https://doi.org/10.1027/1864-1105/a000422

Electronic SupplementaryMaterials

1. Supplementary Methods: Details about the original studies (Samples, Data Acquisition, Preprocessing)

This study builds on two public fMRI datasets. First, the main dataset is the *Queensland Adolescent Twin Brains (QTAB) dataset*, which contains fMRI recordings from twins watching a Pixar short film titled Partly Cloudy in Queensland, Australia[1] (Strike et al., 2022). The second fMRI dataset (*Boston dataset*) was also recorded while participants viewed the Partly Cloudy short film, and it was actually the first study using this particular stimulus (Richardson et al., 201X). This dataset was recorded from non-twin individuals, and it included a broader sample in terms of age range (i.e., younger children, adolescents, and adults). Below, we describe details of each dataset. For further details, we refer readers to the respective original populations.

1.1. Queensland Adolescent Twin Brains (QTAB) dataset: Partly Cloudy

1.1.1. Sample

Twins from the area surrounding Brisbane, Queensland, were recruited through twin registries and online postings. There were two study sessions, and the presentation of the Partly Cloudy video happened at the second session. In the following, we focus only on the part of the study that are relevant to the current investigation, leaving aside additional tasks and data (e.g. saliva samples, etc.) that were also collected during the QTAB study.

1.1.2. Procedure and Data Acquisition

[1] The movie can be viewed at: https://www.youtube.com/watch?v=5VRc8poIwU4

In brief, participants watched the Pixar movie Partly Cloudy while in the scanner without further instruction. The stimulus was presented on a back-projection screen and participants viewed via a mirror that was attached to the scanner head coil.

Scanning took place on a 3T Magnetom Prisma (Siemens Medical Solutions, Erlangen) MRI scanner. Anatomical and functional scans were acquired, and full details are given in the data descriptor paper and on OpenNeuro (accession number ds004146), where the data are provided in the BIDS-compatible data format. For the Partly Cloudy movie, a total of 380 volumes were acquired (duration ca. 5 min 11 sec) with a TR/TE=800/30ms (FA = 52 deg, A>P, 60 slices, and a FOV of 216×216mm, slice acceleration=6).

1.1.3. Preprocessing

Results included in this manuscript come from preprocessing performed using fMRIPrep 21.0.2 (Esteban, et al., 2018; Esteban, et al., 2018), which is based on Nipype 1.6.1 (Gorgolewski et al., 2018).

Preprocessing of B0 inhomogeneity mappings. A total of 1 fieldmaps were found available within the input BIDS structure for this particular subject. A B0-nonuniformity map was estimated based on echo-planar imaging (EPI) references (Andersson et al., 2003).

Anatomical data preprocessing. A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset.The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.3.3 (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh

workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 6.0.5.1:57b01774, RRID:SCR_002823, Zhang, Brady, and Smith 2001). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template version 2009c [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing. For each of the Partly-Cloudy-related BOLD runs per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1:57b01774, Jenkinson et al. 2002). BOLD runs were slice-time corrected to 0.351s (0.5 of slice acquisition range 0s-0.703s) using 3dTshift from AFNI (Cox and Hyde 1997, RRID:SCR_005927). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD reference was then co-registered to the T1w reference using mri_coreg (FreeSurfer) followed by flirt (FSL 6.0.5.1:57b01774, Jenkinson and Smith 2001) with the boundary-based registration (Greve and Fischl 2009) cost-function. Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three

region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by thresholding the corresponding partial volume map at 0.05, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of

temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer). Many internal operations of fMRIPrep use Nilearn 0.8.1 (Abraham et al. 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

Functional data extraction. From the preprocessed functional volumes, we then extracted regional time series that measure the viewers' regional brain responses during movie viewing. As described in the main paper and detailed further in the online code repository, we used a brain atlas comprising 293 regions. Using a masker from the nilearn package, we set a temporal high-pass filter to 0.01 Hz, activated the detrending option, and extracted z-scored data time series. This yielded a data a matrix of 293 regions * 380 time points (volumes), which was extracted for each of 200 viewers.

