in the EMO state and when predicting emotion recognition in the REST, SOCIAL, and EMO states. No specific state or task was driving the effect for the WM domain. The effect of Power versus the EMO-specific network was driven by the predictions of list sorting and emotion recognition in the REST state. No specific state or task was driving the effect in the SOCIAL domain.

TABLE 2 Comparison of prediction accuracies between networks.

Domain of predicted performance	Network A	RMSE mean (SD)	Network B	RMSE mean (SD)	t	p-Value
Significant main effect of network						
WM	Power nodes	1.02 (0.06)	WM	1.06 (0.06)	−3.30	.025
			SOCIAL	1.08 (0.06)	−4.65	<.001
			EMO	1.08 (0.06)	−6.03	<.001
SOCIAL	Power nodes	1.04 (0.07)	WM	1.11 (0.06)	−6.37	<.001
			SOCIAL	1.11 (0.06)	−5.72	<.001
			EMO	1.08 (0.07)	−4.43	<.001
EMO	Power nodes	1.01 (0.06)	WM	1.07 (0.07)	−5.05	<.001
			SOCIAL	1.09 (0.06)	−7.10	<.001
			EMO	1.05 (0.06)	−3.93	.003
Significant <i>post hoc</i> tests for specific network—task—state combinations						
WM	Power (List Sorting, REST state)	1.07 (0.09)	EMO (List Sorting, REST state)	1.19 (0.09)	−4.41	.002
SOCIAL	Power (Soc. Cog., EMO state)	1.02 (0.11)	SOCIAL (Soc. Cog., EMO state)	1.14 (0.11)	−4.80	<.001
	Power (Soc. Satisf., EMO state)	1.01 (0.08)	WM (Soc. Satisf., EMO state)	1.11 (0.08)	−3.82	.017
	Power (Soc. Satisf., WM state)	1.04 (0.08)	WM (Soc. Satisf., WM state)	1.15 (0.08)	−3.68	.027
EMO	Power (Emo. Recog., REST state)	1.00 (0.09)	SOCIAL (Emo. Recog., REST state)	1.13 (0.10)	−4.05	.008
			EMO (Emo. Recog., REST state)	1.11 (0.09)	−4.39	.002
	Power (Emo. Recog., WM state)	0.99 (0.09)	SOCIAL (Emo. Recog., WM state)	1.13 (0.10)	−5.18	<.001
	Power (Emo. Recog., SOCIAL state)	1.00 (0.09)	SOCIAL (Emo. Recog., SOCIAL state)	1.11 (0.09)	−4.47	.002
	Power (Emo. Recog., EMO state)	1.01 (0.09)	SOCIAL (Emo. Recog., EMO state)	1.12 (0.09)	−4.47	.002

Note: Machine-learning-adjusted *t* tests to assess network specificity using the averaged 100 RMSE values obtained from 100-fold cross-validation within the network listed in column “Network A” versus the network listed in column “Network B.” *p*-Values are Bonferroni corrected for multiple comparisons. *Post hoc t* tests between individual predictions of the task in the state (both noted in brackets) and the network listed in column “Network A” versus the state listed in column “Network B.”

Abbreviations: Emo. Recog., Emotional Face Recognition task (out-of-scanner score); Soc. Cog., Social cognition task (in-scanner task); Soc. Satisf., Social Satisfaction Questionnaire (out-of-scanner score).

This superiority of the Power nodes over functional network definitions in all three domains was not present in other ML algorithms. However, we still saw a similar trend when trimming the time series to the shortest task (EMO: 2:16 min), as the Power nodes performed better in all domains and networks, except for the EMO domain where the EMO network did not perform significantly worse.

4 | DISCUSSION

Using state-of-the-art fMRI preprocessing and ML approaches, this study investigated brain–behavior relationships. Specifically, how brain features from specific states and networks, or the task similarity within the behavioral domain, affects this relationship. Based on

previous studies, we hypothesized that brain features obtained from networks and/or states that are corresponding to the target task are more informative about individual behavior than those obtained from other (non-corresponding) states or networks. Additionally, we expected that behavior in the task performed during fMRI data acquisition will be predicted better than similar tasks of the same phenotypic domain. Contrary to expectations, we found no significant differences in predictability (when correcting for multiple comparisons) that would indicate specific benefits of state, task, or network correspondence. Rather, our results show a general benefit of predicting WM scores using (any) task state, relative to rest, and for predicting performance in any domain from whole-brain FC (Power nodes), relative to predefined functional networks. Importantly, however, prediction accuracies were overall quite low, raising the question to what extent the observed differences (or their absence) in prediction

performance between state and network conditions, and task similarity can be meaningfully interpreted.

4.1 | Is there state specificity for brain–behavior prediction?

We expected not only an improvement in predicting behavior based on FC during task states compared to FC at rest as demonstrated by (Greene et al., 2018), but especially in predicting behavior based on FC within corresponding states. However, apart from the overall low prediction performance, our results, show only weak evidence for the former, that is, an advantage of task states compared to resting state, but only for predicting WM scores (see Figure 4a). Previous studies have already reported that predictions of cognitive scores such as intelligence or attention improve when using task fMRI data (vs. resting state) to derive FC patterns (Avery et al., 2020; Greene et al., 2018; Jiang et al., 2020). But also when combining task and rest (Jiang et al., 2020), and specifically when using FC within a WM task state (Avery et al., 2020; Jiang et al., 2020; Sripada et al., 2020; Stark et al., 2021). Our results extend this work to different states and additionally show that this benefit of using task-fMRI data cannot be assumed for behaviors other than WM. Task-fMRI data may lead to better predictions compared to resting-state data, particularly for WM performance, possibly because task-based fMRI has a more constrained setup which potentially enhances reliability. Resting-state fMRI has been shown to be less reliable than stimulating (such as movies or task) fMRI (Noble et al., 2019). Task-based modulations of brain states may therefore contain more information about individual differences in brain functioning and behavior (Greene et al., 2018). This is consistent with a recent emphasis on shifting from solely focusing on resting-state FC (Finn, 2021; Greene et al., 2018) to accelerate progress in human neuroscience. Possibly, predictive performance can be improved by using naturalistic stimuli (Finn & Bandettini, 2021), as such settings are still more constrained than rest but less constrained than certain laboratory tasks. However, using movie data would not readily allow testing of state specificity effects, and therefore would not be a ready-made solution to the current research question.

