Neuroscience

Brain-Cognitive Gaps in relation to Dopamine and Health-related Factors: Insights from AI-Driven Functional Connectome Predictions

Reviewed Preprint
v1 • February 14, 2025
Not revised

Morteza Esmaeili, Erin Beate Bjørkeli, Robin Pedersen, Farshad Falahati, Jarkko Johansson, Kristin Nordin, Nina Karalija, Lars Bäckman, Lars Nyberg, Alireza Salami

Department of Electrical Engineering and Computer Science, University of Stavanger, Stavanger, Norway • Department of Diagnostic Imaging, Akershus University Hospital, Lørenskog, Norway • Institute of Clinical Medicine, University of Oslo, Oslo, Norway • Wallenberg Center for Molecular Medicine (WCMM), Umeå University, Umeå, Sweden • Department of Medical and Translational Biology, Umeå University, Umeå, Sweden • Aging Research Center, Karolinska Institute and Stockholm University, Solna, Sweden • Department of Diagnostics and Intervention, Diagnostic Radiology, Umeå University, Umeå, Sweden • Umeå Center for Functional Brain Imaging (UFBI), Umeå University, Umeå, Sweden

eLife Assessment

This multimodal neuroimaging study leverages fMRI, PET, and deep learning to predict memory performance. The authors introduce the brain-cognition gap to link these different imaging modalities to cognition and evaluate their results in two independent cohorts. The results are a **valuable** addition to the literature and will be of interest to neuroscientists working at the interface of cognition, neuroimaging, and computational modeling. However, the evidence supporting the conclusions remains **incomplete**.

https://doi.org/10.7554/eLife.104053.1.sa3

Abstract

A key question in human neuroscience is to understand how individual differences in brain function are related to cognitive differences. However, the optimal condition of brain function to study between-person differences in cognition remains unclear. Additionally, there is a lack of objective biomarkers to accurately predict cognitive function, with brain age emerging as a potential candidate. Recent research suggests that brain age offers minimal additional information on cognitive decline beyond what chronological age provides, prompting a shift toward approaches focused directly on cognitive prediction. Using a novel deep learning approach, we evaluated the predictive power of the functional connectome during various states (resting state, movie-watching, and n-back) on episodic memory and working memory performance. Our findings show that while task-based connectomes, especially during movie watching, better predict working memory, resting state connectomes are equally effective in predicting episodic memory. Furthermore, individuals with a

negative brain-cognition gap (where brain predictions underestimate actual performance) exhibited lower physical activity, lower education, and higher cardiovascular risk compared to those with a positive gap. This shows that knowledge of the brain-cognition gap provides insights into factors contributing to cognitive resilience. Further lower PET-derived measures of dopamine binding were linked to a greater brain-cognition gap, mediated by regional functional variability. Together, our study introduces the brain-cognitive gap, as a new marker, modulated by the dopamine system, to identify individuals at risk of compromised brain function.

Introduction

Resting state fMRI serves as a tool to map brain function (1, 2). Functional connectivity (FC) at rest, estimated with methods to gauge correlated spontaneous activity between two or more regions, serves as a measure of functional brain integrity (1, 3). Individual differences in FC at rest are associated with differences in various cognitive domains (4–6). Associations of this nature have been reported for several large-scale networks, including the default mode network (DMN) (7–9), the frontoparietal network (10–12), the dorsal attention network (13), and the salience network (14). Moreover, such associations have also been identified in more comprehensive investigations spanning the entire functional brain repertoire of the brain (15, 16). Most previous reports focused on associations, not predictions, utilizing the correlation between FC and behavioral phenotypes, which tend to overfit the data and therefore fail to generalize (17). Proper cross-validation, preferably in an independent sample, is therefore important to assert reliable population-level inferences (18). Recent machine learning-based predictive frameworks offer powerful tools for assessing the predictability of individual behavioral phenotypes based on brain connectivity (19–23). In particular, deep neural networks (DNN) methods have been successfully applied to behavioral and disease prediction (24–26), and have been found to outperform other machine learning approaches (27–29).

The functional connectome has demonstrated predictive utility regarding trait-like cognitive phenotypes (30–33). The predictive-modeling framework of the functional connectome has been applied to various cognitive domains, including fluid intelligence (34), working memory (WM) (35), visuospatial ability (36), attention (37), creativity (38), as well as personality traits (39). Understanding the patterns contributing to predictions could offer insights into the functional organization underlying cognitive phenotypes, serving as biomarkers indicating current or prospective health conditions (40–42). Moreover, the whole-brain functional connectome acts as a fingerprint with accurate identification of subjects from a large population (43) within the same cognitive state (e.g., rest-to-rest) but also across different states (e.g., rest-to-task). Overall, past research suggests that the functional connectome is relatively robust within individuals, is unique across individuals, and can predict cognitive and personality phenotypes.

Despite consensus on the value of resting state functional connectivity for mapping brain function, there is an ongoing debate about whether rest is the optimal brain state for investigating individual differences in neurocognitive function (44–47). A study using data from the human connectome project has shown that resting state fMRI predicts differences in brain activity during various tasks, including social, language, relational, and motor tasks (48). This finding supports the notion that individual differences in neural activation can be predicted from resting state (46). However, results from the same dataset revealed that task-based FC outperforms resting state FC in predicting individual differences in fluid intelligence, with task-based data explaining 20 % of the variance compared to just 6 % explained by rest (49). Consistent with this result, previous studies have investigated trait-as well as state-dependent FC, supporting the utility of an integrative approach (11, 47). More recent studies suggested that naturalistic viewing, such as movie-watching, may serve as a happy medium between unconstraint rest and overlay-

constrained tasks in predicting behavior differences (50, 51). Despite the presence of similar spatiotemporal activity patterns across individuals during movie-watching (52), notable individual differences in activity and functional connectivity (53) persist alongside these idiosyncratic features. This suggests that tasks which align individuals' functional connectome more closely to an optimal level, neither completely unconstrained like rest nor overly synchronized like a task, also render them easiest to identify (44). Hence, it is plausible that FC during naturalistic paradigms improves sensitivity to predict behavioral differences (50, 54).

In this study, we first assess the predictive efficacy of the functional connectome across various states for two commonly used cognitive functions. Specifically, we focused on working memory i.e., the ability to store and manipulate information over shorter durations of time, as well as its integral involvement in problem-solving, reasoning, and other key components of fluid intelligence (e.g., (55–57)). In contrast, episodic memory (EM) entails the recollection of specific experiences and events (58), which is regarded as an important element of mind-wandering (59) during resting state.

Past studies have used neuroimaging data, including resting state to predict brain age (60–62). These studies show that brain age, based on biological phenotypes, and their deviation from the chronological age (known as brain age gap prediction), could serve as a biomarker in characterizing disease risk (62–64). Importantly, an older-appearing brain has been shown to be associated with exacerbated physiological and cognitive aging, and risk of mortality (61). However, a recent study suggests that brain age predictions contribute minimally compared to chronological age for explaining cognitive decline (65), implying that cognitive predictions are more reliable. Building upon this study and extending the concept of brain-age gap to brain-cognition gap, it is plausible that a potential disconnection between the brain's biological state and observed cognitive state might emerge. If the brain predicts lower performance than is observed (i.e., a negative brain-cognition gap), it may be compensating for underlying issues not yet apparent through cognitive assessments. By this view, individuals with negative brain-cognition gaps should be less healthy than those whose brains predict higher cognitive function than their actual performance (i.e., a positive brain-cognition gap). Considering the significance of life-style and cardiovascular risk for maintaining healthy brain function (66, 67), the secondary aim of this study was to assess whether individuals with positive versus negative brain-cognitive gaps differ in terms of physical activity habits and cardiovascular risk factor burden.

The third aim of the current study is to uncover the contribution of dopamine (DA) integrity to brain-cognition gaps. Specifically, we investigate whether individuals with lower DA receptor levels exhibit higher functional variability, and in turn greater brain-cognition gaps. DA is a vital neuromodulator with critical implications for motor function, reward-seeking behavior, and various higher-order cognitive functions (68–75). Insufficient DA modulation can affect neurocognitive functions detrimentally (69, 74, 76–78). DA is important for enhancing the signal-to-noise ratio (SNR) of neural transmission (79, 80), and insufficient modulation may contribute to increased variability in blood-oxygen-level-dependent (BOLD) signal (81, 82) and reduced coherence in the functional connectome (83). Indeed, pharmacological studies have shown that DA depletion increases the variability of the BOLD signal, subsequently leading to less synchronized connectivity within resting state networks (84). Consequently, we expect individuals with inadequate DA levels to exhibit increased regional signal variability, a less unique functional connectome, and greater brain-cognition gaps.

Using data from the DopamiNe, Age, connectoMe, and Cognition (DyNAMiC) study (n =180, 20-79 years, 50% female) (54) (**Fig. 1**), we evaluated the predictive power of the functional connectome during resting state, movie-watching, and n-back tasks for two different cognitive domains: episodic memory and working memory. Based on recent research indicating that movie-watching enhances predictability by highlighting key features of FC (50), we hypothesized that FC during movie-watching would outperform FC during rest, and possibly during task, in

predicting both cognitive measures. To achieve this objective, we employed a deep neural network approach, a specific subtype of artificial intelligence (AI), to predict cognitive scores from the functional connectome. Deep learning approaches overcome the limitation of predictive techniques that solely rely on linear associations between connectivity and behavioral phenotypes (17). Considering the importance of individual characteristics, such as age, in predicting behavior from FC (33), we conducted cross-validation of our model, initially derived from an age-heterogeneous sample, in an age-homogeneous sample (from the Cognition, Brain, and Aging (COBRA) study (85)). We subsequently investigated whether individuals with positive prediction gaps differ from those with negative gaps in terms of lifestyle and cardiovascular disease risk factors. Moreover, we tested the hypothesis of whether individuals with lower striatal dopamine D1-like receptor availability (D1DR), the brain's most abundant DA receptor subtype, have a less distinctive FC pattern (i.e., more regional variability) and, in turn, a larger brain-cognition gap. Finally, we conducted a cross-validation of the link between DA and the prediction gap in an independent cohort with estimates of DA D2-like receptor (D2DR) availability (85).

Results

AI-Driven Predictive Modeling of Cognition Scores from the Functional Connectome

We used fMRI data from the DyNAMiC project, in which each subject underwent rest, movie-watching, and working memory n-back tasks. These data were parcellated into 273 nodes (264 with 9 additional subcortical nodes) using a previously published whole-brain functional atlas (86). The averaged time series of 273 regions were subsequently correlated, to create the FC matrix for each participant and cognitive state (rest, movie-watching, and n-back). We found FC to be significantly correlated between cognitive states (r^2 for rest versus movie-watching =.81; r^2 for rest versus n-back =.76; r^2 for movie-watching versus n-back =.71), suggesting that a common functional architecture is present across various tasks.

We trained a convolutional neural network, DenselyAttention, derived from DenseNet (87) on FC matrices from each condition (resting state, movie-watching, and n-back) to generate cognition-specific prediction models of two memory domains (EM and WM). Model performance was quantified as the correlation between observed and predicted cognitive performance. Each model was then tested on all three conditions to examine the generalizability of each model across cognitive states. For example, the model trained on the resting state data (orange circles in **Fig. 1**) was used to predict EM scores using the test dataset derived from resting state (**Fig. 1a**), movie-watching, and n-back. Significance of our main predictions was assessed via linear correlations, and uncorrected p values are presented in **Tables 1**-**2**.

Resting state and movie-watching models outperform n-back in episodic-memory prediction, with resting state offering the best generalizability

We first started with all cases in which congruent conditions were used for model building and prediction. Only models derived from the rest and movie-watching datasets yielded significant predictions of EM (**Figs. 1a-b** and **Table 1**), with resting state yielding the best-performing model (r^2 =25.0%, p <.0001) followed by movie-watching (r^2 =24.0%, p <.0001) (**Table 1**). While there was no significant difference between resting state and movie watching in predicting episodic memory (Mann–Whitney U test, p =.17), both models yielded better EM prediction than n-back (Mann–Whitney U test, p <.0001; **Fig. 1e**).

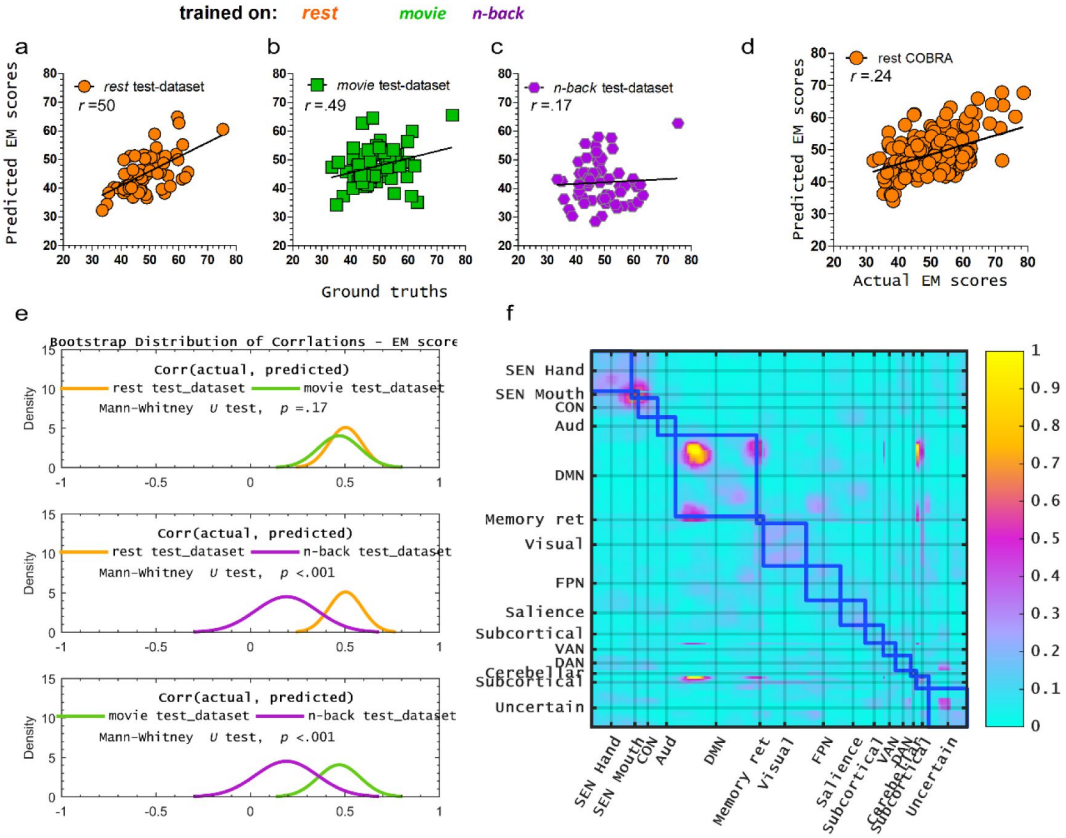


Figure 1.

Model trained on functional connectivity maps acquired at rest predicts (**a**) episodic memory (EM) of the test dataset. The models trained on movie-watching (**b**) but not n-back (**c**) datasets predicted the EM scores of the test dataset. **Table 1** summarizes the p values and correlations for each model. Test datasets were obtained from the same cohort for rest, movie-watching, and n-back. The winning model trained on EM at rest was evaluated on the external COBRA cohort and yielded a significant prediction of EM scores (**d**). Bootstrap distributions of correlations between predicted and actual EM scores indicated no significant difference in the predictive power of EM between models trained on resting state and movie-watching data (**e**). Additionally, the bootstrap distribution revealed that models trained on resting state and movie-watching data yielded higher correlations than those trained on n-back data (**e**). Visualization of features contributing to successful prediction of EM at rest (**f**). A grad-CAM-derived saliency map displays the features that contributed to the model's predictions. The hot spots overlaid on the FC map demonstrate noticeable cross-correlation contributions in "default mode" (DMN) regions. Another important feature visualized by Grad-CAM includes off-diagonal hot spots reflecting inter-connections of the DMN – "subcortical" node.