1.2. Boston, USA Dataset (MIT): Partly Cloudy

The sample for the Boston dataset came from a study of child brain development that included a total sample of 155 viewers, who also watched the same movie: Partly Cloudy. The dataset is

publicly available (OpenNeuro #ds000228) and it contains fMRI recordings from adults and children watching the Partly Cloudy short film (Richardson et al. 2018), along with typical descriptor data like participant sex and age, all also in the BIDS format. This dataset, its sample, and processing details have been described in at least two publications (Richardson et al, 2018, blinded_for_review) and thus we limit the presentation here to the main aspects.

The sample was restricted to exclude the children below the age of 12, leaving 33 viewers (20 female). Participants were recruited from the local Bostonian community, and they were scanned on a 3-Tesla Siemens Tim Trio scanner located at the Martinos Imaging Center at MIT.

BOLD-fMRI data were collected with a gradient-echo EPI sequence in 32 interleaved axial slices aligned with the A/P commissure (EPI factor: 64; TR: 2 s, TE: 30 ms, flip angle: 90°). Given the slower TR in this study compared to the QTAB study, it took 168 volumes to acquire brain responses during movie viewing. Of note, the QTAB study provides further details about the specific stimulus version used, including timing details and format/codecs. However, it is worth noting that while the same movie was shown to both audiences (QTAB and Boston), the specific timing differed minimally (regarding disdaq-volumes, movie credits, and the more precise timing in the QTAB study).

Preprocessed functional data , were downloaded from OpenNeuro's BIDS-derivatives folder (slice-time corrected, realigned, and normalized), and our own analyses of these data were then carried out using functions from the nilearn and BrainIAK packages (Abraham et al. 2014; Kumar et al. 2020). As for the QTAB processing pipeline, the fMRI data were high-pass filtered at 0.01 Hz, detrended, and regional brain activity time series were extracted from the 268 regions provided by the Shen-parcellation (Shen et al., 2013) and z-standardized. One additional detail worth pointing out is that the Boston-data were extracted from 268 regions of the Shen parcellation only,

compared to the QTAB dataset for which we extracted data also from the 268-region parcellation, but added a few anatomically defined subcortical regions. Given that ISC is low in these regions anyway, and our interest here is about the cortical response similarity, this does not matter further. Thus, comparisons between the Boston and QTAB data refer only to the 268 regions common to both datasets. Moreover, it is worth noting that the Boston dataset contains 33 viewers, 268 regional timeseries, and 168 data points, the QTAB dataset contains 200 viewers, als 268 regional timeseries (or 293, if subcortical structures were added), but 380 data points (due to the faster TR of 0.8s vs. 2s).

2. Supplementary Methods: Details about ISC analysis

2.1. Randomization Time-Series and Bootstrapping Analysis

In the literature about ISC, there has been some discussion and confusion about methods for statistical significance testing. Additionally, there are different levels of granularity at which ISC results can be assessed. For instance, ISC can be computed on a pair-to-pair level (as done here) or between individuals and the average among the remainder of the group without this participant. Also, ISC can be computed over time-series extracted from regions (as done here), or it can be computed at the level of voxels. Finally, several papers have examined the statistical properties of ISC, and how they depend on the specific choice within this space (voxels vs. regions, pairs vs. groups, etc). Importantly, it has been suggested that ISC statistics computed across pairs of participants (and typically at the voxel level) suffer from a tangled structure, which can bias parametric tests. Several strategies have been proposed and we thus carry out additional analyses complementing the results in the paper to demonstrate that the results and conclusions are the same. In particular, the main results are computed using a bootstrap statistic in which the ISC

among pairwise viewers is resampled and used as a reference distribution to compute significance. Another strategy is to resample phase-shifted time-courses to create a null distribution. Other options exist as well, but these are the main options (aside from the LOO-approach) (Chen et al, 2016, 2017; De Angelis et al., 2020, Nastase et al., 2019). In Figure E1, we demonstrate that the results are practically identical, especially for the sample of 200 audience and our region-based approach. Moreover, we also show that at a higher statistical threshold (leading to more conservative results), the observed ISC stays significant (Lindquist & Meija, 2015).

