nature portfolio

Peer Review File

The Network Architecture of General Intelligence in the Human Connectome

Corresponding Author: Professor Aron Barbey

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this human neuroimaging study, the authors investigated associations between general intelligence, g, and brain connectome structures by analysing a relatively large MRI dataset. The theme is important, and the method looks solid. In the meantime, at least in my reading, several critical methodological issues are unclear.

Concern 1
In my understanding, one of the main claims of this study is that the whole-brain model had stronger explanatory power than the independent model. In the meantime, usually models with more variables are likely to have greater power to explain the data. Given this, I'm wondering whether the authors have already considered this issue when they compared the two types of model.

Concern 2
I understood that the authors collected diffusion data from the participants, but am not sure how the authors calculated the path length between regions. In fact, another main claim of this study is about the relationship between general intelligence and weak, long-range connections. In addition, analysis on the small-worldness seems to use this path length. Therefore, the definition of the path length is somewhat critical. The length was calculated based on the diffusion MRI data? Or simply calculated based on the coordinates?

Minor concern
- Fig. 2c in line 473 means Fig. 2a?

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The paper presents a well-designed investigation into how general intelligence relates to the brain's network structure. By combining structural and functional connectivity through a (previously published) joint modeling approach, the authors offer insights into the global organization of the brain and its role in supporting intelligence. The results support key ideas from network neuroscience theory (NNT, previously introduced by one of the authors), showing that general intelligence involves multiple brain networks, depends on weak, long-range connections, engages modal control regions, and relies on small-world topology for efficient communication. It is therefore a valuable step toward moving away from older models focused on local brain regions, and instead highlights the importance of the brain's overall network architecture. While the study makes interesting contributions, its primary contribution appears to be the confirmation of results from NNT via a joint structural-functional network model. We also notice several other weaknesses:

* Network Resolution:
Please see "Issues of parcellation in the calculation of structure-function coupling" by Fotiadis et al. for discussions on the impact of parcellation choice in structure-function coupling analyses. In line with this, recent relevant studies such as Di

Plinio et al. [1], which focus on predictive modeling of intelligence and its components, have addressed this issue by employing multiresolution fMRI network analyses to reduce bias introduced by fixed resolution. Similarly, Kopetzky et al. [2], using structural connectomes, accounted for the impact of parcellation by employing five different parcellations for predictive modeling of intelligence and age. This is not addressed and it therefore is essential for the current study to examine whether its findings hold across multiple spatial resolutions.

* Predictive Model Generalizability on Unseen Data:
Similar recent studies have not only employed multiresolution networks for both functional and structural connectomes in relation to intelligence, but have also validated their models on independent datasets, for example, Di Plinio et al. [1] and Kopetzky et al. [2]. We recommend that the current study extend its analysis by validating the findings on independent datasets such as the Human Connectome Project (HCP) and/or the Alzheimer's Disease Neuroimaging Initiative (ADNI). This would also enable the use of datasets with improved spatial and angular (dMRI) resolutions, which would be better suited for the type of analyses performed in the present study.

* Experiments Relating General Intelligence and Modal Controllability:
Modal controllability is strongly anticorrelated with weighted degree, suggesting that the prediction of general intelligence may also be informed by examining sparsely connected brain regions (see Fig. 2 in Gu et al. [3]). While this may seem redundant given the current study's focus on weak, long-range connections, it may provide a complementary perspective, but that needs to be clarified.

Unlike regions with high average controllability (typically hubs) or modal controllability (typically non-hubs), regions with high boundary controllability do not fall clearly into either category. Boundary controllability does not exhibit a strong correlation or anticorrelation with weighted degree, making it a unique and potentially informative feature. Therefore, we recommend exploring boundary controllability, alongside average controllability, for their potential relationship with general intelligence. We also suggest incorporating this comparison into Figure 5. Additionally, representing the analysis at the nodal level, rather than summarizing at the subnetwork level, may help further localize the brain regions most relevant to general intelligence.

* Connectome-based predictive modeling:
While Leave-One-Out Cross-Validation (LOOCV) is theoretically unbiased, it often exhibits high variance, is computationally inefficient, and can be suboptimal for model selection, even in small-sample settings. In the current context, with 145 samples, LOOCV requires training 145 nearly identical models, each excluding a single data point. This results in high correlation between the training sets and can lead to unstable estimates of model performance, particularly when small perturbations in the data significantly affect predictions (e.g., near decision boundaries). Please see Chapter 5 of James et al. [4] for further details. LOOCV may also be unreliable for tuning generalized linear model parameters, such as selecting relevant features or hyperparameter tuning. Because each fold changes by only one sample, the resulting highly correlated training sets introduce noise into model selection [5].