4.2 | Is there state–target similarity?

Surprisingly, prediction accuracy was low even when FC was derived from the exact same task state in which the behavioral data were collected, with no improvement when predicting the same score, as compared to a similar score (see Figure 5). Also using the HCP dataset and WM task data, Stark et al. (2021) reported similar though slightly higher accuracies than ours for n-back (“same task”) and list-sorting (“similar task”) scores, with higher correlations for the former (but not tested for significance). We here extend these insights by also testing effects of task similarity for SOCIAL and EMO scores and by showing that although n-back WM performance seems to be predicted better,

the difference between “same task” and “similar task” score prediction is not significant.

A possible reason for the lack of support of our hypothesis might lie in the nature of the task used in the scanner. That is, a lot of paradigms were developed in experimental contexts (Hedge et al., 2018) and therefore optimized for inducing a robust effect across participants instead of assessing interindividual differences. This might especially be the case for experimental tasks used in the scanner. For example, the emotional face-matching task (Hariri et al., 2002) used for EMO assessment was developed to induce robust amygdala activation, rather than capturing individual emotion processing abilities. Additionally, the tasks used here were rather short and may have lacked enough difficult items for a clearer differentiation between participants. The n-back task, for example, most strongly differentiates between individuals when using high-load conditions (>3-back), both in terms of behavior and brain activity (Lamichhane et al., 2020). Therefore, using behavioral measurements from tasks optimized for obtaining stable group-average effects might have counteracted the successful prediction of interindividual differences.

4.3 | Is there network specificity for brain–behavior predictions?

We based the network specificity hypothesis on the assumption that if networks are reliably engaged during a task, then these networks should play an important role in the task outcome (i.e., specific performance). Importantly, our aim to demonstrate network specificity was based on the idea that a priori task-defined networks improve interpretability (Bzdok et al., 2012; Langner et al., 2018; Müller et al., 2018; Rottschy et al., 2012) as they reflect interactions between regions that are jointly engaged during a specific task and should therefore be biologically meaningful (J. Chen et al., 2021; Nostro et al., 2018; Pläschke et al., 2017). Further, visual inspection of FC within task networks and states averaged across subjects (see supplemental Figures S14–S20) revealed the expected stronger FC within the congruent networks and states, respectively. This was most strongly expressed in WM and EMO, whereas in SOCIAL no clear pattern was visible. Yet, this apparently tighter coupling of congruent networks did not provide enough information for the prediction of individual behavior to translate into a significant improvement.

Nonetheless, our results showed that prediction performance was weak regardless of the networks used (see Figure 3—COD). Comparison of the differences between networks showed that prediction from the whole-brain representation (Power et al., 2011) significantly improved prediction compared to the task specific networks (see Figure 4d–f). The reason for an advantage of the whole-brain connectome remains to be revealed. We assume that some subtle pieces of information in the whole-brain connectome, which are not captured by the task networks, reflect individual processing differences in some parts of the tasks at hand and thus contribute to some extent to behavior prediction. Additionally, the whole-brain connectome has considerably more nodes than the task networks studied here, giving

the model much more features to learn from. These nodes can potentially capture interactions and integration of multiple brain regions, including regions, that are not consistently involved in the investigated mental processes and do not translate into group-level average task networks. However, despite the whole-brain Power nodes performing, on average, significantly better than the task networks, they were not consistently superior (see Supplemental Figure S29 showing the prediction performance sorted by network size).

Nevertheless, our results suggest that there is no network specificity, which is in line with the findings of Heckner et al., 2023 and Pläschke et al., 2017, 2020. Using networks based on group analyses may therefore not be a suitable avenue for assessing individual differences (Finn et al., 2017; O'Connor et al., 2017; Shah et al., 2016). Brain mapping results from group analyses typically reveal regions with low inter-subject variability (Hedge et al., 2018), and group-averaged patterns of brain activation often look quite different from patterns observed on the individual subject level (Miller et al., 2002). In addition, it has been shown that brain regions, for which activation has been found to be associated with behavioral outcomes, are not necessarily those that show up in standard group-average analyses (Ganis et al., 2005). Our results now indicate that this might also apply to networks derived from large samples that also reflect small average effects (HCP-derived networks) or from large-scale meta-analyses. However, improvement may be gained through an individualization of the networks prior to prediction. For example, Kong and colleagues employed a multi-session hierarchical Bayesian model to estimate individual-specific cortical network parcellations, significantly improving prediction performance relative to other parcellations (Kong et al., 2019). Similarly, using a different approach (Li et al., 2019) demonstrated an improvement in prediction accuracy using an iterative search based on a population-based functional atlas in combination with a map of inter-individual variability (D. Wang et al., 2015).

4.4 | Methodological considerations

In this study, we aimed to predict complex behavior based on FC. Generally, our prediction accuracies were rather low. Nevertheless, they are comparable to the accuracies (correlation between predicted and observed score) reported in the literature (Dubois, Galdi, Han, et al., 2018; Greene et al., 2018; Heckner et al., 2023; Kandaleft et al., 2022; Ooi et al., 2022; Tomasi & Volkow, 2020). However, using correlation alone as a measure for prediction accuracy can skew the picture. All measures individually (correlation, COD, RMSE) have been shown to have their drawbacks and therefore it has been suggested that they should be considered together as a whole (Poldrack et al., 2020). Our results emphasize the importance of using more than one measure and especially using more than Pearson's r as a measure for prediction performance, as this metric, when used in isolation, may draw an overly optimistic picture. As illustrated in our plots (see supplementary material), Pearson's r invites the observer to interpret some apparent patterns. Yet, when looking at the model fit given by COD values, it can be easily seen that most models barely fit the data

(see e.g. Figure 3). Surprisingly, only few prediction studies in the neuroimaging literature have reported metrics other than r . However, if they did, results were rather similar to ours, with finding only small amounts of variance explained (COD) and reporting high prediction errors on average (Dubois, Galdi, Han, et al., 2018; Kandaleft et al., 2022; Ooi et al., 2022).