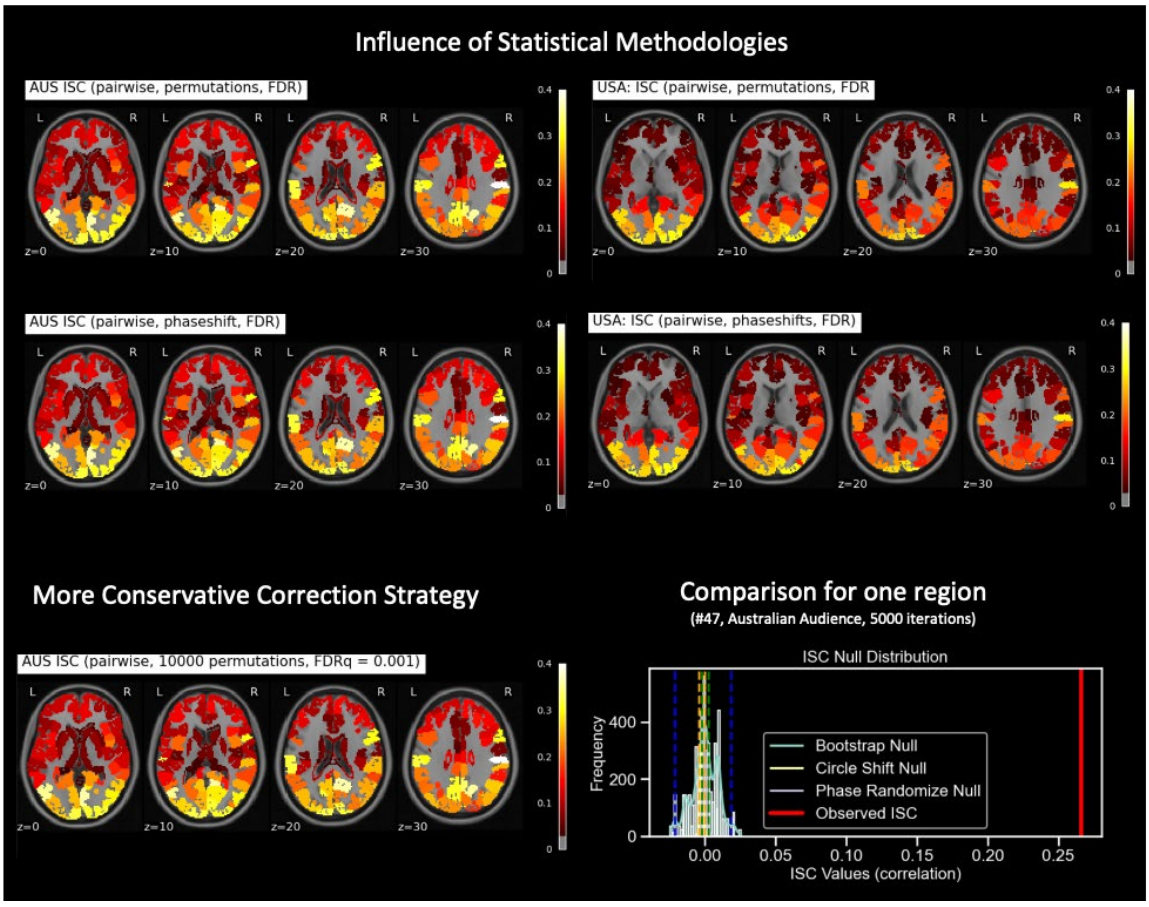

Figure E1. Results for different analysis methods and parameters. Top panels: ISC values per audience, corrected based on the statistical approach described in the title. As can be seen, the different approaches lead to the same pattern of significant ISC results, even at higher thresholds (bottom left panel). Bottom right panel: Exemplary results for the null distribution based on different approaches. As described in the literature, the bootstrap approach is more conservative (broader distribution), but the observed ISC (red line) is far outside of the CI for

all the different approaches (phase-randomization, circular shifting, and bootstrapping). See text for details.

3. Supplementary Results and Control Analyses

3.1. Supplementary Analysis: Structural Similarities

The first analysis was motivated by the notion that twins' brains might be anatomically more similar, which might drive at least some of the resulting functional differences, despite the fact that all brains were normalized into a common MNI space. To that end, we derived a metric of anatomical similarity by correlating the normalized T1 images, i.e., vectorizing the grayscale image values and then comparing similarity in much the same way we compared functional time series. The results show that brain-to-brain anatomical similarity is high, as expected given the normalization (around 0.7-0.8), but varies slightly across individuals. When testing whether anatomical similarity is related to ISC (functional response similarity), we find a number of regions exhibiting positive effects for the twin-comparison (62), but far fewer for the control pairs (6) and others (12). Although these studies are preliminary and should be supplemented with more in-depth analyses such as shared response modeling and hyperalignment approaches (Nastase et al., 2019), they may imply that ISC differences are, at least in part, based on anatomical similarities.