Recommendation: Consider k-Fold Cross-Validation
In contrast, k-fold cross-validation provides a more robust and computationally efficient approach, especially for GLMs. For reference, in a similar analysis, Popp JL et al. [6] employed 5-fold internal cross-validation and validated their models on independent datasets.

* Additional Evaluation Metrics
To improve model evaluation, we recommend supplementing the performance metrics with Mean Absolute Error (MAE), which is less sensitive to outliers compared to RMSE-based measures. Furthermore, residual plots and calibration curves can offer insights into prediction bias and overall model fit. Because the correlation coefficient (r) and coefficient of determination (R2) are mathematically related in linear models, reporting both may not provide independent information, unless the assumption of linearity is explicitly under investigation. For example, Figures 7 and 8 in Kopetzky SJ et al. [2] present both MAE and range-normalized MAE (NMAE), which can serve as useful benchmarks.

* Minor observations:
Structure-Function Coupling Model: While Chu et al. [7] proposed a global modeling approach, it does not account for regional variability in the structure-function relationship, as highlighted in studies such as Gu et al. [8]. Their method, however, has the potential to be extended using joint connectivity matrix independent component analysis (ICA), as suggested by Wu and Calhoun [9]. It would therefore be helpful to briefly discuss why more recent ICA or CCA-based methods, such as those proposed by Fouladivanda et al. [10] were not considered in their modeling approach. Also, the authors state that “Importantly, the joint structure-function model we adopted offers a distinct advantage in recovering weak, long-range connections compared to traditional tractography methods (26)”. This claim requires further clarification. The cited reference (26) is from 2011, while the joint structure-function modeling applied in this study was published in 2018. It is unclear whether this is a general statement or a claim supported by a specific methodological advancement. A more detailed explanation is needed to clarify this point.

While the work by Popp J. et al. [6] is listed in the references, the manuscript does not discuss how their findings relate to the current study. Given that Popp et al. [6] also investigate structure-function coupling and use HCP data to predict cognitive abilities, we recommend including a comparative discussion to contextualize the similarities or differences in methods, findings, or interpretations.

References:

1. Di Plinio, S. et al. Intrinsic brain mapping of cognitive abilities: A multiple-dataset study on intelligence and its components. NeuroImage 309, 121094 (2025).
2. Kopetzky, S. J., Li, Y., Kaiser, M., Butz-Ostendorf, M. & Initiative, for the A. D. N. Predictability of intelligence and age from structural connectomes. PLOS ONE 19, e0301599 (2024).
3. Gu, S. et al. Controllability of structural brain networks. Nat Commun 6, 8414 (2015).
4. James, G., Witten, D., Hastie, T. & Tibshirani, R. An Introduction to Statistical Learning. vol. 112 (Springer, New York, NY, 2013).
5. Cawley, G. C. & Talbot, N. L. C. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. J. Mach. Learn. Res. 11, 2079–2107 (2010).
6. Popp, J. L. et al. Structural-functional brain network coupling predicts human cognitive ability. NeuroImage 290, 120563 (2024).
7. Chu, S.-H., Parhi, K. K. & Lenglet, C. Function-specific and Enhanced Brain Structural Connectivity Mapping via Joint Modeling of Diffusion and Functional MRI. Sci Rep 8, 4741 (2018).
8. Gu, Z., Jamison, K. W., Sabuncu, M. R. & Kuceyeski, A. Heritability and interindividual variability of regional structure-function coupling. Nat Commun 12, 4894 (2021).
9. Wu, L. & Calhoun, V. Joint connectivity matrix independent component analysis: Auto-linking of structural and functional connectivities. Human Brain Mapping 44, 1533–1547 (2023).
10. Fouladivanda, M. et al. A spatially constrained independent component analysis jointly informed by structural and functional network connectivity. Network Neuroscience 8, 1212–1242 (2024).

Please note: I received input about this manuscript from an early career co-reviewer. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
The manuscript presents a significant and methodologically ambitious attempt to validate the Network Neuroscience Theory (NNT) of intelligence by integrating structural and functional neuroimaging data through joint modeling. The authors leverage connectome-based predictive modeling (CPM) and modern parcellation frameworks to associate topological network properties with general intelligence (g). The effort to move beyond regionally localized models of intelligence is commendable and aligns well with the field's trajectory toward whole-brain network approaches.

While I am generally supportive of this paper, several issues need to be addressed:

Tractography is prone to false-positive connections, which are often marked as weak long-range connections (Maier-Hein et al., Nat. Commun., 2017). I am not aware of any statistical method that fully addresses this issue. Therefore, I suggest visualizing these fibers and attempting to interpret their anatomical plausibility to provide additional validation.

The use of the Schaefer 200-7 parcellation is reasonable; it offers a relatively coarse representation that might reduce the incidence of false-positive tracts. However, it may also average over distinct connectivity patterns, potentially masking meaningful information. To strengthen the validity of the results, I recommend including analyses using a more fine-grained parcellation (e.g., 400 parcels, 17 networks), which could provide complementary insights and help confirm the robustness of the findings.