Table 1.

Correlation results for episodic memory (EM) score predictions

Model trained on	rest			movie			n-back		
Model tested on	rest	movie	n-back	rest	movie	n-back	rest	movie	n-back
Correlation	**.50**	**.44**	**.38**	**.28**	**.49**	**.09**	.05	.08	.17
Significance	**<.0001**	**.0004**	**.003**	**.02**	**<.0001**	**.49**	.76	.84	.42
COBRA									
Correlation	**.24**								
Significance	**<.0001**								

Table 2.

Correlation results for working memory (WM) score predictions.

Model trained on	rest			movie			n-back		
Model tested on	rest	movie	n-back	rest	movie	n-back	rest	movie	n-back
Correlation	.02	.08	.12	**.42**	**.57**	**.46**	.20	**.46**	**.47**
Significance	.88	.72	.15	**<.0001**	**<.0001**	**.004**	.12	**.0002**	**<.0001**
COBRA									
Correlation						**0.47**			
Significance						**<.0001**			

To test the generalizability of these models, two types of validation analyses were performed: cross-condition and cross-data set. In the cross-condition analysis, models trained on one condition (e.g., rest) were tested on an incongruent condition (e.g., movie-watching, n-back; **Table 1**). Interestingly, the model trained on resting state significantly predicted EM when tested on movie-watching (r^2 =19.36%, p <.001) and n-back (r^2 =14.44%, p =.003) conditions. In contrast, models trained on movie-watching or n-back could not be generalized to other conditions, unable to significantly predict EM (p's >.1), except for significant generalizability from the movie to the rest condition (**Table 1**).

In a cross-dataset validation analysis, the best-performing model from the age-heterogenous DyNAMiC dataset was tested on the corresponding condition in an age-homogeneous cohort from the COBRA dataset. By doing this, we found that the resting state model derived from DyNAMiC significantly predicted EM performance in the COBRA dataset (r^2 =5.76%, p <.0001).

Overall, our results suggest that both resting state and movie-watching FC effectively predict EM performance, although resting state models provide better generalizability across different cognitive states and datasets. This finding could further be replicated using the Schaefer Parcellation (**Supplementary Figs. 1-2 and Tables S1-S2**).

Next, we aimed to delineate the relative contributions of different brain regions for the best-performing model—the model trained on the "resting state data" in predicting episodic memory. Utilizing the Grad-CAM algorithm, saliency maps were generated for the 120 FC matrices used during the training of the winning model. An averaged and interpolated description of all saliency maps is depicted in **Figure 1f**. The saliency map highlights specific edges, especially within the default-mode network, edges between the default-mode network and subcortical areas, and edges between the default-mode and the cerebellar network. These edges, indicated by a salience intensity of ≥.5, exert a significant influence on the model (**Fig. 1f**).

Movie-watching and n-back models outperform resting state in working-memory predictions, with movie-watching offering the best generalizability

We next investigated whether the superiority of resting state in predicting EM was unique to this domain, considering that previous research demonstrated the advantages of task-based fMRI and naturalistic viewing in predicting fluid intelligence (88). To do so, we compared the predictive power of different states for WM, which is shown to be more directly associated with fluid intelligence compared to EM.

The model derived from the resting state failed to predict and generalize regarding WM (p's >.10; **Fig. 2a** and **Table 2**). By contrast, models trained on movie-watching and n-back yielded significant predictions of WM (**Figs. 2b-c** and **Table 2**), with the movie-watching model emerging as the best-performing model (r^2 =32.49%, p <.0001) followed by n-back (r^2=22.09%, p <.0001). While there was no significant difference between movie-watching and n-back in predicting WM (Mann–Whitney U test, p =.81), these models yielded better WM prediction than resting state (p <0.0001, Mann–Whitney U test; **Fig. 2e**).

In cross-condition validation (**Table 2**), the movie-watching model yielded a significant WM prediction during both resting state (r^2 =17.64%, p <.0001) and n-back (r^2 =21.16%, p =.004). The model derived from n-back yielded a significant WM prediction during movie-watching (r^2 =21.16%, p =.0002), but not during the resting state (r^2 =4%, p =.12).

For cross-dataset validation, the best-performing model in the DyNAMiC dataset (i.e., movie-watching) was applied to predict WM to the COBRA dataset. However, since COBRA does not include a movie-watching paradigm, we applied the model to the n-back task in COBRA. This

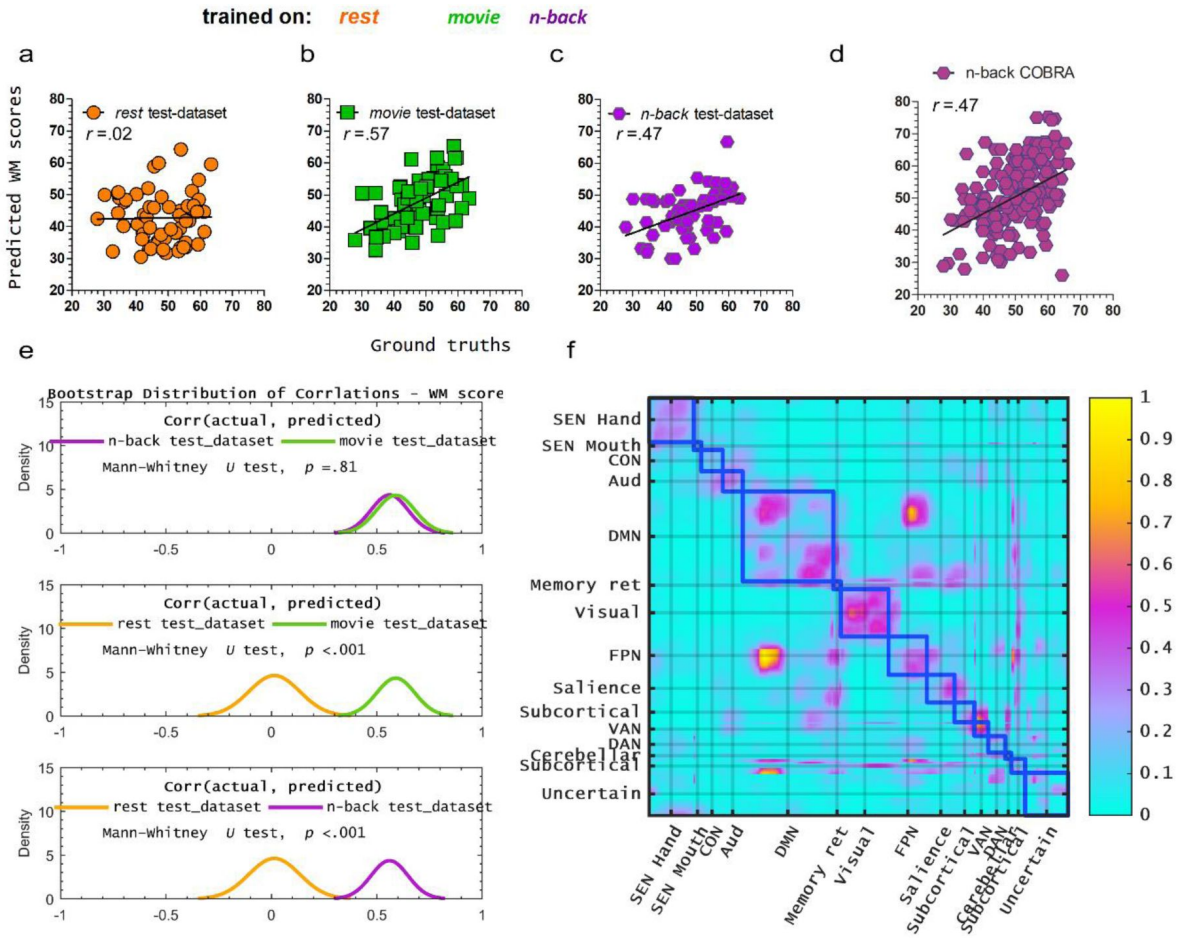


Figure 2.

Model trained on FC maps acquired at rest did not significantly predict (**a**) the working memory (WM) in the test dataset. The model trained on the movie-watching dataset yielded the best-performing model in predicting WM (**b**) while the model trained on the n-back dataset (**c**) was the second-best model. **Table 2** summarizes the p values and correlation power for each model. (**d**) Results of cross-dataset validation, where the best-performing model in the DyNAMiC dataset (i.e., movie-watching) was applied to predict WM to the COBRA dataset. However, since COBRA does not include a movie-watching paradigm, we applied the model to the n-back task in COBRA(**Table 2**). Bootstrap distributions of correlations between predicted and actual WM scores showed no significant difference in predictive power between models trained on movie-watching and n-back data (**e**). The bootstrap distribution revealed that models trained on movie-watching and n-back data exhibited higher correlations than those trained on resting state data (**e**). The Grad-CAM-derived saliency map highlights dominant features in the FC maps that contributed to the model's predictions (**f**). The hot spots overlaid on an FC map demonstrate noticeable cross-correlation contributions in the "VAN", "visual", and to a lesser degree (<.5) "DMN". Other important features visualized by Grad-CAM include off-diagonal hot spots reflecting inter-connections of the "DMN" – with "FPN, Fronto-parietal Task Control", "Subcortical", and "Cerebellar"; "Cerebellar" – "FPN" node.

approach revealed that the model from the DyNAMiC movie-watching condition yielded a significant WM prediction in the COBRA n-back task (r^2 =22.09%, p < 0.0001).

Overall, our results suggest that movie-watching-based WM predictions generalize across different cognitive states and datasets. This finding could be further replicated using a different functional parcellation (**Supplementary Figs. 1-2 and Tables S1-S2**).

The Grad-CAM algorithm generated saliency maps of 120 FC maps from the DyNAMiC dataset employed for training. **Figure 2f** depicts an average of all Grad-CAM generated maps. The saliency map unveils that certain edges, specifically within network connectivity of task-positive regions such as the frontoparietal task control network, dorsal/ventral attention, visual, and subcortical networks as well as between-network connectivity. FC between task-positive dorsal and ventral attention networks, and between the DMN and the fronto-parietal network, contributed to the best-performing model derived from the movie-watching dataset.

The brain-cognition gap is related to physical activity levels and cardiovascular risk factors

Given the importance of lifestyle and cardiovascular risk for maintaining healthy brain function (66, 67, 89), we examined whether individuals with positive versus negative prediction gaps differ in physical activity habits, education, and Framingham cardiovascular disease (CVD) risk score. Our primary focus is on EM predictions derived from resting state data, as this was the common condition in the DyNAMiC and COBRA datasets. We computed the difference between predicted and observed EM scores to generate a gap score. A positive brain-cognition gap indicates that an individual's brain predicted better-than-observed EM performance, whereas a negative gap indicates more compromised brains relative to actual performance. In the test sample of the DyNAMiC data (n =60) and the entire COBRA sample (n =177), we found that individuals with a negative brain-cognition gap exhibited lower levels of physical activity and higher CVD risk scores compared to those with positive gaps (**Fig. 3**). Moreover, individuals with negative brain-cognitive gaps were less educated compared to those with positive brain-cognitive gaps in the DyNAMiC dataset.

Dopamine D1 and D2 receptor availability are associated with brain-cognition gaps

We hypothesized that inadequate DA levels might adversely affect neurocognitive function by reducing the neural signal-to-noise ratio, thereby resulting in a less unique functional connectome, consequently leading to a greater prediction gap. We therefore initially investigated the relationship between DA receptor levels and predictive gaps across different types of DA receptors in the DyNAMiC and COBRA samples.

In the DyNAMiC sample, we found a significant correlation between striatal D1DR and prediction gap (positive gap: r =−.46, p =.03; negative gap: r =.38, p =.01) (**Fig. 4a**), suggesting that lower D1DR is associated with greater brain-cognition gaps.

We replicated our finding in the COBRA sample using D2-like receptor availability (D2DR), revealing a significant relationship between striatal D2DR and prediction gap (positive gap: r =−.38, p =.01; negative gap: r =.39, p =.004) (**Fig. 4b**). Our findings provide support the view that lower D1DR/D2DR is associated with larger brain-cognition prediction gaps.

Figure 3.

The boxplots compare physical activity scores (total hours per week), CVD risk, and educational level (years) between two groups with positive and negative brain-cognition gaps in the DyNAMiC (**top row**) and COBRA (**bottom row**) datasets. The striped cyan-colored boxes represent scores for subjects with a positive brain-cognition gap, while the magenta-colored boxes represent subjects with a negative estimated gap.

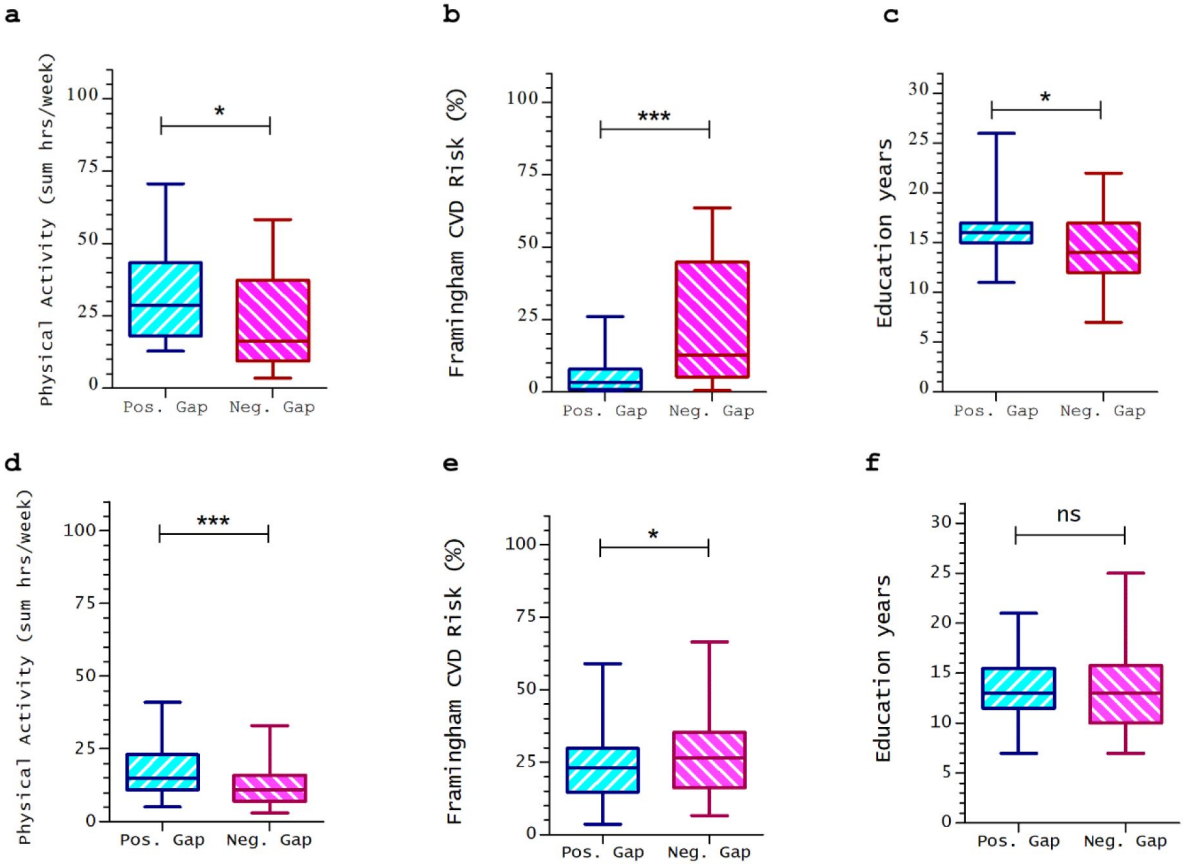
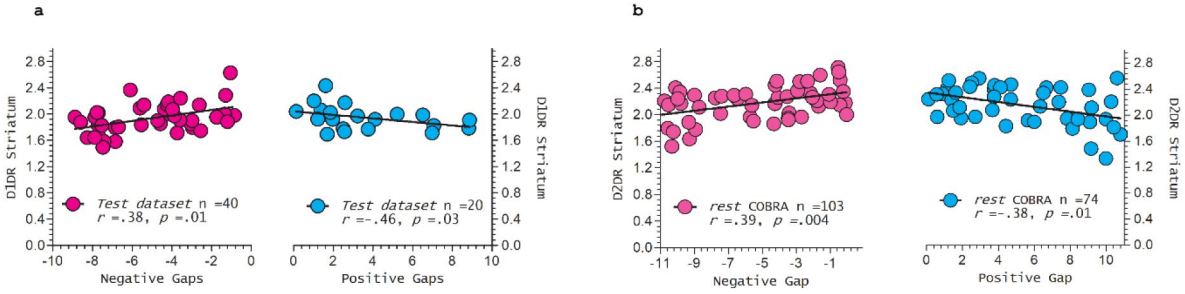


Figure 4.