3.2. Supplementary Analysis: Predicting Kinship status from ISC

The second set of supplementary analyses focused on prediction and classification. Given that we found differences in ISC between twins and non-twins, we can also ask an intriguing question: Can we use the ISC strength between two people to predict whether the pairings are twins or not? This transposes our conventional design, where 'relationship status' represents the independent

variable (IV) and the ISC is the dependent variable (DV). In the resulting transposed design, ISC represents a feature/IV, and the 'relationship status' represents the target variable/DV. Recasting the study in this way turns the analysis into a classification problem. To investigate this research problem, we created a predictive modeling pipeline that used ISC results as features to classify the target variable (twin vs. no-twin relationship status). More specifically, we obtained the ISC from all 293 regions for the 100 twin pairs and computed ISC for another group of 100 randomly selected individuals (again for all 293 regions). This way, we obtain a balanced dataset (100 twins and 100 non-twins) that is easy to interpret, annotate, and analyze. We then used a standard logistic regression model with a leave-one-out cross-validation strategy, and tested model performance against a baseline dummy classifier. As expected given the balanced dataset, the dummy classifier performed at chance level (50%), whereas the logistic regression model using actual data (comprising measured response similarities across all brain regions) performed with an accuracy of 67% on average. In other words, by just comparing the inter-brain similarity during movie reception between two viewers in this sample, we could classify correctly whether the two were twins in nearly 7 out of 10 cases.

3.3. Control Analysis: Cross-continental Response Similarity Analysis based on Time Series

We provide an additional time-series based analysis for the inter-continent comparison. The main inter-continent-comparison was based on comparing the spatial pattern of ISC found in each audience - i.e. computed separately among Australian and US viewers, and then compared. Readers might ask why we did not compare the underlying time series directly, as they watched the same movie. The issue is that the dataset from Brisbane has a different TR (time of repetition) that the one from Boston; in other words, the data in Australia were sampled faster/more often. In addition, the moment at which the movie was started differed a bit between sites, and the movie

differed a bit (in terms of its credits, onset, offset). This all makes it more difficult to align the neural time series across continents that it might initially seem. On top of this all, the fMRI data were acquired with different scanners and processed somewhat differently, which is another challenge. Therefore, the main analysis in the paper is in our view more robust, as it only compares data that are actually comparable (i.e. homogeneously acquired). Comparing the results of these two samples seems fine, but comparing the time series is more challenging for all these reasons. Nevertheless, we also compared the time series directly (i.e. upsampling the Bostonian dataset to match the length of the Brisbane data, and correcting as good as we could for the differences), and we report exemplary results for two regions (other regions show the same pattern), but readers should keep in mind that the way in which we cut and pieced together the time series from different regions is ad hoc. The results clearly support that the different regions also show similar brain response time courses, which is assuring.