Concerns regarding CPM's feature selection should be addressed more explicitly. The method used may lead to model instability and potential overfitting, as the authors themselves noted. Have you considered applying regularization techniques (e.g., Ridge regression, Lasso) during the fitting procedure to mitigate these risks?

A note on the generalizability of findings would be helpful. The current cohort size (N = 145) may limit broader applicability. Marek et al. (2022) suggested that correlation-based brain-behavior studies may require thousands of subjects to achieve stable, generalizable results. While acquiring such large-scale datasets—particularly with both imaging and behavioral data—remains challenging, acknowledging this limitation and, where possible, comparing the variance in your cohort to that of

larger datasets could help reassure readers concerned with sample size effects.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I appreciate the authors' responses to my previous concerns. They fully addressed the second concern and minor point. In the meantime, I am afraid to say that, regarding the first concern, they did not. They, in fact, admit the necessity and importance of performing additional analyses to correct the bias induced by the difference in the feature size and put some sentences to clarify this issue. However, they didn't appear to perform actual additional analyses. At least, the response letter didn't indicate such additional results…

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
We thank the authors for their thoughtful clarifications, the revisions made, and the additional analyses performed in response to the reviewers' comments. These efforts have substantially improved the clarity, quality, and overall contribution of the manuscript. After carefully reviewing the authors' responses as well as the changes incorporated into the revised version, we believe that the main concerns previously raised have been adequately addressed.

Note: I co-reviewed this manuscript with an Early Career Researcher. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
No comments

(Remarks on code availability)
-

In the responses below, reviewer comments are presented in **black** text, our replies are provided in **red**, and excerpts from the revised manuscript are shown in **blue**, accompanied by the corresponding page numbers.

Reviewer #1

In this human neuroimaging study, the authors investigated associations between general intelligence, g, and brain connectome structures by analysing a relatively large MRI dataset. The theme is important, and the method looks solid.

We appreciate the reviewer's positive evaluation of our study.

In the meantime, at least in my reading, several critical methodological issues are unclear.

Concern 1

In my understanding, one of the main claims of this study is that the whole-brain model had stronger explanatory power than the independent model. In the meantime, usually models with more variables are likely to have greater power to explain the data. Given this, I'm wondering whether the authors have already considered this issue when they compared the two types of model.

We thank the reviewer for raising this valid concern. One way to address feature set size bias is to implement a grid search across different feature set sizes of randomly selected edges. These are then used to fit predictive models keeping everything exactly the same as the reported cross-validation procedure (please see our prior work, Wilcox & Barbey, 2023; Anderson & Barbey, 2023). One can then see whether predictive performance increases simply as a function of increasing feature set size. We have added a clarification note of this. These prior results have shown that differences in whole brain modeling performance compared to smaller, more localized models, is not dependent on larger feature set sizes. Please see below for relevant, updated language:

"A common concern regarding the comparison of whole-brain models to more localized, ICN specific models is that performance is biased by the feature set size. Prior research has investigated this potential source of bias by performing grid searches across sets of varying sizes comprised of randomly selected edges. These are then used to fit predictive models in the same fashion as the originally described cross-validation procedure. Results from these investigations have consistently shown that the greater predictive performance of whole-brain models compared to more localized models is not simply due to larger feature set sizes.[5,7]" [Page 18; 401-407].

Concern 2

I understood that the authors collected diffusion data from the participants, but am not sure how the authors calculated the path length between regions. In fact, another main claim of this study is about the relationship between general intelligence and weak, long-range connections. In

addition, analysis on the small-worldness seems to use this path length. Therefore, the definition of the path length is somewhat critical. The length was calculated based on the diffusion MRI data? Or simply calculated based on the coordinates?

We thank the reviewer for the opportunity to clarify this point and have worked to address this in the main text and supplementary materials. In this effort, we have recapitulated all analyses in a Human Connectome Project dataset with new methodologies. These new methods enable us to investigate the physical distance of connections, allowing for a more direct investigation of weak ties. As such, the inclusion of average shortest path length in the investigation of weak ties is no longer in the main text. However, as noted by the reviewer, the calculation of average shortest path length is still used in the investigation of small-worldness. So, we have included language to clarify these points. For easier reference, we provide relevant language from the manuscript below that reflects these changes:

"For clarity, average shortest path length was calculated using the topology of the joint structure-function connectomes. This metric reflects the average shortest number of connections that must be traversed between all pairwise nodes. The anatomical coordinates of cortical regions and their geometric distances were not used. In addition, average local clustering was calculated using the Onnela function. This metric reflects the average intensity of triangles around a node.[51] These relationships were examined using a correlation-based approach." [Page 21; Lines 477-483]

Minor concern

- Fig. 2c in line 473 means Fig. 2a?