Relationship between the gap measured from predicted and actual EM scores and dopamine D1 receptor (**a**). Negative gaps indicate that the predicted EM score was lower than actual EM scores, while the positive gaps indicate higher predicted scores than actual EM scores. Partial correlation analysis showed a significant correlation between D1 receptor values and the measured negative and positive gaps. Relationship between the gap measured from predicted and actual EM scores and dopamine D2 receptor (**b**). Negative gaps indicate that the predicted EM score was lower than actual EM scores, while positive gaps indicate higher predicted scores than actual EM scores. Partial correlation analysis showed a significant link between D2 receptor values to negative gaps and positive gaps.



Regional variability mediates the direct impact of dopamine on brain-cognition gaps

To evaluate whether functional variability mediated the relationship between D1DR and prediction gap connectivity, we conducted additional mediation analyses. We computed BOLD signal entropy, which estimates within-region signal variability (84), that was averaged across all regions of the parcellation. In the DyNAMiC sample, for the group exhibiting a negative brain-cognitive gap, we observed a significant indirect effect of D1DR on the gap mediated by entropy, β =2.41, 95% CI [0.89, 4.51], p <.0001, explaining 56.81% of the total effect of D1DR on the gap. The direct effect of D1DR, however, was not significant, β =1.83, 95% CI [−.86, 3.79], p =.19. Similarly, for those with a positive brain-cognition gap, an indirect effect of D1DR on the gap through entropy was observed, β =-6.78, 95% CI [−11.26, −3.05], p <.0001, accounting for 89.41% of the D1DR effect on the gap. The direct effect of D1DR was again non-significant, β =-0.803, 95% CI [−.48, 2.27], p =.26. These findings suggest that lower D1DR levels contribute to increased signal variability, which in turn leads to reduced specificity of FC and, consequently, a larger brain-cognition gap.

In the COBRA sample, within the negative gap group, we detected a significant indirect effect of D2DR on both the negative and positive gap groups through entropy. For the negative gap, β =2.18, 95% CI [0.01, 4.25], p =.04, accounting for 63.43% of the D2DR effect on the gap; and for the positive gap, β =−.2.18, 95% CI [−3.87, −.04], p =.04, accounting for 61.49% of the D2DR effect on the gap.

Discussion

Using deep learning models, we examined the predictive power of the functional connectome during various states (resting state, movie-watching, and n-back) on two different cognitive domains (EM and WM). Both rest and movie-watching states yielded significant predictions of EM, with the model derived from resting state generalized across states and datasets. Differences between the DyNAMiC and COBRA datasets make cross-dataset prediction a harder problem, as the age ranges of samples significantly vary, and prior studies highlight the importance of individual characteristics like age in predicting behavior from FC (33). Possibly as a consequence of these age differences, model performance decreased when using deep learning models to predict EM from COBRA data. Validation outcomes thus affirm that the models, particularly those constructed from rest data, are robust to the particulars of the dataset. The saliency map generated from the final layer of the deep learning model indicates that certain edges within DMN, between DMN and the subcortical network contributed significantly to the prediction. Building on a recent finding by Kurkela and Ritchy (90), our finding reveals that a fraction of the known-memory subnetwork within the DMN, as well as whole-brain multivariate pattern, which notably encompasses interactions of the DMN with other networks, such as the subcortical network, made a more substantial contribution to prediction. Importantly, this prediction generalizes across conditions and datasets, suggesting that features derived from resting state FC serve as a stable marker of individual differences in EM.

Our findings are in contrast to recent work suggesting that task paradigms, in general, and movie-watching, in particular, outperform resting state data in predicting cognitive performance (49, 50, 91). There are two possible explanations for this discrepancy. First, while previous studies have often demonstrated a superiority of task and naturalistic viewing over resting state in predicting fluid intelligence or WM (49, 50), there are fewer reports of FC predicting EM (e.g., (92, 93)) and, to our knowledge, no study, has compared rest and movie-watching. While we acknowledge that the resting state represents a complex amalgamation of cognitive, emotional, and perceptual processes (94), the good prediction power of the resting state may arise from the presence of mind wandering during rest, which is strongly related to EM (95, 96). EM plays a crucial role in generating mental content during mind wandering, especially episodes

characterized by distinct times and locations (59, 97). Second, most previous studies utilized connectome predictive modeling (CPM), which relies on a linear relationship between FC and behavioral data. In contrast, deep learning makes no assumption about the nature of the relationship, and thus, a deep learning model may outperform CPM. Indeed, the predictive power in the current study is stronger than for CPM-based predictions reported before. For example, across various studies, the correlation between predicted and actual fluid intelligence typically hovers around 0.25 (98–100), while an average correlation of approximately 0.11 was reported across 58 different measures covering various domains (101). However, in the current study, we found that deep learning yields a stronger association (resting state prediction of EM: .31< r <.49; movie-watching prediction of WM: .42< r <.57), even across data-sets (cross-validation of EM: r =.24; cross-validation of WM: r =.47, between FC and cognitive data).

In contrast to the EM prediction, both n-back and movie-watching datasets yielded significant predictions of offline WM performance. Importantly, the models derived from movie-watching and n-back outperformed the resting state in WM prediction. These differences in model accuracy when predicting the same target behaviors (i.e., WM) suggest the presence of trait-state interactions. Specifically, movies and n-back enhance individual differences in WM-relevant connection. Indeed, we found that several WM-related networks (102–105), including the fronto-parietal, the salience, and the dorsal/ventral attention networks contribute to prediction. Additionally, previous research showed that movie-watching alters the propagation of activity across cortical pathways (106), particularly within and between regions involved in audiovisual processing and attention. These alterations lead to a less segregated and more integrated network organization (107). Similarly, the n-back task has been associated with increased integration of task-positive cortico-cortical connectivity (105, 108) and striato-cortical connectivity (103). Taken together, our results confirm previous findings that movie watching paradigm is a great condition for studying individual differences in various cognitive domains. Nonetheless, if movie watching paradigm is not feasible/ available, the resting state still provides a robust means of studying individual differences, particularly in self-referential domains, such as EM.

Our study introduces a novel deep neural network architecture that features dense connections and incorporates an attentional mechanism. We found that the current method outperformed widely used approaches, such as kernel regression (110) and connectome predictive modeling (17), in predicting cognitive data. Our findings contrast with previous work showing similar performance between kernel regression and various DMN architectures for predicting behavioral and demographic variables (29). Unlike the BrainNet convolutional neural network, which focuses on staged transformations, our densely connected model promotes extensive feature reuse, possibly leading to more robust feature extraction. That is, the model explicitly uses the Enhanced Residual Block (ERB) and High-Frequency Attention Block (HFAB), enhancing the focus on relevant features and potentially boosting performance. The architecture includes multiple layers and blocks, balancing depth and computational efficiency.

Applying two different parcellation methods (86, 111) to the DyNAMiC data indicated that parcellation resolution does not significantly impact model performance (see **Figs. S1-S2 and Tables S1-S2**). However, improvements in registration and parcellation, such as individualized parcellations and areal features for alignment, may enhance the delineation of functionally homogeneous regions (112), improving the predictive performance of deep learning models. Therefore, the impact of alignment and parcellation on predictive model performance remains a key area for future research. Additionally, the performance of our approach still needs to be examined on a large-scale publicly available neuroimaging dataset for predicting other types of cognitive data.

We found a significant link between the brain-cognition prediction gap, life-style, and risk factors for vascular disease, such that individuals with a negative gap exhibited lower levels of physical activity and higher cardiovascular risk scores compared to those with a positive gap. This finding

was consistent across both age-heterogeneous and age-homogenous older samples. Brain-cognition prediction gap could serve as a biomarker for identifying individuals at risk (e.g., individuals with a negative gap). Previous studies suggest that the brain age prediction gap is associated with cognitive aging (62), some aspects of physiological aging (61), as well as aging in other organs (113), and even mortality in older age (61). However, a recent study revealed that brain age accounts for only a small portion of cognitive decline compared to chronological age (65), suggesting that cognitive prediction might be more informative. Our findings build upon this concept by extending the brain-age gap to behavioral variables, demonstrating that the brain-cognition gap could provide insights regarding physical activity status, education, and cardio vascular risk – key factors contributing to cognitive reserve (114–117).

Critically, we found that D1DRs and D2DRs were strongly associated with the brain-cognition gap, such that lower dopamine receptor levels were associated with greater gaps. More specifically, in two independent samples, we discovered greater correspondence (i.e., near zero in the gap) between brain function and cognition in individuals with higher D1DR/D2DR whereas lower correspondence (i.e., significantly different gap from zero) was found in individuals with lower D1DR/D2DR availability. Previous computational models proposed that DA modulates neuronal gain, which improves SNR in neural processing, contributing to more coherent activity across large-scale networks (e.g., balanced integration and segregation (83)). Past studies also showed that lower D1DR contributed to more BOLD variability in the subcortical area (81) and less functional segregation of the striatum (118) and the large-scale networks in aging (83), possibly due to increased noise (lower SNR). In support of this notion, we found for the first time that regional variability, estimated using entropy, mediated the impact of DA on the brain-cognition gap. Although the cross-sectional nature of our data warrants caution, this novel finding suggests that lower DA integrity relates to BOLD variability, which in turn is associated with a larger brain-cognition gap.

In summary, our findings reveal that while tasks like movie-watching excel in predicting WM performance, there are features during rest that predict internally-oriented mind-wandering-type tasks, such as EM. Additionally, individuals whose brains predict poorer cognitive performance (i.e., negative gap) exhibit lower physical activity and higher cardiovascular risk compared to those whose brains predict higher cognitive function than their actual performance (i.e., positive gap). This finding suggests that our prediction model offers a novel marker to identify individuals at risk of compromised brain maintenance. Furthermore, individuals with lower DA showed less accurate cognitive prediction (larger brain-cognition gap) due to increased BOLD variability and less unique and cohesive FC.

Materials and methods

Participants

All participants provided written informed consent and studies were conducted in accordance with the Declaration of Helsinki and approved by the Regional Ethical Board and the local Radiation Safety Committee (reference numbers: 2012-57-31M; 2017-404-32M).

This study used data from DyNAMiC (54), which is a longitudinal study with a focus on changes in the brain connectome and the D1DR system. At baseline,180 participants (20-79 years, 50% female) across the adult lifespan underwent all tests between 2017 and 2020 (54) (**Fig. 5**). Rigorous exclusion criteria were used to recruit a sample without neurological conditions and medical treatments affecting brain functioning and cognition. Exclusion criteria included brain injury or neurological disorder, dementia, neurodevelopmental disorder, psychiatric diagnosis, psychopharmacological treatment, history of severe head trauma, substance abuse or dependence,

and illicit drug use. Individuals with other chronic or severe medical conditions (e.g., cancer, diabetes, and Parkinson's disease) were also excluded. Here, we only use data from the baseline measurement.

We used a separate sample as a testing dataset from the COBRA study (85) (**Fig. 5**). COBRA is a longitudinal aging study in which 181 healthy individuals (64-68 years, 45% female) underwent baseline assessments of brain, cognition, health, and lifestyle during 2012–2014 (85). Exclusion criteria at baseline included traumatic brain injury, stroke, dementia, intellectual disability, epilepsy, psychiatric and neurological disorders, diabetes, and cancer medications, severe visual or auditory impairment, claustrophobia, and poor Swedish language skills. In the current study, we used data from 177 subjects from COBRA who underwent both MRI and PET examinations at baseline (78, 85, 102, 103, 119).

Cognitive Measures

The same cognitive test battery was used in DyNAMiC and COBRA (Nevalainen et al 2015; Nordin et al, 2022), and assessed two cognitive domains: episodic memory (EM) and working memory (WM). Each domain was assessed using three separate tests, including letter-, number-, and figure-based material, respectively. Participants completed all tasks on a computer and responded by either typing in words or numbers; using the computer mouse; or pressing keys marked by different colors. Each test included initial practice runs, after which testing followed across several trials. For each of the three tests within a domain, scores were summarized across trials for a measure of overall performance. Finally, these sum scores were z-standardized and averaged to form one composite score per cognitive domain (*T* score: *mean* =50; *sd* =10) representing within-domain performance (e.g., EM or WM performance). Missing values were replaced by the average of the available observed scores.

Episodic Memory (EM)

Tests of EM included word recall, number-word recall, and object-location recall. In word recall, participants viewed 16 Swedish concrete nouns, successively appearing on the computer screen. Each word was presented for 6 s, with an inter-stimulus interval (ISI) of 1 s. Directly following encoding, participants reported as many words as they could using the keyboard. Two trials were completed, with a combined maximum score of 32. In number-word recall, participants encoded pairs of 2-digit numbers and concrete plural nouns (e.g., 46 dogs). During encoding, eight number-word pairs were presented, each for 6 s, with an ISI of 1 s. Directly after encoding, nouns were presented again, in re-arranged order, and participants had to report the 2-digit number associated with each presented noun (e.g., How many dogs?). This test included two trials with a total combined maximum score of 16. The third test assessed object-location memory. Participants viewed a 6 × 6 square grid in which 12 objects were consecutively shown at distinct locations. Each object-position pairing was displayed for 8 s, with an ISI of 1 s. Directly following encoding, all objects were simultaneously displayed next to the grid for the participant to place in their correct position within the grid. If unable to recall the correct position of an object, participants had to guess to the best of their ability. Two trials of this test were completed, making the total combined maximum score 24.

Working Memory (WM)

Working memory was examined with three tests letter updating, number updating, and spatial updating. These tests differed from the working memory *n*-back task performed during fMRI scanning. In letter-updating, capital letters (A–D) were consecutively presented on the computer screen, with participants instructed to always keep the three final letters in memory. Each letter was presented for 1 s, with an ISI of 0.5 s. When prompted, at any given moment, participants provided their responses by typing in three letters using the keyboard and provided a guessing-based answer if they failed to remember the correct letters. Four practice trials were followed by

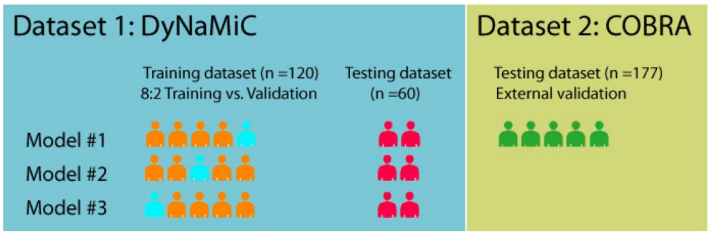


Figure 5.

Overview of the experimental procedure and the use of datasets. We used a 3-fold within-sample (DyNAMiC) cross-validation where we trained our model on 120 subjects (8:2; 80% training:20% validation during training) and tested it in a separate sample of 60 subjects. The winning within-sample model was used for between-sample (COBRA) cross-validation.