There may be several reasons why we did not successfully predict behavioral performance. One is predicting behavioral scores of single tasks or questionnaires, like WM, as opposed to compound scores across many tests, like overall cognition (Akshoomoff et al., 2013; Dubois, Galdi, Paul, & Adolphs, 2018). Studies using compound scores generally report better accuracies (McCormick et al., 2022; Ooi et al., 2022), as they may capture individual abilities better and show higher reliabilities compared to individual test scores (Hedge et al., 2018). However, the interpretation and biological foundation of compound scores is debatable (Dubois, Galdi, Paul, & Adolphs, 2018; McFarland, 2012; Van Der Maas et al., 2006). In this study, we aimed to investigate specificity, and hence we focused on individual tasks or questionnaires at the cost of a potential decrease in prediction performance.

Another reason for the low prediction performance, related to the first explanation, might be the reliability of the predicted measures but also the features, setting an upper bound for detecting relationships (Cohen et al., 2013; Vul et al., 2009; Yarkoni & Braver, 2010). Using the HCP test-retest sample calculation of the correlations between measurement time points 1 and 2 (test-retest reliability) of the scores we used revealed reliabilities between 0.5 and 0.8, with highest reliabilities for the WM domain. In our and other studies, WM or intelligence scores were generally predicted better than other cognitive measures (Avery et al., 2020; Kandaleft et al., 2022; Ooi et al., 2022; Sripada et al., 2020; Takeuchi et al., 2021), which could be because these constructs are measured more accurately than others.

Finally, for ML applications in CV schemes, sample size is an important factor for achieving good prediction performance. The more data is available, the better a model can learn. In our case, our sample size decreased due to our carving out a subsample for a priori network delineation, leaving us with 420 subjects in the training set. This step was essential to assess network specificity using networks as close as possible to the investigated tasks. Other studies using the HCP dataset and similar algorithms have in part achieved slightly better predictions, possibly through larger training sets (Jiang et al., 2020; Ooi et al., 2022). The effects we sought to detect are presumably very small; hence, a substantially larger dataset and/or more reliable behavioral assessments could be required to detect them (Marek et al., 2022).

4.5 | Limitations and outlook

We are aware that there is a plethora of preprocessing pipelines and feature selection models that may improve prediction. We used a well-established preprocessing pipeline (Glasser et al., 2016) and

widely used ML models that previously yielded the highest predictions (Greene et al., 2018; Jiang et al., 2020; Yeung et al., 2022). Given that we saw a similar pattern of prediction accuracies irrespective of the model used, we would not expect a substantial change of the result pattern if other models were used.

Further, we aimed to cover a broad range of task network representations by (i) extracting networks from a high-powered single study using task fMRI data, (ii) using ALE meta-analyses based on previously published neuroimaging results (see supplemental material), as well as (iii) including a whole-brain representation (Power et al., 2011) for comparison with the task networks. We acknowledge that different whole-brain representations, such as the parcellation by Schaefer et al. (2018) could yield different and possibly even better prediction accuracies. Also, the inclusion of data-driven approaches to network definition, like principle component analysis or group independent component analysis, could lead to different results. Testing the influence of such methodological choices is an important research topic and should be addressed more systematically in future studies. Until then, our results should only be generalized to settings that employ the same or similar methods as were used here. We here, hence, limited our analyses to one whole-brain representation, as our focus was on task networks and their interpretability. This also entails using the task-specific networks in their most accurate representation, encompassing their unique spatial distributions as well as their different sizes. We believe that both aspects constitute fundamental inherent characteristics of networks.

Finally, the HCP dataset comprises young and healthy adults, with an above-population average intelligence. As the majority of subjects in the HCP dataset were highly educated, performed generally well on the tests, and as tests are optimized for group effects, the between-subject variability in this dataset is relatively limited, that is, suboptimal for approaches relying on individual differences. Nevertheless, the HCP currently offers the only dataset that allows for the investigation of such complex research questions as state specificity, state-target similarity, and network specificity in brain-based prediction settings, because it covers a vast array of phenotypic domains, both in and outside the scanner, while providing high-quality fMRI data in task and resting states in a large number of participants.

4.6 | Conclusions

Here, using state-of-the-art ML algorithms for out-of-sample prediction analyses, we aimed to investigate the specific influence of the factors state, task, and network on behavior prediction from FC patterns. Based on previous research on brain-behavior relationships, we hypothesized that FC features from corresponding state, tasks, and networks would be more informative than non-corresponding features and hence improve prediction. We only found improvement for using task over resting state fMRI, as well as better predictions for whole brain compared to task specific networks. However, across three behavioral domains, predictive performance was generally poor,

and there were no significant patterns indicating specificity of state, networks, or task similarity, when looking at RMSE and COD. A significant improvement of prediction performance based on task-fMRI (vs. resting-state fMRI) was only observed for the WM domain. Of note, an isolated consideration of Pearson's correlation coefficient as the sole index of model fit would have led us to different and apparently overly optimistic conclusions. Hence, even with maximum state-network-behavior compatibility, the relationship between FC and behavior remains low. This study therefore emphasizes the need for a critical assessment of prediction accuracies and suggests that individual behavior cannot be successfully predicted based solely on FC in task-specific networks.