All figures have been updated in the main text to reflect the new analyses on the Human Connectome Project dataset.

Reviewer #2

The paper presents a well-designed investigation into how general intelligence relates to the brain's network structure. By combining structural and functional connectivity through a (previously published) joint modeling approach, the authors offer insights into the global organization of the brain and its role in supporting intelligence. The results support key ideas from network neuroscience theory (NNT, previously introduced by one of the authors), showing that general intelligence involves multiple brain networks, depends on weak, long-range connections, engages modal control regions, and relies on small-world topology for efficient communication. It is therefore a valuable step toward moving away from older models focused on local brain regions, and instead highlights the importance of the brain's overall network architecture.

While the study makes interesting contributions, its primary contribution appears to be the confirmation of results from NNT via a joint structural-functional network model. We also notice several other weaknesses.

We appreciate the reviewer's complimentary evaluation of our study goals, methodology and findings. We have addressed their remaining comments in detail below.

** Network Resolution:*

Please see "Issues of parcellation in the calculation of structure-function coupling" by Fotiadis et al. for discussions on the impact of parcellation choice in structure-function coupling analyses. In line with this, recent relevant studies such as Di Plinio et al. [1], which focus on predictive modeling of intelligence and its components, have addressed this issue by employing multiresolution fMRI network analyses to reduce bias introduced by fixed resolution. Similarly, Kopetzky et al. [2], using structural connectomes, accounted for the impact of parcellation by employing five different parcellations for predictive modeling of intelligence and age. This is not addressed and it therefore is essential for the current study to examine whether its findings hold across multiple spatial resolutions.

We thank the reviewer for sharing this important point regarding the robustness of results across different parcellation schemes. We addressed the reviewer's concerns by adopting the Glasser parcellation, which boasts 360 cortical regions, as well as the Cole-Anticevic Brain-wide Network Partition, which describes 12 separate networks. These were used in the new analyses carried out with the Human Connectome Project dataset, which is now reported in the main text. Previous analyses that used the Schaefer-200 and Yeo-7 parcellations are now reported in the Supplementary Materials. Comparison across the four schemes suggests that results hold across different parcellations of varying resolutions.

** Predictive Model Generalizability on Unseen Data:*

Similar recent studies have not only employed multiresolution networks for both functional and structural connectomes in relation to intelligence, but have also validated their models on independent datasets, for example, Di Plinio et al. [1] and Kopetzky et al. [2]. We recommend

that the current study extend its analysis by validating the findings on independent datasets such as the Human Connectome Project (HCP) and/or the Alzheimer's Disease Neuroimaging Initiative (ADNI). This would also enable the use of datasets with improved spatial and angular (dMRI) resolutions, which would be better suited for the type of analyses performed in the present study.

We thank the reviewer for highlighting the need to replicate findings, a critical issue in the psychological and brain sciences. To address this, we recapitulated our analyses in the Human Connectome Project dataset. In addition, the Supplementary Materials presents the results of the original analyses carried out in the INSIGHT dataset. The results provide evidence for a conceptual replication, demonstrating that our findings generalize across datasets and are robust to the processing and analysis pipelines, tractography algorithms, behavioral measures, factor analyses, and tests of significance for brain-behavior relationships.

** Experiments Relating General Intelligence and Modal Controllability:*

Modal controllability is strongly anticorrelated with weighted degree, suggesting that the prediction of general intelligence may also be informed by examining sparsely connected brain regions (see Fig. 2 in Gu et al. [3]). While this may seem redundant given the current study's focus on weak, long-range connections, it may provide a complementary perspective, but that needs to be clarified.

We thank the reviewer for raising this thoughtful point regarding the role of weighted degree in establishing modal control mechanisms, and how this might relate to individual differences in general intelligence. The present study was designed to test specific, theoretically motivated predictions of the Network Neuroscience framework. Currently, the theory does not offer predictions regarding the aforementioned brain-behavior relationships. However, we have included language reflecting the reviewer's insight and have highlighted this promising direction for future research:

"Further, while modal control is not strictly a graph theoretic property, at least not in the same way local clustering or characteristic path length are, it does correlate highly with graph theoretic properties, namely weighted degree.[16] This suggests that certain topological properties may be driving factors in the creation and maintenance of structural control mechanisms. Thus, together with our findings, the prediction of general intelligence may also be informed by examining sparsely connected brain regions. Future research investigating this relationship may provide a complementary perspective on the relationship between modal control and *g*." [Page 41; Lines 823-829]

Unlike regions with high average controllability (typically hubs) or modal controllability (typically non-hubs), regions with high boundary controllability do not fall clearly into either category. Boundary controllability does not exhibit a strong correlation or anticorrelation with weighted degree, making it a unique and potentially informative feature. Therefore, we recommend exploring boundary controllability, alongside average controllability, for their potential relationship with general intelligence. We also suggest incorporating this comparison

into Figure 5. Additionally, representing the analysis at the nodal level, rather than summarizing at the subnetwork level, may help further localize the brain regions most relevant to general intelligence.