16 test trials consisting of either 7-, 9-, 11- or 13-letter sequences. The combined maximum number of correct answers across trials was 48 (16 trials × 3 reported letters = 48). The number-updating test followed a columnized 3-back design. Three boxes were presented next to each other on the screen throughout the task, in which a single digit (1–9) was sequentially presented from left to right for 1.5 s with an ISI of 0.5 s. During an ongoing sequence, participants had to judge whether the number currently presented in a specific box matched the last number presented in the same box (i.e., appearing three numbers before). For each presented number they responded yes/no by pressing assigned keys ("yes" = right index finger; "no" = left index finger). Two practice trials were followed by four test trials each consisting of 30 numbers. The combined maximum number of correct answers across trials (after discarding responses to the first three numbers in every trial – as these were not preceded by any numbers to be matched with) was 108 (27 numbers × 4 trials). In spatial-updating, participants viewed three 3 × 3 square grids presented next to each other on the computer screen. At the start of each trial, a blue circular object was displayed at a random location within each grid. After 4 s, the circular objects disappeared, leaving the grids empty. An arrow then appeared below each grid, indicating that the circular object in the corresponding grid was to be mentally moved one step in the direction of the arrow. The arrows appeared stepwise from left to right grids, each presented for 2.5 s (ISI = 0.5 s). Prompted by three new arrows, participants mentally moved the circular objects one more time, resulting in each circular object having moved two steps from its original location at the end of each trial. Participants indicated which square the circular object in each grid had ended up in using the computer mouse. If unsure, they were instructed to guess. The test was performed across 10 test trials, preceded by five practice trials. The combined maximum number of correct location indications was 30.

Measure of Physical Activity and Cardiovascular Disease Risk

Physical Activity

An extensive battery of self-rating questionnaires was administered in DyNAMiC and COBRA (54, 85). Participants were asked to indicate the frequency (number of hours during a typical summer week; options: 1-14 h with 1-h increments, or 15+ hrs) and the intensity (how physically demanding on a scale from 1 ="not at all" to 5 ="extremely") by which they typically engage in a selection of activities relevant to life in northern Sweden. These included 15 specific activities. For the purpose of the present study, we focused on physical activities, and in particular on those activities that are purely physical and that individuals are sufficiently engaged in (i.e., physically demanding >= 2). Each of these activities was performed by at least 20% of the participants at least once a week; e.g., walking, bicycling, jogging, strength training, household tasks, and daily work-related activities. We computed physical activity frequency (sum hrs/week) to generate physical activity scores accordingly.

Cardiovascular Disease risk

The risk of cardiovascular disease was determined via a multivariable score, according to the algorithm developed in the Framingham Heart Study (120). Variables include age, sex, hypertension diagnosis, systolic blood pressure, body-mass index, smoking, and diabetes mellitus. The risk estimates were derived using an algorithm proposed by D'Agostino et al. (120), which employs Cox proportional-hazard regression models to predict the probability of developing any form of cardiovascular disease within 10 years:

$$\hat{p} = 1 - S_0(t)exp(\sum_{i=1}^{m} \beta_i(X_i - \bar{X}_i)) \quad [1]$$

With $S_0(t)$ being the baseline survival at follow-up t (t =10 years), βi the estimated regression coefficient, X_i the log-transformed value of the ith risk factor, X_i the corresponding mean, and m the number of risk factors included. Baseline survival, means, and regression coefficients were

taken from the original algorithm, with DyNAMiC and COBRA participants' risk variables inserted to compute the final scores. Risk score calculators can be found at *framinghamheartstudy.org*.

Image Acquisitions

Structural, functional, and neurochemical brain measures were acquired using MRI and PET at Umeå University Hospital in northern Sweden. For both DyNAMiC and COBRA, all MRI data were collected using a 3T Discovery MR750 MRI system (General Electric, Healthcare, Illinois, USA) equipped with a 32-channel phased-array head coil. PET was conducted in 3D mode with a Discovery PET/CT 690 (General Electric, WI, United States) to assess whole-brain D1DR with [^{11}C]SCH23390) and D2DR with [^{11}C]raclopride at rest in DyNAMiC and COBRA, respectively. Comprehensive descriptions of MRI, PET, and cognitive testing protocols are given elsewhere (54, 85). In this study, we primarily focus on those data directly pertinent to the current investigation.

Functional MRI

For the DyNAMiC dataset, high-resolution anatomical T1-weighted images were collected using a 3-dimensional (3D) fast spoiled gradient-echo sequence with acquisition parameters of 176 sagittal slices, thickness =1 mm, TR =8.2 ms, TE =3.2 ms, flip angle =12°, and a field of view (FOV) =250×250 mm. Whole-brain functional images were acquired during resting state, naturalistic viewing, and an n-back WM task. Functional images were acquired using a T2*-weighted single-shot echo-planar imaging (EPI) sequence, with 330 volumes collected over 12 min. The sequence provided 37 axial slices, slice thickness =3.4 mm, 0.5 mm spacing, TR =2,000 ms, TE =30 ms, flip angle =80°, and FOV =250×250 mm. Ten dummy scans were collected at the start of the sequence. During the resting state, participants were instructed to keep their eyes open and focus on a white fixation cross on a black background displayed on a computer screen through a tilted mirror attached to the head coil. WM was assessed in the scanner (12-min) with a numerical *n*-back task which consisted of blocks of 1-back, 2-back, and 3-back (102, 103). During movie-watching, participants viewed and listened to a 12-minute video consisting of selected and chronologically ordered sections from the Swedish movie *Cockpit* (121). Participants were instructed to view the movie attentively and answered a short multiple-choice questionnaire about the movie after the scanning session.

In COBRA, data were collected using identical scans for resting state and n-back WM. However, the resting state scan was shorter, lasting only 6 minutes (73, 103).

Functional Connectivity Analysis

Functional data from all conditions (i.e., rest, movie, n-back) were pre-processed using the Statistical Parametric Mapping software package (SPM12). The functional images were first corrected for slice-timing differences and in-scanner motion followed by registration to anatomical T1 images. Distortion correction was performed using subject-specific and T1 co-registered field maps. The functional time series were subsequently demeaned and detrended followed by simultaneous nuisance regression and temporal high-pass filtering (threshold at 0.009 Hz) to not re-introduce nuisance signals (122). The nuisance regression model included mean cerebrospinal and white-matter signal and their derivatives, 24-motion parameters (123), a binary vector flagging motion-contaminated volumes exceeding framewise displacement (FD) of 0.2 mm (124), in addition to an 18-parameter RETRICOR model (125, 126) of cardiac pulsation (up to third-order harmonics), respiration (up to fourth-order harmonics), and first-order cardio-respiratory interactions estimated using the PhysIO Toolbox v.5.0 (127). Regression models for n-back included an additional set of finite impulse response (FIR) task regressors (128) to avoid false positive connectivity due to task-evoked activations (129). The FIR regression approach involved fitting mean cross-block responses for each time point within a time-locked window of equal duration to each task block (27 blocks of 20 s), extended by an

additional 18 s following each block to account for the duration of the hemodynamic response function (HRF). Given that this approach linearly fits a set of binary task regressors for each time point, it is nearly identical to subtracting the mean task response, with the difference being that FIR regression is better able to handle overlapping task responses and differences in the shape of the HRF (129). The nuisance-regressed images were subsequently normalized to sample-specific group templates (DARTEL) (130) for each dataset respectively, followed by spatial smoothing using a 6-mm FWHM Gaussian kernel to mitigate DARTEL-induced aliasing and affine-transformed to stereotactic MNI space (ICBM152NLin2009) (131, 132). Functional images from both DyNAMiC and COBRA were preprocessed according to the steps described above with the only exception that the RETRICOR parameters were excluded from the regression models in COBRA due to technical issues related to the respiration and cardiac traces.

Graph Construction

For all fMRI conditions, functional time series were averaged from 273 cortical and subcortical regions, represented by 5-mm radius spheres, based on a widely employed FC parcellation (86). In addition to 264 regions reported in the Power parcellation, we added 9 additional regions including some subcortical regions, such as putamen, caudate, and anterior and posterior hippocampus, identified using independent component analysis (8, 133). These regions were categorized into 14 resting state networks according to a consensus partition (86). To mitigate sampling from non-gray matter voxels, each parcel underwent erosion by a permissive gray matter mask (eroding voxels <.1% threshold). The averaged time series were then subjected to Pearson's correlations, followed by Fisher's r-to-z transformation, resulting in the creation of a 273×273 adjacency matrix for each participant, with coefficients along the main diagonals set to zero.

To further investigate the impact of network parcellation, we replicated our prediction analysis (**Supplementary Material, Figs S1-S2, and Table S1-S2**) using Schaefer parcellation (111), which entails 300 cortical and subcortical regions.

Deep Neural Network Model

Based on convolutional neural networks, deep learning is an advanced form of artificial intelligence that uses multiple layers of "hidden" neural networks. Deep learning methodologies are capable of automatically identifying complex patterns and representations directly from raw data using these multi-layered networks, thus eliminating the need for explicit feature engineering or manual intervention (134). The success of the training and learning phases depends on the model's ability to process high-dimensional input data, extracting meaningful features from complex data. This is done while managing the number of trainable parameters, which are crucial for automated feature learning during the construction of the model.

In this study, the inputs to our deep learning models were subject-specific FC maps with a matrix size of 273×273 (e.g., **Fig. 6a**). We generated different versions of each FC map by replacing portions of networks. For example, we relocated the DMN network toward the last nodes, which were assigned as DAN, cerebellar, subcortical, and uncertain in **Figure 6a**. This approach allowed us to create a diverse set of FC matrices for each individual, each reflecting a different composition of edges and neighboring while maintaining the linear relationships exonerated in the original data. Consequently, we augmented the dataset by producing a total of 3,600 FC maps from the initial set of FC maps.

Following the DenseNet framework (87), we incorporated the Enhanced Residual Block (ERB) and High-Frequency Attention Block (HFAB) into the dense layers (**Fig. 6b**), termed DenselyAttention, to facilitate feature reuse in each layer. DenseNet (87) diverges from traditional methods like deepening layers or widening network structure by focusing on feature

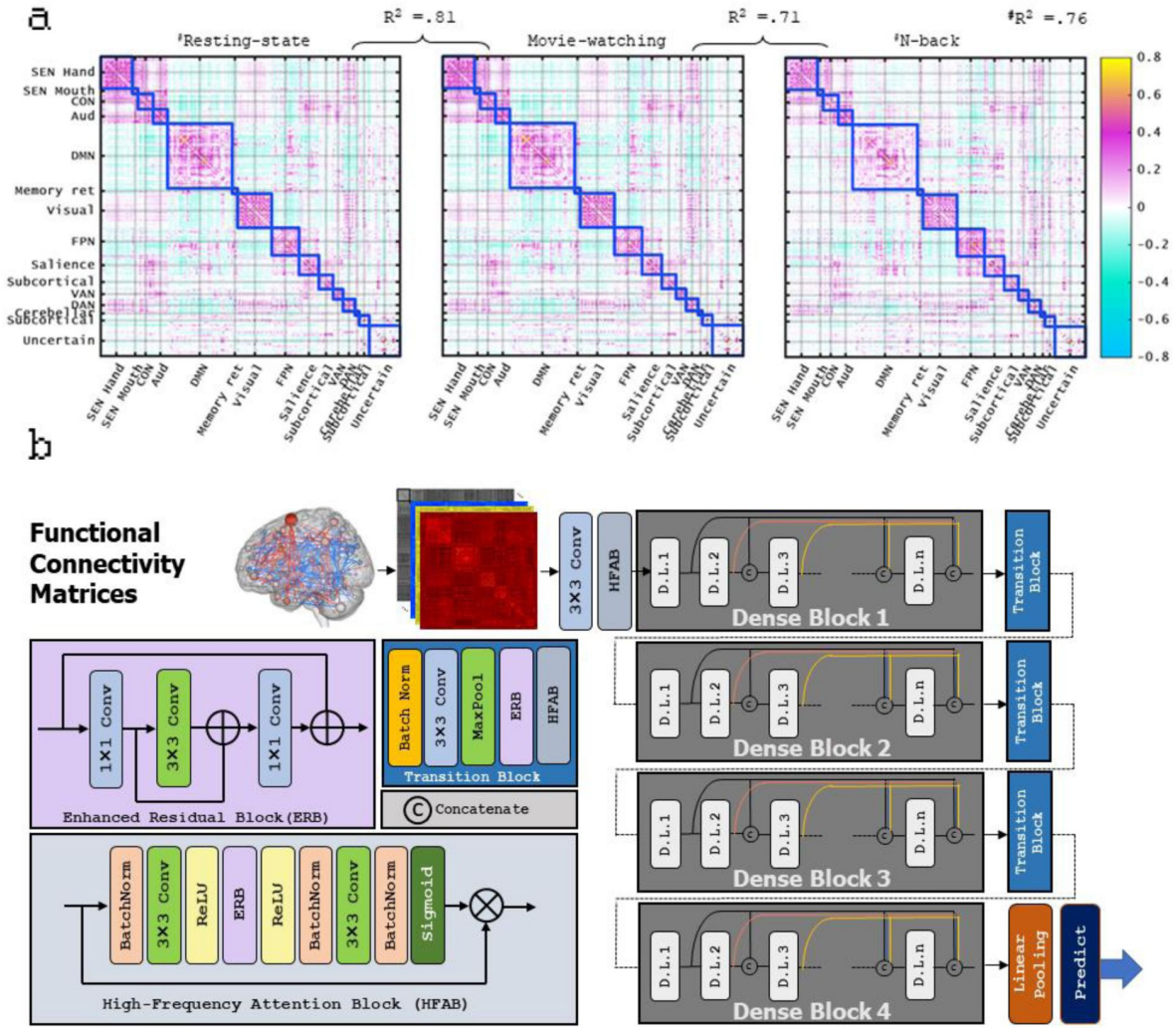


Figure 6.

(a) Example of functional connectivity map across three different cognitive states. SEN hand: SENsory hand; SEN Mouth: SENsory Mouth; CON: Cingulo-Operculum control Network; Aud: Auditory; DMN: Default Mode Network; Memory Ret: Memory Retrieval network; Visual: Visual; FPN: Fronto-Parietal Network; Salience: Salience control network; Subcortical (upper row): subcortical network included in original Power parcellation; VAN: Ventral Attention Network; DAN: Dorsal Attention Network; Cerebellar: Cerebellar network; Subcortical (lower row): additional Subcortical regions, including hippocampus and caudate, added to the original Power Parcellation; Uncertain: Regions with less known network assignment. **(b)** DenselyAttention architecture. Enhanced Residual Block (ERB) and High-Frequency Attention Block (HFAB) into the Transition Block. Note that each "D.L." layer in the table corresponds to the sequence BatchNormalization-ReLU-Conv3×3.

reuse and bypass settings. This results in fewer parameters than similar dense networks such as ResNet (135), enhances feature reuse, improves feature propagation, makes training easier, and reduces issues of gradient vanishing and model degradation.

The architecture of DenseNet is characterized by its dense connectivity pattern, which entails direct connections from each layer to all subsequent layers within its dense block (as illustrated in **Fig. 6b**). This design ensures that every layer has access to the feature maps generated by preceding layers, thereby facilitating a seamless and efficient gradient flow throughout the network. In essence, the knowledge acquired at each layer is propagated forward, enabling the model to effectively capture intricate patterns and dependencies within the data, ultimately enhancing its ability to learn and generalize (87). Additionally, dense blocks provide a regularizing effect, reducing overfitting, particularly on tasks with smaller training datasets (87). This is suitable for this study which includes relatively small samples. Each sequence combines operations of batch normalization (BN), rectified linear unit (ReLU), and a 3×3 convolution (Conv). Batch normalization can effectively prevent overfitting, as described by equation 1:

$$BN(x_n) = \alpha\left(\frac{x_n - \mu_B}{\sqrt{\sigma_\beta^2 + \epsilon}}\right) + \beta, \quad [1]$$

Where μ_β, mean; σ_β, standard deviation; ε, random noise; α and β are adaptable variables in training. We utilized the Rectified Linear Unit (ReLU) activation function (136), which activates neurons by directly outputting the input if it is positive, or zero otherwise, as outlined in equation 2.

$$ReLu(x) = \max(0, x), \quad [2]$$

Where x is the input to a neuron. ReLU benefits the performance of networks with dense layers by decreasing the computation and selectively optimizing parameters.