AUTHOR CONTRIBUTIONS

Nevena Kraljević: Conceptualization, methodology, software, formal analysis, writing—original draft, writing—review and editing, visualization. Robert Langner: Conceptualization, methodology, supervision, writing—review and editing. Vincent Küppers: Writing—review and editing: methodology. Federico Raimondo: Methodology, software, writing—review and editing. Kaustubh R. Patil: Methodology, software, writing—review and editing. Simon B. Eickhoff: Funding acquisition, resources, supervision, writing—review and editing. Veronika I. Müller: Conceptualization, methodology, supervision, writing—review and editing.

ACKNOWLEDGMENTS

The authors would like to sincerely thank all participants and contributors to the data provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The work was supported by: Helmholtz Portfolio Theme “Supercomputing and Modeling for the Human Brain”; European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 945539 (HBP SGA3). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interest.

DATA AVAILABILITY STATEMENT

The data used in this study were obtained from the Human Connectome Project (HCP) database, which is publicly accessible at <https://www.humanconnectome.org/>. However, some parameters needed application for restricted access, which all authors handling the data had granted by the HCP. The three meta-analytically defined networks will be openly available via the ANIMA-database (Reid et al., 2016; <https://anima.fz-juelich.de/>).

ORCID

Nevena Kraljević  <https://orcid.org/0000-0003-0869-648X>

REFERENCES

- Akshoomoff, N., Beaumont, J. L., Bauer, P. J., Dikmen, S. S., Gershon, R. C., Mungas, D., Slotkin, J., Tulskey, D., Weintraub, S., Zelazo, P. D., & Heaton, R. K. (2013). NIH toolbox cognition function battery (CFB): Composite scores of crystallized, fluid, and overall cognition. *Mono-graphs of the Society for Research in Child Development*, 78(4), 119–132. <https://doi.org/10.1111/mono.12038>
- Avery, E. W., Yoo, K., Rosenberg, M. D., Greene, A. S., Gao, S., Na, D. L., Scheinost, D., Constable, R. T., & Chun, M. M. (2020). Distributed patterns of functional connectivity predict working memory performance in novel healthy and memory-impaired individuals. *Journal of Cognitive Neuroscience*, 32(2), 241–255. https://doi.org/10.1162/jocn_a_01487
- Babakhanyan, I., McKenna, B. S., Casaletto, K. B., Nowinski, C. J., & Heaton, R. K. (2018). National Institutes of Health toolbox emotion battery for English- and Spanish-speaking adults: Normative data and factor-based summary scores. *Patient Related Outcome Measures*, 9, 115–127. <https://doi.org/10.2147/PROM.S151658>
- Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., ... WU-Minn HCP Consortium. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, 80, 169–189. <https://doi.org/10.1016/j.neuroimage.2013.05.033>
- Bouckaert, R. R., & Frank, E. (2004). Evaluating the replicability of significance tests for comparing learning algorithms. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Advances in knowledge discovery and data mining* (Vol. 3056, pp. 3–12). Springer. https://doi.org/10.1007/978-3-540-24775-3_3
- Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function*, 217(4), 783–796. <https://doi.org/10.1007/s00429-012-0380-y>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, 12(3), 314–325. <https://doi.org/10.1006/nimg.2000.0612>
- Chaudhary, S., Zhang, S., Zhorinitsky, S., Chen, Y., Chao, H. H., & Li, C.-S. R. (2023). Age-related reduction in trait anxiety: Behavioral and neural evidence of automaticity in negative facial emotion processing. *NeuroImage*, 276, 120207. <https://doi.org/10.1016/j.neuroimage.2023.120207>
- Chen, J., Müller, V. I., Dukart, J., Hoffstaedter, F., Baker, J. T., Holmes, A. J., Vatansever, D., Nickl-Jockschat, T., Liu, X., Derntl, B., Kogler, L., Jardri, R., Gruber, O., Aleman, A., Sommer, I. E., Eickhoff, S. B., & Patil, K. R. (2021). Intrinsic connectivity patterns of task-defined brain networks allow individual prediction of cognitive symptom dimension of schizophrenia and are linked to molecular architecture. *Biological Psychiatry*, 89(3), 308–319. <https://doi.org/10.1016/j.biopsych.2020.09.024>