We thank the reviewer for raising this point regarding the role of structural control mechanisms in intelligence. We agree that examining the relationship between boundary and average controllability could provide valuable insights into the neurobiology of intelligence. The present study was designed to test specific predictions derived from Network Neuroscience Theory, specifically regarding difficult-to-reach states. Currently, NNT does not provide direct predictions for the roles of average and boundary controllability in enabling access to difficult-to-reach states. However, to incorporate the reviewer's important insights, we now address these metrics as a future direction for research in the field:

"In addition, other structural control mechanisms may also provide complimentary insights, specifically average and boundary controllability. Within Network Control Theory, average controllability is hypothesized to enable easy-to-reach states,[16] which may reflect access to general knowledge structures,[4] boundary controllability is hypothesized to enable states reflecting network integration.[16]" [Pages 41-42; Lines 823-834].

We agree with the reviewer that providing modal control information at the nodal level would help to further localize the brain regions most relevant to general intelligence. To address this, we mapped region-wise modal control values across the cortical surface that describe which regions, and to what extent, contribute to the modal control profiles used in brain-behavior analyses for the Human Connectome Project dataset. This works to localize the brain regions most relevant to general intelligence, and, as noted in the main text, coincides with prior work. These results can be found in Fig. 6.

** Connectome-based predictive modeling:*

While Leave-One-Out Cross-Validation (LOOCV) is theoretically unbiased, it often exhibits high variance, is computationally inefficient, and can be suboptimal for model selection, even in small-sample settings. In the current context, with 145 samples, LOOCV requires training 145 nearly identical models, each excluding a single data point. This results in high correlation between the training sets and can lead to unstable estimates of model performance, particularly when small perturbations in the data significantly affect predictions (e.g., near decision boundaries). Please see Chapter 5 of James et al. [4] for further details. LOOCV may also be unreliable for tuning generalized linear model parameters, such as selecting relevant features or hyperparameter tuning. Because each fold changes by only one sample, the resulting highly correlated training sets introduce noise into model selection [5].

Recommendation: Consider k-Fold Cross-Validation

In contrast, k-fold cross-validation provides a more robust and computationally efficient approach, especially for GLMs. For reference, in a similar analysis, Popp JL et al. [6] employed 5-fold internal cross-validation and validated their models on independent datasets.

We thank the reviewer for their insights. As recommended, we have incorporated 5-fold cross-validation in the Human Connectome Project analyses reported in the main text, taking advantage of the dataset's much larger sample. The LOOCV results with the INSIGHT dataset are now reported in the appendix. We observe that our findings are robust across methods.

* *Additional Evaluation Metrics*

To improve model evaluation, we recommend supplementing the performance metrics with Mean Absolute Error (MAE), which is less sensitive to outliers compared to RMSE-based measures. Furthermore, residual plots and calibration curves can offer insights into prediction bias and overall model fit. Because the correlation coefficient (r) and coefficient of determination (R2) are mathematically related in linear models, reporting both may not provide independent information, unless the assumption of linearity is explicitly under investigation. For example, Figures 7 and 8 in Kopetzky SJ et al. [2] present both MAE and range-normalized MAE (NMAE), which can serve as useful benchmarks.

We thank the reviewer for highlighting an important issue in the connectome-based predictive modeling literature. Within cross-validation frameworks, it is possible for correlations between observed and predicted values to be misleading when outliers drive performance. So, it is crucial to adopt additional measures that are robust to outliers, as noted by the reviewer. The coefficient of determination formula used in cross-validation frameworks is one such measure. Importantly, this is not the square of the observed vs. predicted correlation. Rather, it must be calculated using the mean *only from the training set* applied to fit the model. In other words, while it is possible to receive a high observed vs. predicted correlation due to an outlier, the coefficient of determination in such contexts would be very low and can even be negative, rendering the measure robust to outliers. To further address concerns regarding outliers, we have replaced all correlation measures of performance within the lesion analyses with thus-calculated coefficients of determination. For clarity, we have added language to the manuscript that addresses how the coefficients are calculated. This text is copied below for easier reference:

"We used the cross-validation formula to generate the coefficient of determination: $R^2 = 1 - \frac{\sum (y_{obs} - y_{pred})^2}{\sum (y_{obs} - \bar{y}_{obs,train})^2}$. Note that the observed mean comes from the training set and does not include observed values from the test set. Thus, R^2 does not simply equal the correlation squared. This is important to account for, because prediction outliers can artificially inflate the observed vs. predicted correlation, thereby giving the impression that the model performed well. The coefficient of determination formula used here is robust to such outliers and provides a more accurate measure of model performance. For example, R^2 values are typically negative when highly positive performance correlations are due solely to extreme outliers." [Page 17; Lines 375-383].