Four dense blocks facilitate a stepwise down-sampling in the network. These blocks are connected with transition layers, which consist of a 1×1 convolutional layer followed by a 2×2 average pooling layer.

Additionally, ERB and HFAB pairs are introduced for targeted high-frequency features and residual block enhancement. The ERB-HFAB pairs are stacked sequentially at the beginning of each dense block loop (**Fig. 6b**), with ERB and HFAB having 16 feature maps.

High-Frequency Attention Block (HFAB)

Our approach to attentional mechanisms, specifically the HFAB, introduces a sequential attention branch inspired by edge detection. This branch rescales each position based on its neighboring pixels, efficiently focusing on high-frequency areas. The HFAB employs a 3×3 convolution to enhance both the receptive field and computational efficiency. Batch normalization is seamlessly integrated into the attention branch, introducing global statistics during inference without additional computational cost.

Enhanced Residual Block (ERB)

We present ERB as an alternative to the traditional residual block. As illustrated in **Figure 6b**, ERB comprises a re-parameterization block (RepBlock) and a ReLU. In the training stage, the RepBlock utilizes a 1×1 convolution to either expand or contract the number of feature maps, employing a 3×3 convolution to extract features in a higher-dimensional space. Furthermore, two

skip connections are integrated to mitigate training complexities. During inference, all linear transformations can be unified, facilitating the conversion of each RepBlock into a singular 3×3 convolution. Essentially, ERB capitalizes on the advantages of residual learning.

We implemented model training using Tensorflow 2.11.0 (137) and Keras 2.11.0 (138) as programming interfaces and trained on a fifth-generation MacBook Pro (Apple M1 MAX silicon chips, 10-core CPU, 24-core GPU, 16-core Neural Engine, 64 GB memory). For regression tasks, selecting an appropriate loss function is crucial for guiding the optimization process and ensuring accurate predictions. In this study, we opted for the mean squared error (MSE) loss function due to its suitability for regression problems. To minimize the loss function, we trained the network using the stochastic gradient descent (SGD) optimizer with a learning rate of $8e^{-5}$ and a Nesterov momentum of 0.9 (139). The number of epochs was set to 100, and the batch size was set to 74. We also added dropout (140) after each convolutional layer, except the first one, with a rate of 0.15.

We conducted training on distinct, identical datasets extracted from DyNAMiC, each comprising FC maps generated from 120 subjects. To prepare each training dataset, we randomly shuffled data for training and validation patches, allocating 80% for training and 20% for validation. For testing, we utilized all FC maps of 60 subjects from the same sample (the age-heterogeneous DyNAMiC study) enabling us to assess model performance through three-fold cross-validation (**Fig. 5**). Each cross-validation fold was a new training with unseen validation set to the model. Based on its performance on the testing dataset, we selected and employed the winning model for all subsequent analyses. This model, which demonstrated the best performance on the testing dataset, underwent cross-validation in an independent sample from the age-homogeneous COBRA study.

To explore and visually represent the crucial features of the deep learning models contributing to the prediction of cognitive scores, we used the Grad-CAM (Gradient-weighted Class Activation Mapping) technique (141). This method interprets the model's decisions by highlighting the regions of the input image with the most significant impact on the model's output. Grad-CAM conducts a backward pass to compute the gradients of the target class score with respect to the feature maps of the final convolutional layer. We present the average heatmaps calculated for all input data to the model (141).

Positron Emission Tomography (PET)

The scanning sessions started with a 5-minute low-dose helical CT scan (20 mA, 120 kV, 0.8 s per revolution), obtained for attenuation correction. During scanning, a thermoplastic mask was attached to the bed surface to minimize head movement.

In DyNAMiC, a 60-minute scan was performed following 350 MBq (337 ± 27 MBq) in list-mode format. Offline re-binning of list-mode data was conducted to achieve a total of 49 frames with increasing length. In COBRA, a 55-min, 18-frame dynamic PET scan was acquired during rest following intravenous bolus injection of approximately 250 MBq 11C-raclopride (264 ± 19 MBq). For both studies, attenuation- and decay-corrected images (47 slices, field of view = 25 cm, 256×256-pixel transaxial images, voxel size = 0.977× 0.97×3.27 mm^3) were reconstructed with the iterative VUE Point HD-SharpIR algorithm (GE; 6 iterations, 24 subsets, 3.0 mm post filtering; full-width-at-half-maximum (FWHM): 3.2 mm). The estimation of receptor availability, or binding potential relative to non-displaceable binding (BP_{ND}), was carried out following previously described procedures with the cerebellum as a reference region (74). PET images were motion-corrected and co-registered with the structural T1-weighted images from the corresponding session using Statistical Parametric Mapping software (SPM12, Wellcome Department of Imaging Science, Functional Imaging Laboratory, London, UK). Motion-corrected PET data were resliced to match the spatial dimensions of MR data (1 mm^3 isotropic, 256×256×256). The mean of the first five frames was used as a source for co-registration. In DyNAMiC, frame-to-frame head motion correction, with translations ranging from 0.23 to 4.22 mm (mean ± sd = 0.95 ± 0.54 mm), revealed

a trend-level difference across age-groups (age <40 and age ≥ 40 years), as determined by Student's t-test (t =2.0, p =.047; mean ± sd for younger individuals = 1.07 ± 0.52, mean ± sd for older individuals =0.90 ± 0.55). Partial-volume-effect (PVE) correction was achieved using the symmetric geometric transfer matrix (SGTM) method for regional correction, implemented in FreeSurfer (142). An estimated point-spread-function of 2.5 mm FWHM was utilized. Regional estimates of BP_{ND} were calculated within the automated FreeSurfer segmentations employing the simplified reference tissue model (SRTM (143)). In the current study, we focused on the striatal BP_{ND}, calculated as an average of BP_{ND} across the caudate and putamen.

The direct impact of dopamine on brain-cognition gaps through mediation analysis

To evaluate whether functional variability mediates the relationship between D1DR and prediction gap connectivity, we conducted additional mediation analyses. We first computed entropy, which estimates within-region signal variability (84), and then averaged this measure across all regions in the parcellation. Following the mediation analysis framework proposed by Baron and Kenny (144), our goal was to determine whether the association between D1/D2 receptors and brain-cognition gaps is mediated by regional variability (entropy) or if the indirect effect exceeds the direct association between D1/D2 receptors and brain-cognition gaps. To assess the statistical significance of this mediation effect, we employed the bootstrapping method as outlined by Preacher and Hayes (145).

Statistical significance analysis

Statistical analyses were carried out using SPSS (IBM Corp., V24.0.0, Armonk, NY, USA), MATLAB (The MathWorks Inc., V9.13.0 (R2022b), Natick, MA, USA), and GraphPad Prism (GraphPad Software, Inc. V5.01, CA, USA). We performed partial correlations between predicted and actual scores, as well as linear regression analyses. To investigate the relationship between generated gap variables and DA receptor availability, we controlled for age (in DyNAMiC) using partial correlation. The Mann–Whitney U test was used to calculate the mean differences in prediction accuracy. The level of statistical significance was set at $p \leq .05$.

Supplementary Materials

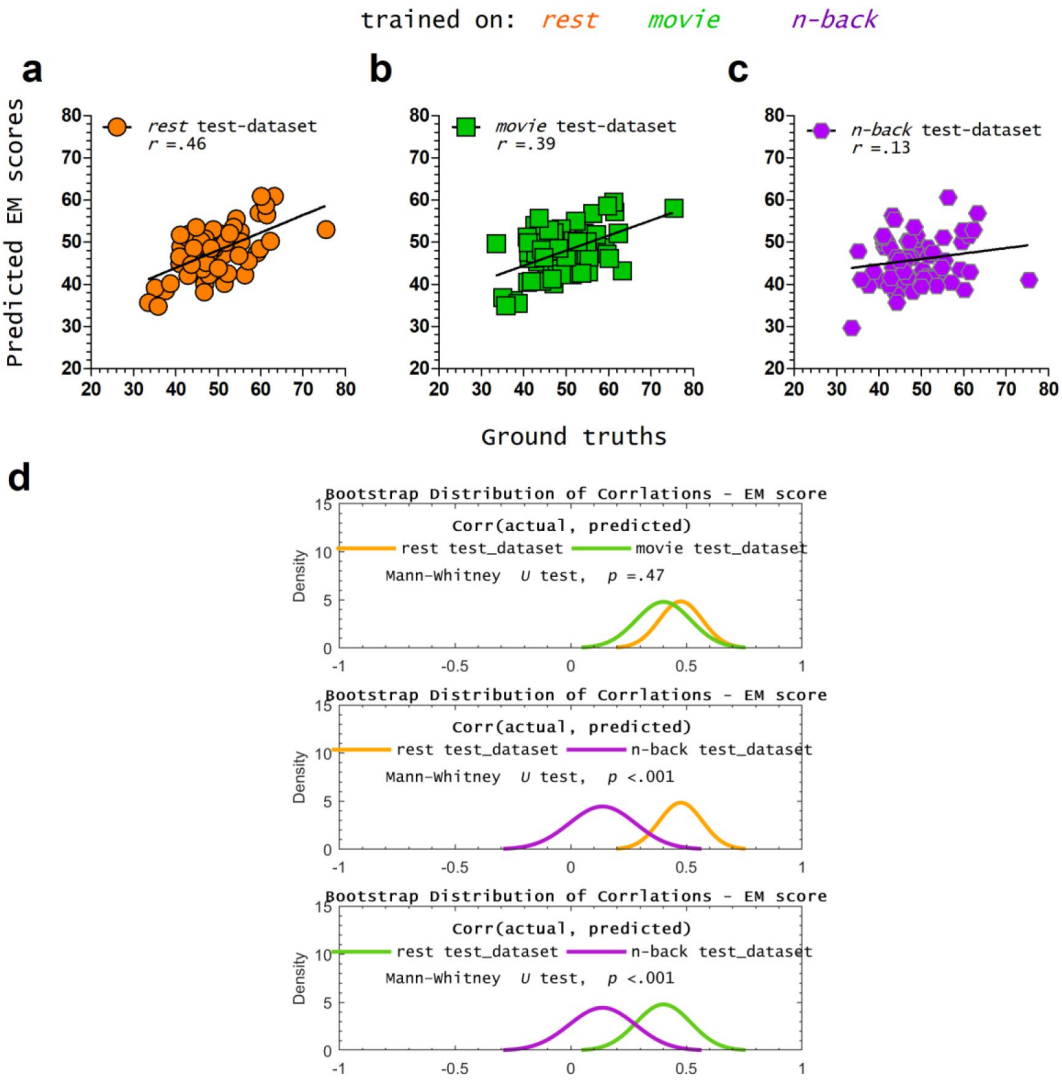


Figure S1.

Model trained on functional connectivity maps acquired at rest predicts (**a**) episodic memory (EM) of the test dataset. The models trained on movie-watching (**b**) but not n-back (**c**) datasets predicted EM scores of the test dataset significantly. **Table S1** summarizes the *p* values and correlation power for each model. Test datasets were obtained from the same cohort for rest, movie-watching, and n-back. Bootstrap distributions of correlations between predicted and actual EM scores indicated no significant difference in the predictive power of EM between models trained on resting-state and movie-watching datasets (**d**). Additionally, the bootstrap distribution revealed that models trained on resting-state and movie-watching data yielded higher correlations than those trained on n-back data (**d**).

Model trained on		rest			movie			n-back	
Model tested on	rest	movie	n-back	rest	movie	n-back	rest	movie	n-back
Correlation	**.46**	**.38**	**.36**	**.34**	**.39**	.09	.06	.15	.14
Significance	**.0002**	**.004**	**.005**	**.007**	**.002**	.47	.67	.25	.28
COBRA									
Correlation	**N.A.**								
Significance	**N.A.**								

Table S1.

Correlation results for episodic memory (EM) score predictions – DyNAMiC dataset with Schaefer300 parcellation.

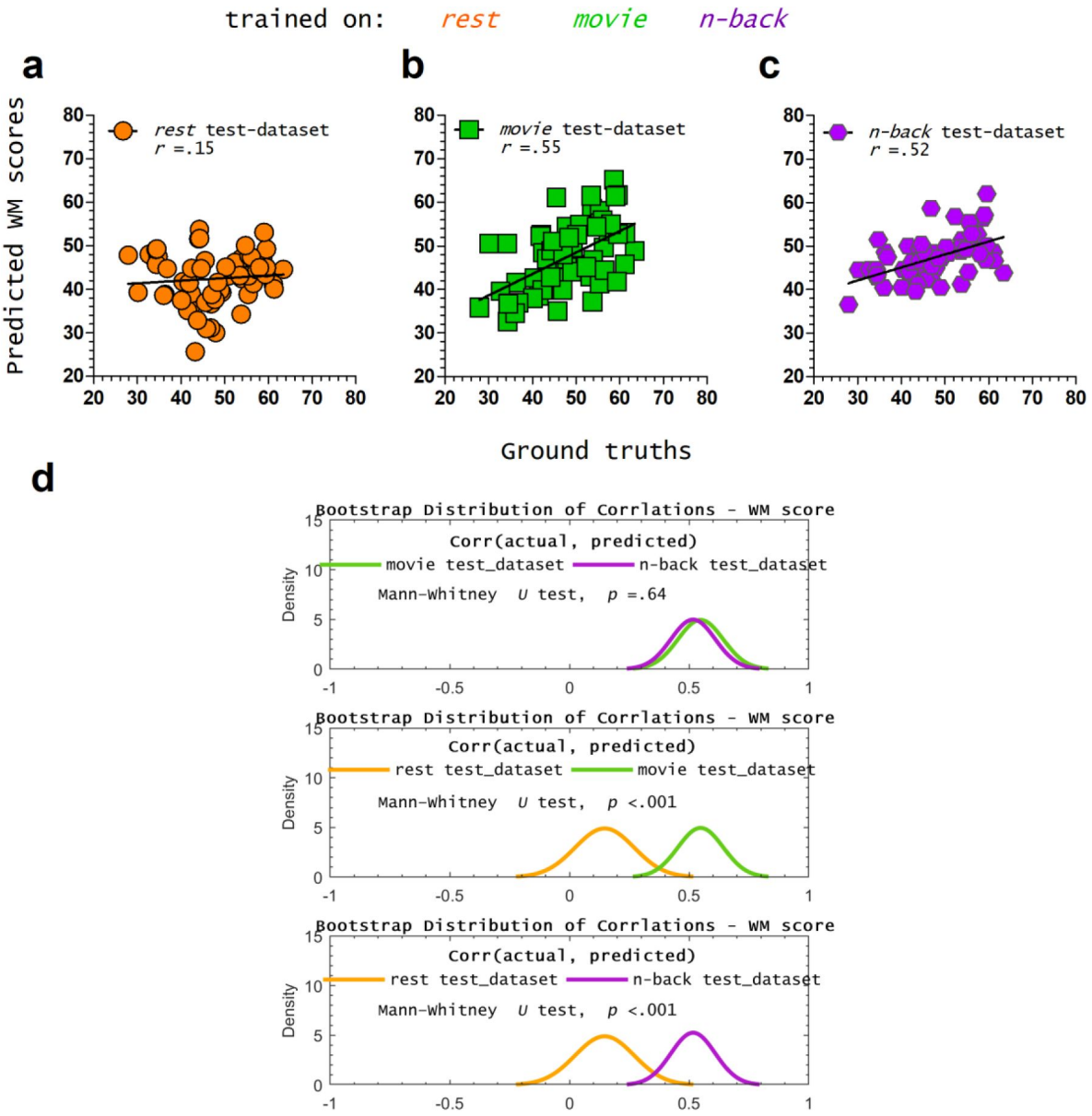


Figure S2.