- Chén, O. Y., Cao, H., Reinen, J. M., Qian, T., Gou, J., Phan, H., De Vos, M., & Cannon, T. D. (2019). Resting-state brain information flow predicts cognitive flexibility in humans. *Scientific Reports*, 9(1), 3879. <https://doi.org/10.1038/s41598-019-40345-8>
- Chen, T., Gau, S. S.-F., Wu, Y.-Y., & Chou, T.-L. (2023). Neural substrates of theory of mind in adults with autism spectrum disorder: An fMRI study of the social animation task. *Journal of the Formosan Medical Association*, 122(7), 621–628. <https://doi.org/10.1016/j.jfma.2022.10.009>
- Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., Davatzikos, C., Gur, R. C., Gur, R. E., Bassett, D. S., & Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, 154, 174–187. <https://doi.org/10.1016/j.neuroimage.2017.03.020>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Taylor & Francis. Retrieved from <https://books.google.de/books?id=fAnSOgbdFXIC>
- Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, 163, 115–124. <https://doi.org/10.1016/j.neuroimage.2017.07.059>
- Cui, Z., Su, M., Li, L., Shu, H., & Gong, G. (2018). Individualized prediction of reading comprehension ability using gray matter volume. *Cerebral Cortex*, 28(5), 1656–1672. <https://doi.org/10.1093/cercor/bhx061>
- Daamen, M., Bäuml, J. G., Scheef, L., Sorg, C., Busch, B., Baumann, N., Bartmann, P., Wolke, D., Wohlschläger, A., & Boecker, H. (2015). Working memory in preterm-born adults: Load-dependent compensatory activity of the posterior default mode network. *Human Brain Mapping*, 36(3), 1121–1137. <https://doi.org/10.1002/hbm.22691>
- Dubois, J., Galdi, P., Han, Y., Paul, L. K., & Adolphs, R. (2018). Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. *Personality Neuroscience*, 1, e6. <https://doi.org/10.1017/pen.2018.8>
- Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756), 20170284. <https://doi.org/10.1098/rstb.2017.0284>
- Esteban, O., Ciric, R., Finc, K., Durnez, J., Ghosh, S., & Poldrack, R. A. (2019). *Task fMRI analysis using FSL and data preprocessed with fMRI-Prep [computer software]*. Zenodo. <https://doi.org/10.5281/zenodo.2635848>
- Finn, E. S. (2021). Is it time to put rest to rest? *Trends in Cognitive Sciences*, 25(12), 1021–1032. <https://doi.org/10.1016/j.tics.2021.09.005>
- Finn, E. S., & Bandettini, P. A. (2021). Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage*, 235, 117963. <https://doi.org/10.1016/j.neuroimage.2021.117963>
- Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*, 160, 140–151. <https://doi.org/10.1016/j.neuroimage.2017.03.064>
- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11), 1664–1671. <https://doi.org/10.1038/nn.4135>
- Franke, K., Luders, E., May, A., Wilke, M., & Gaser, C. (2012). Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI. *NeuroImage*, 63(3), 1305–1312. <https://doi.org/10.1016/j.neuroimage.2012.08.001>
- Fuentes-Claramonte, P., Martín-Subero, M., Salgado-Pineda, P., Alonso-Lana, S., Moreno-Alcázar, A., Argila-Plaza, I., Santo-Angles, A., Albajes-Eizaguirre, A., Anguera-Camós, M., Capdevila, A., Sarró, S., McKenna, P. J., Pomarol-Clotet, E., & Salvador, R. (2019). Shared and differential default-mode related patterns of activity in an autobiographical, a self-referential and an attentional task. *PLoS One*, 14(1), e0209376. <https://doi.org/10.1371/journal.pone.0209376>
- Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2005). Understanding the effects of task-specific practice in the brain: Insights from individual-differences analyses. *Cognitive, Affective, & Behavioral Neuroscience*, 5(2), 235–245. <https://doi.org/10.3758/CABN.5.2.235>
- Gao, S., Greene, A. S., Constable, R. T., & Scheinost, D. (2019). Combining multiple connectomes improves predictive modeling of phenotypic