We appreciate the reviewer's note about the risk of prediction bias and model overfitting when using standard linear regression. To address this, we adopt an elastic net approach instead in our new HCP analyses that incorporates regularization. Combined with the 5-

fold cross-validation procedure that produces less overlap between training sets, this enhances the generalizability of our findings.

** Minor observations:*
Structure-Function Coupling Model: While Chu et al. [7] proposed a global modeling approach, it does not account for regional variability in the structure-function relationship, as highlighted in studies such as Gu et al. [8]. Their method, however, has the potential to be extended using joint connectivity matrix independent component analysis (ICA), as suggested by Wu and Calhoun [9]. It would therefore be helpful to briefly discuss why more recent ICA or CCA-based methods, such as those proposed by Fouladivanda et al. [10] were not considered in their modeling approach. Also, the authors state that "Importantly, the joint structure-function model we adopted offers a distinct advantage in recovering weak, long-range connections compared to traditional tractography methods (26)". This claim requires further clarification. The cited reference (26) is from 2011, while the joint structure-function modeling applied in this study was published in 2018. It is unclear whether this is a general statement or a claim supported by a specific methodological advancement. A more detailed explanation is needed to clarify this point.

We appreciate these suggestions for the refinement of Chu et al's joint structure-function model. We applied recent scripts provided by Chu and colleagues and are unsure why the noted ICA/CCA alternatives were not built into their method. Given our manuscript's focus on improving the prediction of intelligence rather than enhancing the joint modeling approaches, we note the more recent analysis frameworks as possible advancements to the Chu et al. method. We provide the updated text below:

"Other methods seek to improve the estimation of ICNs by developing multi-modal independent component analyses.[62,63] These methods can produce multi-modal, data-driven parcellation schemes, link structural and functional connectivity information, enhance structure-function coupling between ICNs, and improve network distinctions." [Pages 37-38; Lines 738-742]

We thank the reviewer for the opportunity to include the correct citation. As the reviewer notes, the citation was not intended to be (26), which refers to a paper demonstrating issues with recovering weak, long-range connections that traverse through cross-fiber regions, but rather should have been the more recent Chu et al. citation. This has been corrected, thank you.

While the work by Popp J. et al. [6] is listed in the references, the manuscript does not discuss how their findings relate to the current study. Given that Popp et al. [6] also investigate structure-function coupling and use HCP data to predict cognitive abilities, we recommend including a comparative discussion to contextualize the similarities or differences in methods, findings, or interpretations.

We thank the reviewer for bringing this to our attention. We have added language in the manuscript to compare our approach to the one adopted by Popp and colleagues. Structure-function coupling merely describes the coherence between independently-generated structural and functional connectomes. This is typically reported as a correlation

describing how similar a region's connections are between structural and functional connectomes. In contrast, we adopt a method that jointly models the modalities from the ground up to give rise to a single connectome, rather than comparing independently-generated, modality-specific connectomes. The aim is to more accurately model the global topology of the brain, for example, by recovering weak, long-range connections. Another benefit is the ability to use brain function to better inform the information flow bandwidth of structural connections, which can only be hypothesized from tractography algorithms applied to diffusion-weighted data. To summarize, while their approach investigates the relationship between structure-function coupling and cognitive abilities, we investigate the relationship between function-informed structural connections and general intelligence. Please see updated text below:

"Notably, our HCP whole-brain model outperformed a recent study predicting intelligence scores also in held-out HCP participants from resting-state and diffusion-weighted data, but using structure-function coupling methods[60] (12% versus 6%). In addition, our HCP whole-brain model similarly outperformed another coupling model deployed on the HCP data that incorporated task-based functional images rather than solely focusing on resting-state images[61] (12% versus < 1 – 7%). These comparisons encourage allowing structural and functional data to inform each other rather than merely considering their coherence. The joint modeling approach supports a more accurate model of the brain's global topology, for example, through the recovery of weak, long-range connections. Another benefit is a function-informed measure of information flow bandwidth associated with structural connections that can only ever be 'hypothesized' by tractography algorithms applied to diffusion-weighted data. [Page 37; Lines 723-733].

Reviewer #3

We appreciate the reviewer's co-review of the manuscript with reviewer #2 and have included our response above, thank you.

Reviewer #4

The manuscript presents a significant and methodologically ambitious attempt to validate the Network Neuroscience Theory (NNT) of intelligence by integrating structural and functional neuroimaging data through joint modeling. The authors leverage connectome-based predictive modeling (CPM) and modern parcellation frameworks to associate topological network properties with general intelligence (g). The effort to move beyond regionally localized models of intelligence is commendable and aligns well with the field's trajectory toward whole-brain network approaches.