Model trained on functional connectivity maps acquired at rest did not significantly predict (**a**) the working memory (WM) of the test dataset. The corresponding model best performed in WM scores predictions while trained on the movie-watching data set (**b**). The model trained on the n-back dataset (**c**) was the second-best, predicting WM scores. **Table S2** summarizes the *p* values and correlation power for each model. Test datasets for resting-state, movie-watching, and n-back tasks were derived from the same cohort. Bootstrap distributions of correlations between predicted and actual WM scores showed no significant difference in predictive power between models trained on movie-watching and n-back data (**d**). Furthermore, the bootstrap distribution revealed that models trained on movie-watching and n-back data exhibited higher correlations than those trained on the resting state dataset (**d**).

Model trained on		rest			movie			n-back	
Model tested on	rest	movie	n-back	rest	movie	n-back	rest	movie	n-back
Correlation	.16	.12	.10	**.33**	**.55**	**.43**	**.38**	**.46**	**.52**
Significance	.27	.36	.43	**.01**	**<.0001**	**.006**	**.003**	**.0002**	**<.0001**
COBRA									
Correlation						**N.A.**			
Significance						**N.A.**			

Table S2.

Correlation results for working memory (WM) score predictions – DyNAMiC dataset with Schaefer300 parcellation.

Data availability

The scripts used for developing the model are available at *https://github.com/MorEsm/AI-based-Prediction-of-Cognitive-Function*. (We are working on this channel to provide open-source scripts).

Acknowledgements

This work was funded by the Swedish Research Council (grant number 2016-01936 to A.S.), Knut and Alice Wallenberg Foundation (Wallenberg Fellow grant to A.S.), bank of Sweden (RJ, grant number P20-0515 to A.S.), StratNeuro grant at Karolinska Institute (A.S.). Morteza Esmaeili and Erin B. Bjørkeli were supported by the Southern Eastern Norway Regional Health Authority (Helse Sør-Øst RHF, HSØ, ME: grant number HSØ 2018047; E.B.B. HSØ 2021023).

Additional information

Author Contributions

M.E., E.B.B., and A.S. (and COBRA PIs) study design, method development, data measurement, and analysis. All authors participated in manuscript writing or editing.

References

1. Biswal B., Yetkin F. Z., Haughton V. M., Hyde J. S. 1995) **Functional connectivity in the motor cortex of resting human brain using echo-planar MRI** *Magn Reson Med* **34**:537–541

2. Buckner R. L., Krienen F. M. 2013) **The evolution of distributed association networks in the human brain** *Trends Cogn Sci* **17**:648–665

3. van den Heuvel M. P., Hulshoff Pol H. E. 2010) **Exploring the brain network: a review on resting-state fMRI functional connectivity** *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* **20**:519–534

4. Ferreira L. K., Busatto G. F. 2013) **Resting-state functional connectivity in normal brain aging** *Neurosci Biobehav Rev* **37**:384–400

5. Damoiseaux J. S. 2017) **Effects of aging on functional and structural brain connectivity** *Neuroimage* **160**:32–40

6. Fox M. D., Greicius M. 2010) **Clinical applications of resting state functional connectivity** *Front Syst Neurosci* **4**

7. Andrews-Hanna J. R., Snyder A. Z., Vincent J. L., Lustig C., Head D., Raichle M. E., Buckner R. L. 2007) **Disruption of large-scale brain systems in advanced aging** *Neuron* **56**:924–935

8. Salami A., Wahlin A., Kaboodvand N., Lundquist A., Nyberg L. 2016) **Longitudinal Evidence for Dissociation of Anterior and Posterior MTL Resting-State Connectivity in Aging: Links to Perfusion and Memory** *Cereb Cortex* **26**:3953–3963

9. Kaboodvand N., Backman L., Nyberg L., Salami A. 2018) **The retrosplenial cortex: A memory gateway between the cortical default mode network and the medial temporal lobe** *Hum Brain Mapp* **39**:2020–2034

10. Seeley W. W., Menon V., Schatzberg A. F., Keller J., Glover G. H., Kenna H., Reiss A. L., Greicius M. D. 2007) **Dissociable intrinsic connectivity networks for salience processing and executive control** *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**:2349–2356

11. Avelar-Pereira B., Backman L., Wahlin A., Nyberg L., Salami A. 2017) **Age-Related Differences in Dynamic Interactions Among Default Mode, Frontoparietal Control, and Dorsal Attention Networks during Resting-State and Interference Resolution** *Front Aging Neurosci* **9**

12. Avelar-Pereira B., Backman L., Wahlin A., Nyberg L., Salami A. 2020) **Increased functional homotopy of the prefrontal cortex is associated with corpus callosum degeneration and working memory decline** *Neurobiol Aging* **96**:68–78

13. Machner B., Braun L., Imholz J., Koch P. J., Munte T. F., Helmchen C., Sprenger A. 2021) **Resting-State Functional Connectivity in the Dorsal Attention Network Relates to Behavioral Performance in Spatial Attention Tasks and May Show Task-Related Adaptation** *Front Hum Neurosci* **15**

14. Schimmelpfennig J., Topczewski J., Zajkowski W., Jankowiak-Siuda K. 2023) **The role of the salience network in cognitive and affective deficits** *Front Hum Neurosci* **17**

15. Ferreira L. K., Regina A. C., Kovacevic N., Martin Mda G., Santos P. P., Carneiro Cde G., Kerr D. S., Amaro E., McIntosh A. R., Busatto G. F. 2016) **Aging Effects on Whole-Brain Functional Connectivity in Adults Free of Cognitive and Psychiatric Disorders** *Cereb Cortex* **26**:3851–3865

16. Smith S. M., Elliott L. T., Alfaro-Almagro F., McCarthy P., Nichols T. E., Douaud G., Miller K. L. 2020) **Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations** *Elife* **9**

17. 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** *Nat Protoc* **12**:506–518

18. Vul E., Harris C., Winkielman P., Pashler H. 2009) **Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition** *Perspect Psychol Sci* **4**:274–290

19. Bzdok D., Eickenberg M., Varoquaux G., Thirion B. 2017) **Hierarchical Region-Network Sparsity for High-Dimensional Inference in Brain Imaging** *Inf Process Med Imaging* **10265**:323–335

20. Bzdok D., Meyer-Lindenberg A. 2018) **Machine Learning for Precision Psychiatry: Opportunities and Challenges** *Biol Psychiatry Cogn Neurosci Neuroimaging* **3**:223–230

21. Bzdok D., Krzywinski M., Altman N. 2018) **Machine learning: supervised methods** *Nat Methods* **15**:5–6

22. Sui J., Jiang R., Bustillo J., Calhoun V. 2020) **Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises** *Biol Psychiatry* **88**:818–828

23. Vieira S., Liang X., Guiomar R., Mechelli A. 2022) **Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies** *Clin Psychol Rev* **97**

24. Plis S. M., Hjelm D. R., Salakhutdinov R., Allen E. A., Bockholt H. J., Long J. D., Johnson H. J., Paulsen J. S., Turner J. A., Calhoun V. D. 2014) **Deep learning for neuroimaging: a validation study** *Front Neurosci* **8**

25. van der Burgh H. K., Schmidt R., Westeneng H. J., de Reus M. A., van den Berg L. H., van den Heuvel M. P. 2017) **Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis** *Neuroimage Clin* **13**:361–369

26. Nguyen H., Nguyen V., Nguyen T., Larsen M. E., O'Dea B., Nguyen D. T., Le T., Phung D., Venkatesh S., Christensen H. 2018) **Jointly predicting affective and mental health scores using deep neural networks of visual cues on the web** *Web Information Systems Engineering–WISE 2018: 19th International Conference, Proceedings, Part II* :100–110

27. Li H., Jiang G., Zhang J., Wang R., Wang Z., Zheng W. S., Menze B. 2018) **Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images** *Neuroimage* **183**:650–665

28. Choi Y., Kwon Y., Lee H., Kim B. J., Paik M. C., Won J.-H., Crimi A., Menze B., Maier O., Reyes M., Winzeck S., Handels H. 2016) **Ensemble of Deep Convolutional Neural Networks for Prognosis of Ischemic Stroke** *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* Cham: Springer International Publishing :231–243

29. He T., Kong R., Holmes A. J., Nguyen M., Sabuncu M. R., Eickhoff S. B., Bzdok D., Feng J., Yeo B. T. T. 2020) **Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics** *Neuroimage* **206**

30. Kong J., Wolcott E., Wang Z., Jorgenson K., Harvey W. F., Tao J., Rones R., Wang C. 2019) **Altered resting state functional connectivity of the cognitive control network in fibromyalgia and the modulation effect of mind-body intervention** *Brain Imaging Behav* **13**:482–492

31. Sripada C., Angstadt M., Taxali A., Clark D. A., Greathouse T., Rutherford S., Dickens J. R., Shedden K., Gard A. M., Hyde L. W., Weigard A., Heitzeg M. 2021) **Brain-wide functional connectivity patterns support general cognitive ability and mediate effects of socioeconomic status in youth** *Transl Psychiatry* **11**

32. Chen J., Tam A., Kebets V., Orban C., Ooi L. Q. R., Asplund C. L., Marek S., Dosenbach N. U. F., Eickhoff S. B., Bzdok D., Holmes A. J., Yeo B. T. T. 2022) **Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study** *Nat Commun* **13**

33. Omidvarnia A., Sasse L., Larabi D. I., Raimondo F., Hoffstaedter F., Kasper J., Dukart J., Petersen M., Cheng B., Thomalla G., Eickhoff S. B., Patil K. R. 2023) **Is resting state fMRI better than individual characteristics at predicting cognition?** *bioRxiv*

34. Jiang L., Qiao K., Li C. 2021) **Distance-based functional criticality in the human brain: intelligence and emotional intelligence** *BMC Bioinformatics* **22**

35. Jangraw D. C., Gonzalez-Castillo J., Handwerker D. A., Ghane M., Rosenberg M. D., Panwar P., Bandettini P. A. 2018) **A functional connectivity-based neuromarker of sustained attention generalizes to predict recall in a reading task** *Neuroimage* **166**:99–109

36. Chen Q., Beaty R. E., Cui Z., Sun J., He H., Zhuang K., Ren Z., Liu G., Qiu J. 2019) **Brain hemispheric involvement in visuospatial and verbal divergent thinking** *Neuroimage* **202**

37. Rosenberg M. D., Finn E. S., Scheinost D., Papademetris X., Shen X., Constable R. T., Chun M. M. 2016) **A neuromarker of sustained attention from whole-brain functional connectivity** *Nat Neurosci* **19**:165–171

38. Beaty R. E., Kenett Y. N., Christensen A. P., Rosenberg M. D., Benedek M., Chen Q., Fink A., Qiu J., Kwapil T. R., Kane M. J., Silvia P. J. 2018) **Robust prediction of individual creative ability from brain functional connectivity** *Proc Natl Acad Sci U S A* **115**:1087–1092

39. Hsu W. T., Rosenberg M. D., Scheinost D., Constable R. T., Chun M. M. 2018) **Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals** *Soc Cogn Affect Neurosci* **13**:224–232

40. Gabrieli J. D. E., Ghosh S. S., Whitfield-Gabrieli S. 2015) **Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience** *Neuron* **85**:11–26

41. Finn E. S., Todd Constable R. 2016) **Individual variation in functional brain connectivity: implications for personalized approaches to psychiatric disease** *Dialogues Clin Neurosci*

18:277–287

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

43. 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** *Nat Neurosci* **18**:1664–1671

44. 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

45. Grady C. L., Ryan J. D., Hannula D. E., Duff M. C. 2017) **Age-Related Differences in the Human Hippocampus: Behavioral, Structural and Functional Measures** *The Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition* Cham: Springer International Publishing :167–208

46. Campbell K. L., Schacter D. L. 2017) **Aging and the Resting State: Cognition is not Obsolete** *Lang Cogn Neurosci* **32**:692–694

47. Geerligs L., Tsvetanov K. A. 2017) **The use of resting state data in an integrative approach to studying neurocognitive ageing - Commentary on Campbell and Schacter (2016)** *Lang Cogn Neurosci* **32**:684–691

48. Tavor I., Parker Jones O., Mars R. B., Smith S. M., Behrens T. E., Jbabdi S. 2016) **Task-free MRI predicts individual differences in brain activity during task performance** *Science* **352**:216–220

49. Greene A. S., Gao S., Scheinost D., Constable R. T. 2018) **Task-induced brain state manipulation improves prediction of individual traits** *Nat Commun* **9**

50. Finn E. S., Bandettini P. A. 2021) **Movie-watching outperforms rest for functional connectivity-based prediction of behavior** *Neuroimage* **235**

51. Finn E. S., Glerean E., Hasson U., Vanderwal T. 2022) **Naturalistic imaging: The use of ecologically valid conditions to study brain function** *Neuroimage* **247**

52. Hasson U., Nir Y., Levy I., Fuhrmann G., Malach R. 2004) **Intersubject synchronization of cortical activity during natural vision** *Science* **303**:1634–1640

53. Geerligs L., Rubinov M., Cam C., Henson R. N. 2015) **State and Trait Components of Functional Connectivity: Individual Differences Vary with Mental State** *The Journal of neuroscience : the official journal of the Society for Neuroscience* **35**:13949–13961

54. Nordin K., Gorbach T., Pedersen R., Panes Lundmark V., Johansson J., Andersson M., McNulty C., Riklund K., Wahlin A., Papenberg G., Kalpouzos G., Backman L., Salami A. 2022) **DyNAMiC: A prospective longitudinal study of dopamine and brain connectomes: A new window into cognitive aging** *J Neurosci Res* **100**:1296–1320

55. Cattell R. B. 1971) **Abilities: Their structure, growth, and action** Publisher Houghton Mifflin Company

56. Kyllonen P. C., Dennis I., Tapsfield P. 2013) **Is working memory capacity Spearman's g?** *Human abilities* Psychology Press :49–75

57. Unsworth N., Fukuda K., Awh E., Vogel E. K. 2014) **Working memory and fluid intelligence: capacity, attention control, and secondary memory retrieval** *Cogn Psychol* **71**:1–26

58. Tulving E., Schacter D. L. 1990) **Priming and human memory systems** *Science* **247**:301–306

59. Smallwood J., Schooler J. W. 2015) **The science of mind wandering: empirically navigating the stream of consciousness** *Annu Rev Psychol* **66**:487–518

60. Dosenbach N. U., Nardos B., Cohen A. L., Fair D. A., Power J. D., Church J. A., Nelson S. M., Wig G. S., Vogel A. C., Lessov-Schlaggar C. N., Barnes K. A., Dubis J. W., Feczko E., Coalson R. S., Pruett J. R., Barch D. M., Petersen S. E., Schlaggar B. L. 2010) **Prediction of individual brain maturity using fMRI** *Science* **329**:1358–1361

61. Cole J. H., Franke K. 2017) **Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers** *Trends Neurosci* **40**:681–690

62. Dunas T., Wahlin A., Nyberg L., Boraxbekk C. J. 2021) **Multimodal Image Analysis of Apparent Brain Age Identifies Physical Fitness as Predictor of Brain Maintenance** *Cereb Cortex* **31**:3393–3407

63. Farnsworth von Cederwald B., Josefsson M., Wahlin A., Nyberg L., Karalija N. 2022) **Association of Cardiovascular Risk Trajectory With Cognitive Decline and Incident Dementia** *Neurology* **98**:e2013–e2022

64. Jonasson L. S., Nyberg L., Kramer A. F., Lundquist A., Riklund K., Boraxbekk C. J. 2016) **Aerobic Exercise Intervention, Cognitive Performance, and Brain Structure: Results from the Physical Influences on Brain in Aging (PHIBRA) Study** *Front Aging Neurosci* **8**

65. Tetereva A., Pat N. 2024) **Brain age has limited utility as a biomarker for capturing fluid cognition in older individuals** *eLife* **12**

66. Nyberg L., Lovden M., Riklund K., Lindenberger U., Backman L. 2012) **Memory aging and brain maintenance** *Trends Cogn Sci* **16**:292–305