- measures. *NeuroImage*, 201, 116038. <https://doi.org/10.1016/j.neuroimage.2019.116038>
- Glasser, M. F., Smith, S. M., Marcus, D. S., Andersson, J. L. R., Auerbach, E. J., Behrens, T. E. J., Coalson, T. S., Harms, M. P., Jenkinson, M., Moeller, S., Robinson, E. C., Sotiropoulos, S. N., Xu, J., Yacoub, E., Ugurbil, K., & Van Essen, D. C. (2016). The Human Connectome Project's neuroimaging approach. *Nature Neuroscience*, 19(9), 1175–1187. <https://doi.org/10.1038/nn.4361>
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80, 105–124. <https://doi.org/10.1016/j.neuroimage.2013.04.127>
- Greene, A. S., Gao, S., Noble, S., Scheinost, D., & Constable, R. T. (2020). How tasks change whole-brain functional organization to reveal brain-phenotype relationships. *Cell Reports*, 32(8), 108066. <https://doi.org/10.1016/j.celrep.2020.108066>
- Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, 9(1), 1. <https://doi.org/10.1038/s41467-018-04920-3>
- Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., Zsoldos, E., Ebmeier, K. P., Filippini, N., Mackay, C. E., Moeller, S., Xu, J., Yacoub, E., Baselli, G., Ugurbil, K., Miller, K. L., & Smith, S. M. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, 95, 232–247. <https://doi.org/10.1016/j.NEUROIMAGE.2014.03.034>
- Gur, R. C., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., Kohler, C., Siegel, S. J., & Gur, R. E. (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology*, 25(5), 766–776. [https://doi.org/10.1016/S0893-133X\(01\)00278-0](https://doi.org/10.1016/S0893-133X(01)00278-0)
- Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., Brensinger, C., & Gur, R. E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods*, 187(2), 254–262. <https://doi.org/10.1016/j.jneumeth.2009.11.017>
- Hamdan, S., More, S., Sasse, L., Komeyer, V., Patil, K. R., & Raimondo, F. (2023). Julearn: An easy-to-use library for leakage-free evaluation and inspection of ML models. <https://doi.org/10.48550/ARXIV.2310.12568>
- Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., & Weinberger, D. R. (2002). The amygdala response to emotional stimuli: A comparison of faces and scenes. *NeuroImage*, 17(1), 317–323. <https://doi.org/10.1006/nimg.2002.1179>
- He, L., Liu, W., Zhuang, K., Meng, J., & Qiu, J. (2021). Executive function-related functional connectomes predict intellectual abilities. *Intelligence*, 85, 101527. <https://doi.org/10.1016/j.intell.2021.101527>
- He, N., Rolls, E. T., Zhao, W., & Guo, S. (2020). Predicting human inhibitory control from brain structural MRI. *Brain Imaging and Behavior*, 14(6), 2148–2158. <https://doi.org/10.1007/s11682-019-00166-9>
- Heckner, M. K., Cieslik, E. C., Patil, K. R., Gell, M., Eickhoff, S. B., Hoffstaedter, F., & Langner, R. (2023). Predicting executive functioning from functional brain connectivity: Network specificity and age effects. *Cerebral Cortex*, 33, 6495–6507. <https://doi.org/10.1093/cercor/bhac520>
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Hennion, S., Delbeuck, X., Koelkebeck, K., Brion, M., Tyvaert, L., Plomhause, L., Derambure, P., Lopes, R., & Szurhaj, W. (2016). A functional magnetic resonance imaging investigation of theory of mind impairments in patients with temporal lobe epilepsy. *Neuropsychologia*, 93, 271–279. <https://doi.org/10.1016/j.neuropsychologia.2016.11.007>
- Herrmann, L., Vicheva, P., Kasties, V., Danyeli, L. V., Szyck, G. R., Denzel, D., Fan, Y., Meer, J. V. D., Vester, J. C., Eskoetter, H., Schultz, M., & Walter, M. (2020). fMRI revealed reduced amygdala activation after Nx4 in mildly to moderately stressed healthy volunteers in a randomized, placebo-controlled, cross-over trial. *Scientific Reports*, 10(1), 3802. <https://doi.org/10.1038/s41598-020-60392-w>
- Itahashi, T., Kosibaty, N., Hashimoto, R.-I., & Aoki, Y. Y. (2021). Prediction of life satisfaction from resting-state functional connectome. *Brain and Behavior*, 11(9), e2331. <https://doi.org/10.1002/brb3.2331>
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Jiang, R., Zuo, N., Ford, J. M., Qi, S., Zhi, D., Zhuo, C., Xu, Y., Fu, Z., Bustillo, J., Turner, J. A., Calhoun, V. D., & Sui, J. (2020). Task-induced brain connectivity promotes the detection of individual differences in brain-behavior relationships. *NeuroImage*, 207, 116370. <https://doi.org/10.1016/j.neuroimage.2019.116370>
- Kandaleft, D., Murayama, K., Roesch, E., & Sakaki, M. (2022). Resting-state functional connectivity does not predict individual differences in the effects of emotion on memory. *Scientific Reports*, 12(1), 14481. <https://doi.org/10.1038/s41598-022-18543-8>
- Kennedy, K. M., Boylan, M. A., Rieck, J. R., Foster, C. M., & Rodrigue, K. M. (2017). Dynamic range in BOLD modulation: Lifespan aging trajectories and association with performance. *Neurobiology of Aging*, 60, 153–163. <https://doi.org/10.1016/j.neurobiolaging.2017.08.027>
- Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., Sun, N., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2019). Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. *Cerebral Cortex*, 29(6), 2533–2551. <https://doi.org/10.1093/cercor/bhy123>
- Lamichhane, B., Westbrook, A., Cole, M. W., & Braver, T. S. (2020). Exploring brain-behavior relationships in the N-back task. *NeuroImage*, 212, 116683. <https://doi.org/10.1016/j.neuroimage.2020.116683>
- Langner, R., Leiberg, S., Hoffstaedter, F., & Eickhoff, S. B. (2018). Towards a human self-regulation system: Common and distinct neural signatures of emotional and behavioural control. *Neuroscience and Biobehavioral Reviews*, 90(May), 400–410. <https://doi.org/10.1016/j.neubiorev.2018.04.022>
- Larabi, D. I., Gell, M., Amico, E., Eickhoff, S. B., & Patil, K. R. (2021). Spatially localized fMRI metrics as predictive and highly distinct state-independent fingerprints. *bioRxiv* [Preprint]. <https://doi.org/10.1101/2021.08.03.454862>
- Li, M., Wang, D., Ren, J., Langs, G., Stoecklein, S., Brennan, B. P., Lu, J., Chen, H., & Liu, H. (2019). Performing group-level functional image analyses based on homologous functional regions mapped in individuals. *PLOS Biology*, 17(3), e2007032. <https://doi.org/10.1371/journal.pbio.2007032>
- Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoun, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., ... Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902), 654–660. <https://doi.org/10.1038/s41586-022-04492-9>
- McCormick, E. M., Arnemann, K. L., Ito, T., Hanson, S. J., & Cole, M. W. (2022). Latent functional connectivity underlying multiple brain states. *Network Neuroscience*, 1–21, 570–590. https://doi.org/10.1162/netn_a_00234
- McFarland, D. J. (2012). A single g factor is not necessary to simulate positive correlations between cognitive tests. *Journal of Clinical and Experimental Neuropsychology*, 34(4), 378–384. <https://doi.org/10.1080/13803395.2011.645018>