While I am generally supportive of this paper, several issues need to be addressed.

We are grateful for the reviewer's positive evaluation of our study goals, methodology and findings. We have addressed their remaining comments in detail below.

Tractography is prone to false-positive connections, which are often marked as weak long-range connections (Maier-Hein et al., Nat. Commun., 2017). I am not aware of any statistical method that fully addresses this issue. Therefore, I suggest visualizing these fibers and attempting to interpret their anatomical plausibility to provide additional validation.

We thank the reviewer for highlighting this important methodological concern. Given the number of the connections we studied (up to thousands), it was infeasible to include a visual evaluation of each link within the current manuscript. Therefore, we incorporated several alternative refinements to the analysis instead that reduce false positives:

1. **We adopted a new tractography algorithm (multi-tissue, multi-shell spherical deconvolution) that is designed to more accurately model tracts through cross-fiber areas. These areas represent a major source of false-positive connections seen in structural connectomes (Markov et al., 2013).**

2. **Tracts that appear in a single or small subset of participants are likely results of random noise or fluctuations in the estimation of local direction within fiber orientation density functions (fODFs). A skewness threshold removes those connections representing possible false positives. Our cross-validation framework for the HCP dataset includes a skewness threshold.**

3. **The tractography algorithm we incorporated includes strict termination criteria, ensuring the exclusion of tracts with unrealistic angles, lengths (i.e., extremely long inter-hemispheric connections), and pathways (i.e., travelling through ventricles and/or adjacent sulci). This helps further reduce the false positive rate.**

4. **Finally, we implemented the SIFT2 method in our HCP analysis, which filters the tractograms used to generate structural connectomes. This method solves a global optimization problem to find a set of connection weights across all connections that best match the estimated fODFs across all voxels. In this approach, false positives are given weights close to zero, thereby effectively silencing them when generating structural connectomes.**

Considering the combined effects of the discussed steps as well as the robustness of findings across the HCP and INSIGHT datasets, we contend that the findings are unlikely to be the result of false positive connections.

The use of the Schaefer 200-7 parcellation is reasonable; it offers a relatively coarse representation that might reduce the incidence of false-positive tracts. However, it may also average over distinct connectivity patterns, potentially masking meaningful information. To strengthen the validity of the results, I recommend including analyses using a more fine-grained parcellation (e.g., 400 parcels, 17 networks), which could provide complementary insights and help confirm the robustness of the findings.

We thank the reviewer for highlighting the importance of the parcellation scheme. For our Human Connectome Project analyses in the main text, we adopted the Glasser-360 parcellation scheme, as well as the Cole-Anticevic Brain-wide Network Partition scheme. Both parcellations are greater in resolution compared to the Schaefer 200 and Yeo 7 parcellations applied to the INSIGHT data. Nonetheless, the convergent findings across the schemes suggests that the choice of parcellation approaches did not affect the main results.

Concerns regarding CPM's feature selection should be addressed more explicitly. The method used may lead to model instability and potential overfitting, as the authors themselves noted. Have you considered applying regularization techniques (e.g., Ridge regression, Lasso) during the fitting procedure to mitigate these risks?

We thank the reviewer for their insightful suggestion. We have addressed this issue by adopting an elastic net model for the HCP analysis that incorporates regularization.

A note on the generalizability of findings would be helpful. The current cohort size (N = 145) may limit broader applicability. Marek et al. (2022) suggested that correlation-based brain-behavior studies may require thousands of subjects to achieve stable, generalizable results. While acquiring such large-scale datasets—particularly with both imaging and behavioral data—remains challenging, acknowledging this limitation and, where possible, comparing the variance in your cohort to that of larger datasets could help reassure readers concerned with sample size effects.

We thank the reviewer for highlighting the importance of larger sample sizes in connectomic studies. To ensure generalizability, we have recapitulated our analyses in the much larger Human Connectome Project. The robustness of the results across the two datasets and despite broadly different processing and analysis pipelines suggests that the findings are generalizable.

References

Wilcox, R. R. & Barbey, A. K. Connectome-based predictive modeling of fluid intelligence: evidence for a global system of functionally integrated brain networks. *Cereb. Cortex* **33**, 10322–10331 (2023).

Anderson, E. D. & Barbey, A. K. Investigating cognitive neuroscience theories of human intelligence: A connectome-based predictive modeling approach. *Hum. Brain Mapp.* **44**, 1647–1665 (2023).

Markov, N. T. *et al.* Weight Consistency Specifies Regularities of Macaque Cortical Networks. *Cereb. Cortex* **21**, 1254–1272 (2011).

In the responses below, reviewer comments are presented in **black** text, our replies are provided in **red**, and excerpts from the revised manuscript are shown in **blue**, accompanied by the corresponding page numbers.