67. Karalija N., Wahlin A., Ek J., Rieckmann A., Papenberg G., Salami A., Brandmaier A. M., Kohncke Y., Johansson J., Andersson M., Axelsson J., Oradd G., Riklund K., Lovden M., Lindenberger U., Backman L., Nyberg L. 2019) **Cardiovascular factors are related to dopamine integrity and cognition in aging** *Ann Clin Transl Neurol* **6**:2291–2303

68. Backman L., Lindenberger U., Li S. C., Nyberg L. 2010) **Linking cognitive aging to alterations in dopamine neurotransmitter functioning: recent data and future avenues** *Neurosci Biobehav Rev* **34**:670–677

69. Bäckman L., Nyberg L., Soveri A., Johansson J., Andersson M., Dahlin E., Neely A. S., Virta J., Laine M., Rinne J. O. 2011) **Effects of working-memory training on striatal dopamine release** *Science* **333**

70. Volkow N. D., Gur R. C., Wang G. J., Fowler J. S., Moberg P. J., Ding Y. S., Hitzemann R., Smith G., Logan J. 1998) **Association between decline in brain dopamine activity with age and cognitive and motor impairment in healthy individuals** *Am J Psychiatry* **155**:344–349

71. Juarez E. J., Samanez-Larkin G. R. 2019) **Exercise, Dopamine, and Cognition in Older Age** *Trends Cogn Sci* **23**:986–988

72. de Boer L., Axelsson J., Riklund K., Nyberg L., Dayan P., Backman L., Guitart-Masip M. 2017) **Attenuation of dopamine-modulated prefrontal value signals underlies probabilistic reward learning deficits in old age** *Elife* **6**

73. Nyberg L., Karalija N., Salami A., Andersson M., Wahlin A., Kaboovand N., Kohncke Y., Axelsson J., Rieckmann A., Papenberg G., Garrett D. D., Riklund K., Lovden M., Lindenberger U., Backman L. 2016) **Dopamine D2 receptor availability is linked to hippocampal-caudate functional connectivity and episodic memory** *Proc Natl Acad Sci U S A* **113**:7918–7923

74. Nordin K., Nyberg L., Andersson M., Karalija N., Riklund K., Backman L., Salami A. 2021) **Distinct and Common Large-Scale Networks of the Hippocampal Long Axis in Older Age: Links to Episodic Memory and Dopamine D2 Receptor Availability** *Cereb Cortex* **31**:3435–3450

75. Papenberg G., Karalija N., Salami A., Rieckmann A., Andersson M., Axelsson J., Riklund K., Lindenberger U., Lovden M., Nyberg L., Backman L. 2020) **Balance between Transmitter Availability and Dopamine D2 Receptors in Prefrontal Cortex Influences Memory Functioning** *Cereb Cortex* **30**:989–1000

76. Cools R., D'Esposito M. 2011) **Inverted-U-shaped dopamine actions on human working memory and cognitive control** *Biol Psychiatry* **69**:e113–125

77. Zahrt J., Taylor J. R., Mathew R. G., Arnsten A. F. 1997) **Supranormal stimulation of D1 dopamine receptors in the rodent prefrontal cortex impairs spatial working memory performance** *The Journal of neuroscience : the official journal of the Society for Neuroscience* **17**:8528–8535

78. Karalija N., Papenberg G., Johansson J., Wahlin A., Salami A., Andersson M., Axelsson J., Kuznetsov D., Riklund K., Lovden M., Lindenberger U., Backman L., Nyberg L. 2024) **Longitudinal support for the correlative triad among aging, dopamine D2-like receptor loss, and memory decline** *Neurobiol Aging* **136**:125–132

79. Servan-Schreiber D., Printz H., Cohen J. D. 1990) **A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior** *Science* **249**:892–895

80. Seamans J. K., Yang C. R. 2004) **The principal features and mechanisms of dopamine modulation in the prefrontal cortex** *Prog Neurobiol* **74**:1–58

81. Guitart-Masip M., Salami A., Garrett D., Rieckmann A., Lindenberger U., Backman L. 2016) **BOLD Variability is Related to Dopaminergic Neurotransmission and Cognitive Aging** *Cereb Cortex* **26**:2074–2083

82. Li S. C., Rieckmann A. 2014) **Neuromodulation and aging: implications of aging neuronal gain control on cognition** *Curr Opin Neurobiol* **29**:148–158

83. Pedersen R., Johansson J., Salami A. 2023) **Dopamine D1-signaling modulates maintenance of functional network segregation in aging** *Aging Brain* **3**

84. Shafiei G., Zeighami Y., Clark C. A., Coull J. T., Nagano-Saito A., Leyton M., Dagher A., Misic B. 2019) **Dopamine Signaling Modulates the Stability and Integration of Intrinsic Brain Networks** *Cereb Cortex* **29**:397–409

85. Nevalainen N., Riklund K., Andersson M., Axelsson J., Ogren M., Lovden M., Lindenberger U., Backman L., Nyberg L. 2015) **COBRA: A prospective multimodal imaging study of dopamine, brain structure and function, and cognition** *Brain Res* **1612**:83–103

86. 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**:665–678

87. Huang G., Liu Z., van der Maaten L., Weinberger K. Q. 2016) **Densely Connected Convolutional Networks** *arXiv*

88. Zhao W., Makowski C., Hagler D. J., Garavan H. P., Thompson W. K., Greene D. J., Jernigan T. L., Dale A. M. 2023) **Task fMRI paradigms may capture more behaviorally relevant information than resting-state functional connectivity** *Neuroimage* **270**

89. Kohncke Y., Papenberg G., Jonasson L., Karalija N., Wahlin A., Salami A., Andersson M., Axelsson J. E., Nyberg L., Riklund K., Backman L., Lindenberger U., Lovden M. 2018) **Self-rated intensity of habitual physical activities is positively associated with dopamine D(2/3) receptor availability and cognition** *Neuroimage* **181**:605–616

90. Kurkela K., Ritchey M. 2024) **Intrinsic functional connectivity among memory networks does not predict individual differences in narrative recall** *Imaging Neuroscience* **2**:1–17

91. Finn E. S. 2021) **Is it time to put rest to rest?** *Trends in Cognitive Sciences* **25**:1021–1032

92. Zhu Y., Zang F., Wang Q., Zhang Q., Tan C., Zhang S., Hu T., Qi L., Xu S., Ren Q., Xie C. 2021) **Connectome-based model predicts episodic memory performance in individuals with subjective cognitive decline and amnestic mild cognitive impairment** *Behav Brain Res* **411**

93. Wahlheim C. N., Christensen A. P., Reagh Z. M., Cassidy B. S. 2022) **Intrinsic functional connectivity in the default mode network predicts mnemonic discrimination: A connectome-based modeling approach** *Hippocampus* **32**:21–37

94. Gonzalez-Castillo J., Kam J. W. Y., Hoy C. W., Bandettini P. A. 2021) **How to Interpret Resting-State fMRI: Ask Your Participants** *The Journal of neuroscience : the official journal of the Society for Neuroscience* **41**:1130–1141

95. Stawarczyk D., D'Argembeau A. 2015) **Neural correlates of personal goal processing during episodic future thinking and mind-wandering: An ALE meta-analysis** *Hum Brain Mapp* **36**:2928–2947

96. Blonde P., Sperduti M., Makowski D., Piolino P. 2022) **Bored, distracted, and forgetful: The impact of mind wandering and boredom on memory encoding** *Q J Exp Psychol (Hove)* **75**:53–69

97. Karapanagiotidis T., Bernhardt B. C., Jefferies E., Smallwood J. 2017) **Tracking thoughts: Exploring the neural architecture of mental time travel during mind-wandering** *Neuroimage* **147**:272–281

98. Dubois J., Galdi P., Paul L. K., Adolphs R. 2018) **A distributed brain network predicts general intelligence from resting-state human neuroimaging data** *Philos Trans R Soc Lond B Biol Sci* **373**

99. Li J., Biswal B. B., Meng Y., Yang S., Duan X., Cui Q., Chen H., Liao W. 2020) **A neuromarker of individual general fluid intelligence from the white-matter functional connectome** *Transl Psychiatry* **10**

100. Pervaiz U., Vidaurre D., Woolrich M. W., Smith S. M. 2020) **Optimising network modelling methods for fMRI** *Neuroimage* **211**

101. 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 Biol* **17**

102. Salami A., Garrett D. D., Wahlin A., Rieckmann A., Papenberg G., Karalija N., Jonasson L., Andersson M., Axelsson J., Johansson J., Riklund K., Lovden M., Lindenberger U., Backman L., Nyberg L. 2019) **Dopamine D(2/3) Binding Potential Modulates Neural Signatures of Working Memory in a Load-Dependent Fashion** *The Journal of neuroscience : the official journal of the Society for Neuroscience* **39**:537–547

103. Salami A., Rieckmann A., Karalija N., Avelar-Pereira B., Andersson M., Wahlin A., Papenberg G., Garrett D. D., Riklund K., Lovden M., Lindenberger U., Backman L., Nyberg L. 2018) **Neurocognitive Profiles of Older Adults with Working-Memory Dysfunction** *Cereb Cortex* **28**:2525–2539

104. Eriksson J., Vogel E. K., Lansner A., Bergstrom F., Nyberg L. 2015) **Neurocognitive Architecture of Working Memory** *Neuron* **88**:33–46

105. Liang X., Zou Q., He Y., Yang Y. 2016) **Topologically Reorganized Connectivity Architecture of Default-Mode, Executive-Control, and Salience Networks across Working Memory Task Loads** *Cereb Cortex* **26**:1501–1511

106. Gilson M., Deco G., Friston K. J., Hagmann P., Mantini D., Betti V., Romani G. L., Corbetta M. 2018) **Effective connectivity inferred from fMRI transition dynamics during movie viewing points to a balanced reconfiguration of cortical interactions** *Neuroimage* **180**:534–546

107. Betzel R. F., Byrge L., Esfahlani F. Z., Kennedy D. P. 2020) **Temporal fluctuations in the brain's modular architecture during movie-watching** *Neuroimage* **213**

108. Zhang H., Zhao R., Hu X., Guan S., Margulies D. S., Meng C., Biswal B. B. 2022) **Cortical connectivity gradients and local timescales during cognitive states are modulated by cognitive loads** *Brain Struct Funct* **227**:2701–2712

109. 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 Behav* **11**

110. Murphy K. P. 2012) **Machine Learning: A Probabilistic Perspective** MIT Press

111. 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** *Cereb Cortex* **28**:3095–3114

112. Dworetsky A., Seitzman B. A., Adeyemo B., Nielsen A. N., Hatoum A. S., Smith D. M., Nichols T. E., Neta M., Petersen S. E., Gratton C. 2024) **Two common and distinct forms of variation in human functional brain networks** *Nat Neurosci* **27**:1187–1198

113. Tian Y. E., Cropley V., Maier A. B., Lautenschlager N. T., Breakspear M., Zalesky A. 2023) **Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality** *Nat Med* **29**:1221–1231

114. Arida R. M., Teixeira-Machado L. 2020) **The Contribution of Physical Exercise to Brain Resilience** *Front Behav Neurosci* **14**

115. Klil-Drori S., Cinalioglu K., Rej S. 2022) **Brain Health and the Role of Exercise in Maintaining Late-Life Cognitive Reserve: A Narrative Review Providing the Neuroprotective Mechanisms of Exercise** *The American Journal of Geriatric Psychiatry* **30**

116. Nithianantharajah J., Hannan A. J. 2009) **The neurobiology of brain and cognitive reserve: Mental and physical activity as modulators of brain disorders** *Progress in Neurobiology* **89**:369–382

117. Cabeza R., Albert M., Belleville S., Craik F. I. M., Duarte A., Grady C. L., Lindenberger U., Nyberg L., Park D. C., Reuter-Lorenz P. A., Rugg M. D., Steffener J., Rajah M. N. 2018) **Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing** *Nat Rev Neurosci* **19**:701–710

118. Korkki S. M., Johansson J., Nordin K., Pedersen R., Bäckman L., Rieckmann A., Salami A. 2024) **Dedifferentiation of caudate functional organization is linked to reduced D1 dopamine receptor availability and poorer memory function in aging** *bioRxiv*

119. Nyberg L., Andersson M., Lundquist A., Baare W. F. C., Bartres-Faz D., Bertram L., Boraxbekk C. J., Brandmaier A. M., Demnitz N., Drevon C. A., Duezel S., Ebmeier K. P., Ghisletta P., Henson R., Jensen D. E. A., Kievit R. A., Knights E., Kuhn S., Lindenberger U., Plachti A., Pudas S., Roe J. M., Madsen K. S., Sole-Padulles C., Sommerer Y., Suri S., Zsoldos E., Fjell A. M., Walhovd K. B. 2023) **Individual differences in brain aging: heterogeneity in cortico-hippocampal but not caudate atrophy rates** *Cereb Cortex* **33**:5075–5081

120. D'Agostino R. B., Vasan R. S., Pencina M. J., Wolf P. A., Cobain M., Massaro J. M., Kannel W. B. 2008) **General cardiovascular risk profile for use in primary care: the Framingham Heart Study** *Circulation* **117**:743–753

121. Johansson J., Nordin K., Pedersen R., Karalija N., Papenberg G., Andersson M., Korkki S. M., Riklund K., Guitart-Masip M., Rieckmann A., Backman L., Nyberg L., Salami A. 2023) **Biphasic patterns of age-related differences in dopamine D1 receptors across the adult lifespan** *Cell Rep* **42**

122. Hallquist M. N., Hwang K., Luna B. 2013) **The nuisance of nuisance regression: spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity** *Neuroimage* **82**:208–225

123. Friston K. J., Williams S., Howard R., Frackowiak R. S., Turner R. 1996) **Movement-related effects in fMRI time-series** *Magn Reson Med* **35**:346–355

124. Power J. D., Barnes K. A., Snyder A. Z., Schlaggar B. L., Petersen S. E. 2012) **Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion** *Neuroimage* **59**:2142–2154

125. Glover G. H., Li T. Q., Ress D. 2000) **Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR** *Magn Reson Med* **44**:162–167

126. Hutton C., Josephs O., Stadler J., Featherstone E., Reid A., Speck O., Bernarding J., Weiskopf N. 2011) **The impact of physiological noise correction on fMRI at 7 T** *Neuroimage* **57**:101–112

127. Kasper L., Bollmann S., Diaconescu A. O., Hutton C., Heinzle J., Iglesias S., Hauser T. U., Sebold M., Manjaly Z. M., Pruessmann K. P., Stephan K. E. 2017) **The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data** *J Neurosci Methods* **276**:56–72

128. Fair D. A., Schlaggar B. L., Cohen A. L., Miezin F. M., Dosenbach N. U., Wenger K. K., Fox M. D., Snyder A. Z., Raichle M. E., Petersen S. E. 2007) **A method for using blocked and event-related fMRI data to study "resting state" functional connectivity** *Neuroimage* **35**:396–405

129. Cole M. W., Ito T., Schultz D., Mill R., Chen R., Cocuzza C. 2019) **Task activations produce spurious but systematic inflation of task functional connectivity estimates** *Neuroimage* **189**:1–18

130. Ashburner J. 2007) **A fast diffeomorphic image registration algorithm** *Neuroimage* **38**:95–113

131. Fonov V., Evans A. C., Botteron K., Almli C. R., McKinstry R. C., Collins D. L., Brain G. 2011) **Development Cooperative, Unbiased average age-appropriate atlases for pediatric studies** *Neuroimage* **54**:313–327

132. Fonov V. S., Evans A. C., McKinstry R. C., Almli C. R., Collins D. L. 2009) **Unbiased nonlinear average age-appropriate brain templates from birth to adulthood** *NeuroImage* **47**

133. Salami A., Adolfsson R., Andersson M., Blennow K., Lundquist A., Adolfsson A. N., Scholl M., Zetterberg H., Nyberg L. 2022) **Association of APOE varepsilon4 and Plasma p-tau181 with Preclinical Alzheimer's Disease and Longitudinal Change in Hippocampus Function** *J Alzheimers Dis* **85**:1309–1320