- Miller, M. B., Van Horn, J. D., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., Grafton, S., & Gazzaniga, M. S. (2002). Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience*, 14(8), 1200–1214. <https://doi.org/10.1162/089892902760807203>
- More, S., Antonopoulos, G., Hoffstaedter, F., Caspers, J., Eickhoff, S. B., & Patil, K. R. (2023). Brain-age prediction: A systematic comparison of machine learning workflows. *NeuroImage*, 270, 119947. <https://doi.org/10.1016/j.neuroimage.2023.119947>
- Mossad, S. I., Vandewouw, M. M., de Villa, K., Pang, E. W., & Taylor, M. J. (2022). Characterising the spatial and oscillatory unfolding of theory of mind in adults using fMRI and MEG. *Frontiers in Human Neuroscience*, 16, 921347. <https://doi.org/10.3389/fnhum.2022.921347>
- Mueller, S., Wang, D., Fox, M. D., Yeo, B. T. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J., & Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3), 586–595. <https://doi.org/10.1016/j.neuron.2012.12.028>
- Müller, V. I., Cieslik, E. C., Laird, A. R., Fox, P. T., Radua, J., Mataix-Cols, D., Tench, C. R., Yarkoni, T., Nichols, T. E., Turkeltaub, P. E., Wager, T. D., & Eickhoff, S. B. (2018). Ten simple rules for neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, 84, 151–161. <https://doi.org/10.1016/j.neubiorev.2017.11.012>
- Nadeau, C., & Bengio, Y. (1999). Inference for the generalization error. *Advances in Neural Information Processing Systems*, 12, 307–313.
- Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, 203, 116157. <https://doi.org/10.1016/j.neuroimage.2019.116157>
- Nord, C. L., Gray, A., Charpentier, C. J., Robinson, O. J., & Roiser, J. P. (2017). Unreliability of putative fMRI biomarkers during emotional face processing. *NeuroImage*, 156, 119–127. <https://doi.org/10.1016/j.neuroimage.2017.05.024>
- Nostro, A. D., Müller, V. I., Varikuti, D. P., Pläschke, R. N., Hoffstaedter, F., Langner, R., Patil, K. R., & Eickhoff, S. B. (2018). Predicting personality from network-based resting-state functional connectivity. *Brain Structure and Function*, 223(6), 2699–2719. <https://doi.org/10.1007/s00429-018-1651-z>
- O'Connor, D., Potler, N. V., Kovacs, M., Xu, T., Ai, L., Pellman, J., Vanderwal, T., Parra, L. C., Cohen, S., Ghosh, S., Escalera, J., Grant-Villagas, N., Osman, Y., Bui, A., Craddock, R. C., & Milham, M. P. (2017). The healthy brain network serial scanning initiative: A resource for evaluating inter-individual differences and their reliabilities across scan conditions and sessions. *GigaScience*, 6(2), giw011. <https://doi.org/10.1093/gigascience/giw011>
- Ooi, L. Q. R., Chen, J., Zhang, S., Kong, R., Tam, A., Li, J., Dhamala, E., Zhou, J. H., Holmes, A. J., & Yeo, B. T. T. (2022). Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI. *NeuroImage*, 263, 119636. <https://doi.org/10.1016/j.neuroimage.2022.119636>
- Patil, I., Calò, M., Fornasier, F., Young, L., & Silani, G. (2017). Neuroanatomical correlates of forgiving unintentional harms. *Scientific Reports*, 7(1), 45967. <https://doi.org/10.1038/srep45967>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pläschke, R. N., Cieslik, E. C., Müller, V. I., Hoffstaedter, F., Plächti, A., Varikuti, D. P., Goosses, M., Latz, A., Caspers, S., Jockwitz, C., Moebus, S., Gruber, O., Eickhoff, C. R., Reetz, K., Heller, J., Südmeyer, M., Mathys, C., Caspers, J., Grefkes, C., ... Eickhoff, S. B. (2017). On the integrity of functional brain networks in schizophrenia, Parkinson's disease, and advanced age: Evidence from connectivity-based single-subject classification. *Human Brain Mapping*, 38(12), 5845–5858. <https://doi.org/10.1002/hbm.23763>
- Pläschke, R. N., Patil, K. R., Cieslik, E. C., Nostro, A. D., Varikuti, D. P., Plächti, A., Lösche, P., Hoffstaedter, F., Kalenscher, T., Langner, R., & Eickhoff, S. B. (2020). Age differences in predicting working memory performance from network-based functional connectivity. *Cortex*, 132, 441–459. <https://doi.org/10.1016/j.cortex.2020.08.012>
- Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of best practices for evidence for prediction: A review. *JAMA Psychiatry*, 77(5), 534–540. <https://doi.org/10.1001/jamapsychiatry.2019.3671>
- Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, 72(4), 665–678. <https://doi.org/10.1016/j.neuron.2011.09.006>
- Reid, A. T., Bzdok, D., Genon, S., Langner, R., Müller, V. I., Eickhoff, C. R., Hoffstaedter, F., Cieslik, E.-C., Fox, P. T., Laird, A. R., Amunts, K., Caspers, S., & Eickhoff, S. B. (2016). ANIMA: A data-sharing initiative for neuroimaging meta-analyses. *NeuroImage*, 124, 1245–1253. <https://doi.org/10.1016/j.neuroimage.2015.07.060>
- Rosenberg, M. D., Scheinost, D., Greene, A. S., Avery, E. W., Kwon, Y. H., Finn, E. S., Ramani, R., Qiu, M., Constable, R. T., & Chun, M. M. (2020). Functional connectivity predicts changes in attention observed across minutes, days, and months. *Proceedings of the National Academy of Sciences*, 117(7), 3797–3807. <https://doi.org/10.1073/pnas.1912226117>
- Rosenberg, M. D., Zhang, S., Hsu, W.-T., Scheinost, D., Finn, E. S., Shen, X., Constable, R. T., Li, C.-S. R., & Chun, M. M. (2016). Methylphenidate modulates functional network connectivity to enhance attention. *Journal of Neuroscience*, 36(37), 9547–9557. <https://doi.org/10.1523/JNEUROSCI.1746-16.2016>
- Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, 60(1), 830–846. <https://doi.org/10.1016/j.neuroimage.2011.11.050>
- Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90, 449–468. <https://doi.org/10.1016/j.neuroimage.2013.11.046>
- Salsman, J. M., Butt, Z., Pilkonis, P. A., Cyranowski, J. M., Zill, N., Hendrie, H. C., Kupst, M. J., Kelly, M. A. R., Bode, R. K., Choi, S. W., Lai, J.-S., Griffith, J. W., Stoney, C. M., Brouwers, P., Knox, S. S., & Cella, D. (2013). Emotion assessment using the NIH toolbox. *Neurology*, 80(11 Suppl 3), S76–S86. <https://doi.org/10.1212/WNL.0b013e3182872e11>
- Sasse, L., Larabi, D. I., Omidvarnia, A., Jung, K., Hoffstaedter, F., Jocham, G., Eickhoff, S. B., & Patil, K. R. (2023). Intermediately synchronised brain states optimise trade-off between subject specificity and predictive capacity. *Communications Biology*, 6(1), 705. <https://doi.org/10.1038/s42003-023-05073-w>
- Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughhead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64, 240–256. <https://doi.org/10.1016/j.neuroimage.2012.08.052>
- Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, 28(9), 3095–3114. <https://doi.org/10.1093/cercor/bhx179>
- Schumann, F., Steinborn, M. B., Flehmig, H. C., Kürten, J., Langner, R., & Huestegge, L. (2022). On doing multi-act arithmetic: A multitrait-multimethod approach of performance dimensions in integrated multitasking. *Frontiers in Psychology*, 13, 946626. <https://doi.org/10.3389/fpsyg.2022.946626>