Reviewer #1

I appreciate the authors' responses to my previous concerns. They fully addressed the second concern and minor point. In the meantime, I am afraid to say that, regarding the first concern, they did not. They, in fact, admit the necessity and importance of performing additional analyses to correct the bias induced by the difference in the feature size and put some sentences to clarify this issue. However, they didn't appear to perform actual additional analyses. At least, the response letter didn't indicate such additional results…

Thank you for the opportunity to address this issue. We have now performed an additional permutation-based analysis to explain whether the larger feature set of the whole-brain model could artificially inflate performance. Specifically, we evaluated model performance across features sets of increasing size populated with randomly permuted connections. The results show that increasing the number of input features does not improve predictive accuracy and can even reduce performance, confirming that feature dimensionality does not contribute to the reported findings. These results, along with an accompanying figure, have been included in the revised manuscript (pages 28-29; lines 579-601) and are presented below.

Feature set size bias

It is critical to verify that the predictive efficacy of the whole-brain model compared to smaller, localized models (i.e., inclusion models) was not an artifact of its larger feature set. To examine this issue, we conducted a permutation-based control analysis within our previously described predictive modeling framework. Specifically, we performed a grid search across feature sets of increasing size, each populated with randomly permuted connections to isolate the effect of feature dimensionality. Our results (Fig. 5) demonstrate that increasing the number of input features does not improve predictive accuracy as measured by a Pearson correlation ($r_{Pearson\ correlation\ vs.\ set\ size}$ = -0.08, $t(159)$ = -1.02, p = 0.309; Fig. 5a) and coefficient of determination ($r_{coefficient\ of\ determination\ vs.\ set\ size}$ = -0.72, $t(159)$ = -13.22, p < 2.2e-16***; Fig. 5b). In fact, we show that larger feature sets can lead to decreases in performance compared to smaller sets (Fig. 5b).

This finding highlights an important methodological point: Pearson correlations between observed and predicted values in a cross-validation framework can provide misleading estimators of model performance. As mentioned previously, it is essential to include less biased metrics, such as the coefficient of determination (R^2), for robust model assessment.

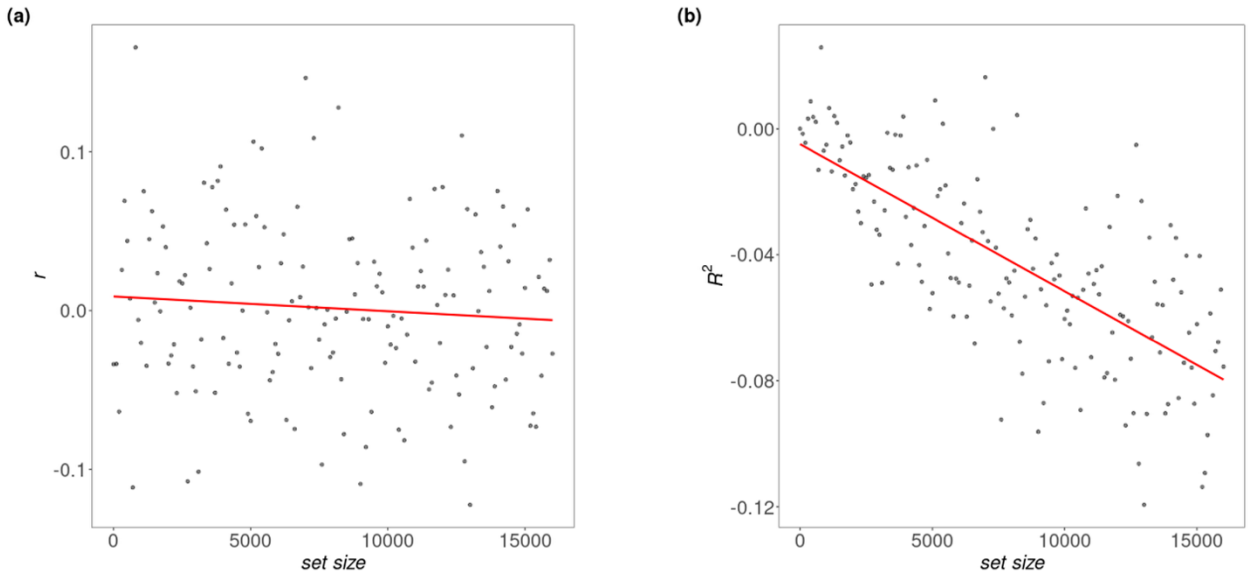

Figure 5. Results from a permutation-based control analysis testing the effect of feature dimensionality on predictive modeling performance. A grid search was conducted across feature sets of increasing size using randomly permuted connections. These feature sets were used to train an elastic net model to predict *g* scores on a held-out test sample using 5-fold cross-validation. Performance, as measured by **(a)** Pearson correlation (r) and **(b)** coefficient of determination (R^2), is plotted against the number of input features.