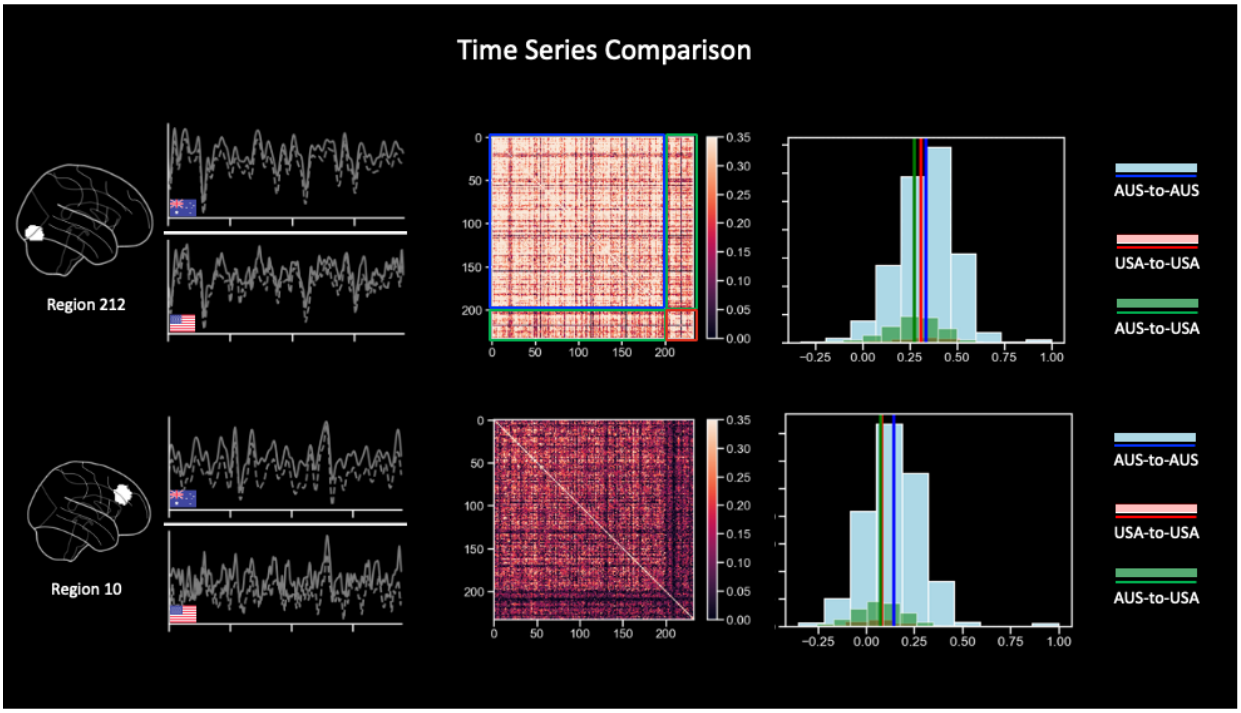

Figure E2. Inter-Continental ISC Comparison based on Time Series. See text for details.

3.4. Control Analysis: Twin-Twin ISC vs. Matched Others (Age/Sex) vs. Random Others

Lastly, as reported in the main paper, we also compared the twin-ISC to a "control-ISC" that correlated data from a given viewer to another viewer the same age and sex. This was done to have viewers who are the same age and sex, but don't share the twin-status except for those two similarity characteristics. Note that compared to the traditional pairwise (all-to-all) ISC comparison, the ISC among twins is also pairwise, but there are only 100 twin-pairs. Thus, we constructed a list of "matched pairs" in which the 100 twin_A-individuals were randomly matched with a matching person of the same age and gender, but not their own twin. Although we could in principle permute this matching procedure, the unbalanced nature of age and gender distribution makes this challenging. Additionally, if there are e.g. only a few twin dyads of 14-year old females, the possible ways of matching them are small in the first place. Thus, we simply constructed a few matching datasets and never found differences in the results. Moreover, as the "random others" comparison, we also simply randomly drew a person for every twin (see more complex analysis in the main paper that computes ISC across all possible pairings). This analysis thus compares ISC among twins against two reference groups, one group of age/sex-matched controls (rather similar), and one group of random others (less similar). The results, which are presented in Figure E3, show a clear pattern: ISC among twins is highest and the ISC of matched controls is more similar to the random others, and never higher than ISC among the real twins.