- Shah, L. M., Cramer, J. A., Ferguson, M. A., Birn, R. M., & Anderson, J. S. (2016). Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain and Behavior*, 6(5), e00456. <https://doi.org/10.1002/brb3.456>
- Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, 12(3), 506–518. <https://doi.org/10.1038/nprot.2016.178>
- Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A. Z., Vu, A. T., ... Glasser, M. F. (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage*, 80, 144–168. <https://doi.org/10.1016/j.neuroimage.2013.05.039>
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23, S208–S219. <https://doi.org/10.1016/j.neuroimage.2004.07.051>
- Snoek, L., Van Der Miesen, M. M., Beemsterboer, T., Van Der Leij, A., Eigenhuis, A., & Steven Scholte, H. (2021). The Amsterdam open MRI collection, a set of multimodal MRI datasets for individual difference analyses. *Scientific Data*, 8(1), 85. <https://doi.org/10.1038/s41597-021-00870-6>
- Sripada, C., Angstadt, M., Rutherford, S., Kessler, D., Kim, Y., Yee, M., & Levina, E. (2019). Basic units of inter-individual variation in resting state connectomes. *Scientific Reports*, 9(1), 1900. <https://doi.org/10.1038/s41598-018-38406-5>
- Sripada, C., Angstadt, M., Rutherford, S., Taxali, A., & Shedden, K. (2020). Toward a “treadmill test” for cognition: Improved prediction of general cognitive ability from the task activated brain. *Human Brain Mapping*, 41(12), 3186–3197. <https://doi.org/10.1002/hbm.25007>
- Stark, G. F., Avery, E. W., Rosenberg, M. D., Greene, A. S., Gao, S., Scheinost, D., Todd Constable, R., Chun, M. M., & Yoo, K. (2021). Using functional connectivity models to characterize relationships between working and episodic memory. *Brain and Behavior*, 11(8), e02105. <https://doi.org/10.1002/brb3.2105>
- Takeuchi, H., Taki, Y., Nouchi, R., Yokoyama, R., Kotozaki, Y., Nakagawa, S., Sekiguchi, A., Iizuka, K., Hanawa, S., Araki, T., Miyauchi, C. M., Sakaki, K., Sassa, Y., Nozawa, T., Ikeda, S., Yokota, S., Magistro, D., & Kawashima, R. (2021). General intelligence is associated with working memory-related functional connectivity change: Evidence from a large-sample study. *Brain Connectivity*, 11(2), 89–102. <https://doi.org/10.1089/brain.2020.0769>
- Tian, Y., & Zalesky, A. (2021). Machine learning prediction of cognition from functional connectivity: Are feature weights reliable? *NeuroImage*, 245, 118648. <https://doi.org/10.1016/j.neuroimage.2021.118648>
- Tomasi, D., & Volkow, N. D. (2020). Network connectivity predicts language processing in healthy adults. *Human Brain Mapping*, 41(13), 3696–3708. <https://doi.org/10.1002/hbm.25042>
- Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge University Press.
- Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113, 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80, 62–79. <https://doi.org/10.1016/j.neuroimage.2013.05.041>
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4(3), 274–290. <https://doi.org/10.1111/j.1745-6924.2009.01125.x>
- Wang, D., Buckner, R. L., Fox, M. D., Holt, D. J., Holmes, A. J., Stoecklein, S., Langs, G., Pan, R., Qian, T., Li, K., Baker, J. T., Stufflebeam, S. M., Wang, K., Wang, X., Hong, B., & Liu, H. (2015). Parcellating cortical functional networks in individuals. *Nature Neuroscience*, 18(12), 1853–1860. <https://doi.org/10.1038/nn.4164>
- Wang, X., Baeken, C., Fang, M., Qiu, J., Chen, H., & Wu, G.-R. (2019). Predicting trait-like individual differences in fear of pain in the healthy state using gray matter volume. *Brain Imaging and Behavior*, 13(5), 1468–1473. <https://doi.org/10.1007/s11682-018-9960-7>
- Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, 21(4), 1732–1747. <https://doi.org/10.1016/j.neuroimage.2003.12.023>
- Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., & Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45(1), S173–S186. <https://doi.org/10.1016/j.neuroimage.2008.10.055>
- Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, 14(6), 1370–1386. <https://doi.org/10.1006/nimg.2001.0931>
- Yarkoni, T., & Braver, T. S. (2010). Cognitive neuroscience approaches to individual differences in working memory and executive control: Conceptual and methodological issues. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of individual differences in cognition* (pp. 87–107). Springer. https://doi.org/10.1007/978-1-4419-1210-7_6
- Yeung, A. W. K., More, S., Wu, J., & Eickhoff, S. B. (2022). Reporting details of neuroimaging studies on individual traits prediction: A literature survey. *NeuroImage*, 256, 119275. <https://doi.org/10.1016/j.neuroimage.2022.119275>
- Yoo, K., Rosenberg, M. D., Hsu, W.-T., Zhang, S., Li, C.-S. R., Scheinost, D., Constable, R. T., & Chun, M. M. (2018). Connectome-based predictive modeling of attention: Comparing different functional connectivity features and prediction methods across datasets. *NeuroImage*, 167, 11–22. <https://doi.org/10.1016/j.neuroimage.2017.11.010>
- Zhuang, K., Yang, W., Li, Y., Zhang, J., Chen, Q., Meng, J., Wei, D., Sun, J., He, L., Mao, Y., Wang, X., Vatansever, D., & Qiu, J. (2021). Connectome-based evidence for creative thinking as an emergent property of ordinary cognitive operations. *NeuroImage*, 227, 117632. <https://doi.org/10.1016/j.neuroimage.2020.117632>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kraljević, N., Langner, R., Küppers, V., Raimondo, F., Patil, K. R., Eickhoff, S. B., & Müller, V. I. (2024). Network and state specificity in connectivity-based predictions of individual behavior. *Human Brain Mapping*, 45(8), e26753. <https://doi.org/10.1002/hbm.26753>