134. Janiesch C., Zschech P., Heinrich K. 2021) **Machine learning and deep learning** *Electronic Markets* **31**:685–695

135. He K., Zhang X., Ren S., Sun J. 2015) **Deep Residual Learning for Image Recognition** *arXiv*

136. Nair V., Hinton G. E. 2010) **paper presented at the Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa** *Israel*

137. 2022) **T. Developers. (https://www.tensorflow.org/, 2022).** https://www.tensorflow.org/

138. 2015) **F. Chollet. (https://keras.io, 2015).** https://keras.io

139. Sutskever I., Martens J., Dahl G., Hinton G. 2013) **On the importance of initialization and momentum in deep learning** *Proceedings of the 30th International Conference on Machine Learning, Proceedings of Machine Learning Research* :III-1139–III-1147

140. Srivastava N., Hinton G. E., Krizhevsky A., Sutskever I., Salakhutdinov R. 2014) **Dropout: a simple way to prevent neural networks from overfitting** *Journal of Machine Learning Research* **15**:1929–1958

141. Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. 2020) **Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization** *International Journal of Computer Vision* **128**:336–359

142. Greve D. N., Salat D. H., Bowen S. L., Izquierdo-Garcia D., Schultz A. P., Catana C., Becker J. A., Svarer C., Knudsen G. M., Sperling R. A., Johnson K. A. 2016) **Different partial volume correction methods lead to different conclusions: An (18)F-FDG-PET study of aging** *Neuroimage* **132**:334–343

143. Lammertsma A. A., Hume S. P. 1996) **Simplified reference tissue model for PET receptor studies** *Neuroimage* **4**:153–158

144. Baron R. M., Kenny D. A. 1986) **The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations** *J Pers Soc Psychol* **51**:1173–1182

145. Preacher K. J., Hayes A. F. 2004) **SPSS and SAS procedures for estimating indirect effects in simple mediation models** *Behav Res Methods Instrum Comput* **36**:717–731

Author information

Morteza Esmaeili
Department of Electrical Engineering and Computer Science, University of Stavanger, Stavanger, Norway, Department of Diagnostic Imaging, Akershus University Hospital, Lørenskog, Norway
ORCID iD: 0000-0003-0686-3571

Erin Beate Bjørkeli
Department of Diagnostic Imaging, Akershus University Hospital, Lørenskog, Norway, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Robin Pedersen
Wallenberg Center for Molecular Medicine (WCMM), Umeå University, Umeå, Sweden, Department of Medical and Translational Biology, Umeå University, Umeå, Sweden, Aging Research Center, Karolinska Institute and Stockholm University, Solna, Sweden

Farshad Falahati
Wallenberg Center for Molecular Medicine (WCMM), Umeå University, Umeå, Sweden, Department of Medical and Translational Biology, Umeå University, Umeå, Sweden, Aging Research Center, Karolinska Institute and Stockholm University, Solna, Sweden

Jarkko Johansson
Wallenberg Center for Molecular Medicine (WCMM), Umeå University, Umeå, Sweden, Department of Diagnostics and Intervention, Diagnostic Radiology, Umeå University, Umeå, Sweden
ORCID iD: 0000-0002-4501-4735

Kristin Nordin
Wallenberg Center for Molecular Medicine (WCMM), Umeå University, Umeå, Sweden, Department of Medical and Translational Biology, Umeå University, Umeå, Sweden, Aging Research Center, Karolinska Institute and Stockholm University, Solna, Sweden
ORCID iD: 0000-0003-4157-1638

Nina Karalija
Umeå Center for Functional Brain Imaging (UFBI), Umeå University, Umeå, Sweden, Department of Diagnostics and Intervention, Diagnostic Radiology, Umeå University, Umeå, Sweden

Lars Bäckman
Aging Research Center, Karolinska Institute and Stockholm University, Solna, Sweden

Lars Nyberg
Department of Medical and Translational Biology, Umeå University, Umeå, Sweden, Umeå Center for Functional Brain Imaging (UFBI), Umeå University, Umeå, Sweden, Department of Diagnostics and Intervention, Diagnostic Radiology, Umeå University, Umeå, Sweden
ORCID iD: 0000-0002-3367-1746

Alireza Salami
Wallenberg Center for Molecular Medicine (WCMM), Umeå University, Umeå, Sweden, Department of Medical and Translational Biology, Umeå University, Umeå, Sweden, Aging Research Center, Karolinska Institute and Stockholm University, Solna, Sweden, Umeå Center for Functional Brain Imaging (UFBI), Umeå University, Umeå, Sweden
ORCID iD: 0000-0002-4675-8437

For correspondence: Alireza.Salami@ki.se

Editors

Reviewing Editor
Guido van Wingen

Senior Editor
Andre Marquand
Radboud University Nijmegen, Nijmegen, Netherlands

Reviewer #1 (Public review):

Summary:

The authors attempted to identify whether a new deep-learning model could be applied to both resting and task state fMRI data to predict cognition and dopaminergic signaling. They found that resting state and moving watching conditions best predict episodic memory, but only movie watching predicts both episodic and working memory. A negative 'brain gap' (where the model trained on brain connectivity predicts worse performance than what is actually observed) was associated with less physical activity, poorer cardiovascular function, and lower D1R availability.

Strengths:

The paper should be of broad interest to the journal's readership, with implications for cognitive neuroscience, psychiatry, and psychology fields. The paper is very well-written and clear. The authors use two independent datasets to validate their findings, including two of the largest databases of dopamine receptor availability to link brain functional connectivity/activity with neurochemical signaling.

Weaknesses:

The deep learning findings represent a relatively small extension/enhancement of knowledge in a very crowded field.

It's unclear from these results how much utility the brain gaps provide above and beyond observed performance. It would be helpful to take a median split of the dataset on observed performance and plot aside the current Figure 3 results to see how the cardiovascular and physical activity measures differ based on actual performance. Could the authors perform additional analyses describing how much additional variance is explained in these measures by including brain gaps?

Some of the imaging findings require deeper analysis. For Figure 1f - Which default mode regions have high salience? DMN is a huge network with subregions having differing functions.

Along the same lines, were the striatal D1R findings regionally specific at all? It would be informative to test whether the three nuclei (Accumbens, Caudate, Putamen) and/or voxelwise models would show something above and beyond what is achieved from averaging D1R across the striatum. What about cortical D1R, which is highly abundant, strongly associated with cognitive (especially WM) performance, and has much unique variance beyond striatal D1R? https://www.science.org/doi/full/10.1126/sciadv.1501672. The PET findings are one of the unique strengths of this paper and are underexplored. It's also unclear if the measure of brain entropy should simply be averaged across all regions.

It is not clear from the text that the authors met the preconditions for mediation analysis (that is, demonstrating significant correlations between D1R and entropy, in addition to the correlation with brain gap. The authors should report this as well.

Was age controlled for in the mediation analysis? I would not consider this result valid unless that is the case.

The discussion section is long, but the authors would do better to replace some less helpful sections (e.g., the paragraph on methodological tweaks to parcellations and model alignment) with a couple of other important points, including:

(1) Discuss the 'sweet-spot' of movie watching for behavior prediction in the context of studies showing that task states 'quench' neural variability: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007983. This may not be mutually exclusive of the discussion on dopamine and signal-to-noise ratio, but it would be helpful for the authors to discuss their potential overlap vs. unique contributions to the observed findings.

(2) The argument that dopamine signaling increases signal-to-noise ratio is based on some preclinical data as well as correlational data using fMRI with pharmacological challenges. It is less clear how PET-derived estimates of D1R and D2R availability equate to 'dopamine signaling' as it is thought of in this context. Presumably, based on these data, higher D1R or D2R availability would be related to greater levels of tonic dopaminergic signaling. However, in the case of the COBRA dataset with D2R estimates, those are based on raclopride -- which competes with endogenous dopamine for the D2 receptor. Therefore, someone with higher levels of endogenous dopamine signaling should theoretically have lower raclopride binding and lower D2R estimates. I'm not arguing that the authors' logic is flawed or that D1R and D2R are not good measures of dopamine signaling, but I'd ask the authors to dig into the literature and describe more direct potential links for how greater receptor availability might be associated with greater dopamine signaling (and hence lower entropy). Adding this to the discussion would be very valuable for PET research.

https://doi.org/10.7554/eLife.104053.1.sa2

Reviewer #2 (Public review):

Summary:

The authors developed a deep learning model based on a DenseNet CNN architecture to predict two cognitive functions: working memory and episodic memory, from functional connectivity matrices. These matrices were recorded under three conditions: during rest, a working memory task, and a movie, and were treated as images for the CNN algorithm. They tested their model's performance across different conditions and a separate dataset with a different age distribution (using the same MRI scanner, scanning configurations, and cognitive tests). They also calculated the "brain cognition gap" based on the model trained on resting functional connectivity to predict working memory. Extending from the commonly used index "brain age," the brain cognition gap was defined as the difference between the working memory score predicted by their model (predicted working memory) and the working memory score based on the working memory test itself (observed working memory). This brain cognition gap was found to be associated with physical activity, education, and cardiovascular risk. The authors also conducted additional mediation tests to examine whether regional functional variability mediated the relationship between PET-derived measures of dopamine and the brain cognition gap.

Strengths:

The major strength of this manuscript is the extensive effort the authors have put into creating a new 'biomarker' that links deep learning with fMRI, PET, physical activity, education, and cardiovascular risk across two studies. This effort is impressive.

Weaknesses:

There are several weaknesses in the current methods and results, making many of the claims unconvincing. These weaknesses include:

(1) The lack of baseline models to benchmark the predictive performance of their DenseNet models.

(2) The inappropriate calculation of the brain cognition gap due to the lack of control for regression-toward-the-mean and the influence of the working memory itself (a common practice in brain age studies).

(3) The lack of benchmarking of the brain cognition gap against the 'corrected' brain age gap and the direct prediction of physical activity, education, and cardiovascular risk.

(4) Minimal justification for their PET mediation analysis.

Regarding the impact of the work on the field and the utility of the methods and data to the community, I see its potential. However, addressing all the weaknesses listed above is crucial and likely to change the conclusions of the results.

It is important to note that many statements in the manuscript are overstated, making the contribution of the manuscript seem exaggerated.

For instance, the abstract claims "there is a lack of objective biomarkers to accurately predict cognitive function," and the discussion states, "across various studies, the correlation between predicted and actual fluid intelligence typically hovers around 0.25 (98-100)." However, a meta-analysis by Vieira and colleagues (2022 https://doi.org/10.1016/j.intell.2022.101654) found over 37 studies up to 2020 predicting cognitive abilities from fMRI with machine learning, with 24 studies published in 2019-20 alone. Since 2020, with the rise of machine learning and AI, even more studies have likely been published on this topic, all

claiming to show objective biomarkers to accurately predict cognitive function. Vieira and colleagues also found an average performance of these objective biomarkers in predicting general cognition at r = .42, similar to what was found in this manuscript. Based on this alone, it is unclear how novel or superior their method is without a proper systematic benchmark.

Similarly, the authors claim superior performance of deep learning and mischaracterize machine learning algorithms: "In particular, deep neural networks (DNN) methods have been successfully applied to behavioral and disease prediction (24-26), and have been found to outperform other machine learning approaches (27-29)," and "Deep learning approaches overcome the limitation of predictive techniques that solely rely on linear associations between connectivity and behavioral phenotypes (17)." However, the superiority of deep learning is debatable. Studies show comparable performance between machine learning (such as kernel regression) and deep learning (such as fully-connected neural networks, BrainNetCNN, Graph CNN (GCNN), and temporal CNN), e.g., He and colleagues (2019) and Vieira and colleagues (2024) https://doi.org/10.1016/j.neuroimage.2019.116276 and Vieira and colleagues' https://doi.org/10.1101/2024.03.07.583858.

Moreover, many non-deep learning predictive techniques are non-linear, e.g., XGBoost, CatBoost, random forest, kernel ridge, and support vector regression with non-linear kernels (such as RBF and polynomial). Thus, stating that machine learning can only model linear relationships is incorrect. Moreover, for the small amount of data the authors had, some might argue that a linear algorithm might be more appropriate to balance the bias-variance trade-off in prediction. Again, without a proper systematic benchmark, it is unclear how well their DenseNet algorithm performs compared to other algorithms.

Regarding the Brain Age literature, the authors also misinterpreted recent findings: "However, a recent study suggests that brain age predictions contribute minimally compared to chronological age for explaining cognitive decline (65), implying that cognitive predictions are more reliable." In this study, Tetereva and colleagues (2024) (https://doi.org/10.7554/eLife.87297.4) showed that non-deep-learning machine learning can make good predictions from MRI on both chronological age (with r up to .88) and fluid cognition (with r up to .627). Using the combination of functional connectivity matrices across rest and tasks to predict fluid cognition, they found performance at r = .565, comparable to what was found in the current manuscript with deep learning. Nonetheless, while brain age predicted chronological age well (and brain cognition predicted fluid cognition well), it was problematic to predict fluid cognition from brain age. They showed that, because brain age, by design, shared so much common variance with chronological age, brain age and chronological age captured the same variance of fluid cognition. When chronological age was controlled for in the prediction of fluid cognition, brain age no longer had high predictive ability. In the case of the current manuscript, the brain cognition gap is not appropriately controlled for cognition (to be more precise, a working memory score). I expect the performance in predicting physical activity, education, and cardiovascular risk will drop dramatically once cognition is controlled for. There are at least two ways to control cognition according to Tetereva and colleagues' study (see more in the recommendations).

The authors mentioned, "The third aim of the current study is to uncover the contribution of dopamine (DA) integrity to brain-cognition gaps." However, I fail to see how mediation analysis would test this. The authors also mentioned, "Insufficient DA modulation can affect neurocognitive functions detrimentally (69, 74, 76-78)." They should test if DA levels are related to working memory scores in their study, and if so, whether the relationship is mediated by the "corrected" brain-cognition gaps. Note see more on the recommendation for the calculation of the "corrected" brain-cognition gaps.

https://doi.org/10.7554/eLife.104053.1.sa1

Reviewer #3 (Public review):

Summary:

This paper by Esmaeili and co-authors presents a connectome prediction study to predict episodic memory and relate prediction errors to other phonotypic variables.

Strengths:

(1) A primary and external validation dataset.

(2) Novel use of prediction errors (i.e., brain-cognitive gap).

(3) A wide range of data was investigated.

Weaknesses:

(1) Lack of comparisons to other methods for prediction.

(2) Several different points are being investigated that don't allow any particular one to shine through.

(3) Some choices of analysis are not well-motivated.

(4) How do the n-back connectomes perform for prediction if the authors do not regress task activations from the n-back task?

(5) I am a little concerned about overfitting with the convolutional neural net. For example, the drop-off in prediction performance in the external sample is stark. How does the deep learning approach used here compare to something simpler, like a connectome-based predictive model or ridge regression?

(6) It may be nice to try the other models in the validation dataset. This would also provide a sense of the overfitting that may be going on with overfitting.

(7) While predictive models increase the power over association studies, they still require large samples to prevent overfitting. Do the authors have a sense of the power their main and external validation sample sizes provide?

(8) I am not sure that the Mann-Whitney is the correct test for comparing the distributions of prediction performances. The distributions are dependent on each other as they are each predicting the same outcomes. Using the typical degrees of freedom formula would overestimate the degrees of freedom.

(9) The brain cognition gap is interesting. It is very similar conceptually to the brain age gap. When associating the brain age gap with other phenotypes, typically age is regressed from the brain age gap and the other phenotype. In other words, age is typically associated with a brain age gap as individuals at the tail ages often show the largest gaps. Is the brain cognition gap correlated with episodic memory and do the group differences hold if episodic memory is controlled for?

(10) I have the same question for the dopamine results. Particularly, in the correlations that are divided by brain cognition gap sign. I could see these types of patterns arise due to a correlation with a third variable.

https://doi.org/10.7554/eLife.104053.1.sa0