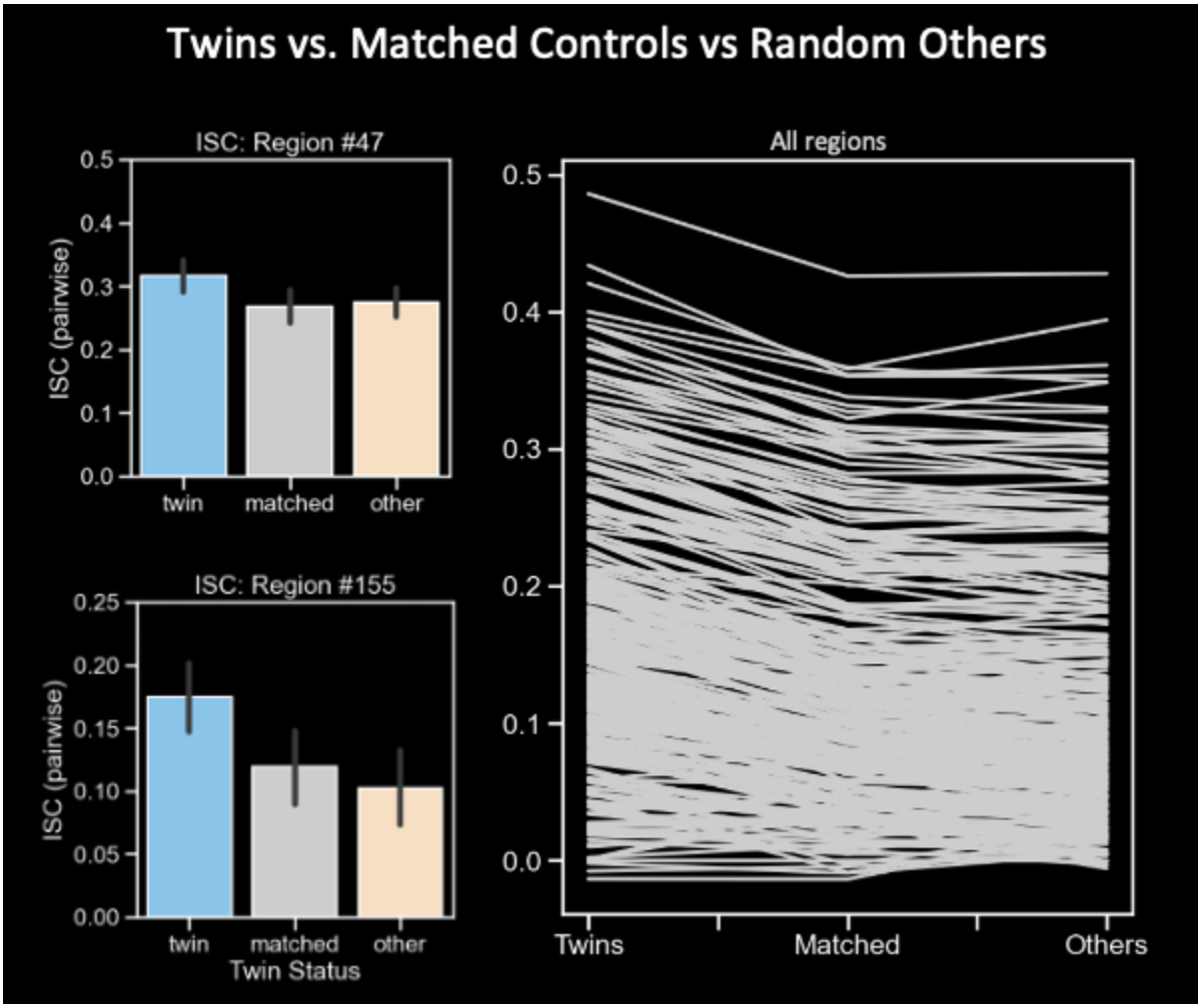

Figure E3. ISC among real twins, matched controls (matched for age and sex), and random others. See text for details.

3.5. Control Analysis: ISC among Fraternal Twins

Lastly, we turn to the issue of twin status. In brief, the twins scanned here were of two types - fraternal (dizygotic, DZ) and identical (monozygotic, MZ). Identical twins are starting our from the same cell and thus have a highly similar genome, whereas fraternal twins are starting from two eggs and are thus more like siblings, yet born at the same time and usually raised jointly. Monozygotic twins are thus genetically more similar than fraternal ones, yet both twins are usually

raised in the same shared environment. The data provided by the QTAB study are shared in two ways – one completely open via OpenNeuro (including fMRI data and basic information about age and sex, and which persons are twins). However, the QTAB also provides restricted data (with more in-depth genetic information, as well as other functional metrics), but for these a researcher needs to apply for access and comply with the imposed privacy requirements. Critically, the twin status, i.e. whether a pair of twins is DZ or MZ is part of the private data. Hence, we cannot publicly share it. However, we were still interested in exploring the potential influence of the additional similarity that MZ twins have over DZ ones, and we thus devised the following strategy: Given that the QTAB dataset provides sex-related information, we can certainly know that if a twin pair includes both sexes, it must be DZ. This was the case in 24 out of 100 twin pairs. For the remaining 76, it is overall more likely that they are MZ (because statistically there should be ca. as many same-gender DZ twin pairs). In any case, we can assume that in the remaining 76 twin pairs, the MZ twins are the majority. This gives us the opportunity to compare the more similar group of "likely MZ" twin pairs against the still genetically similar, yet somewhat less similar group of "certainly DZ". When running this analysis, we find that indeed the "likely MZ" twin pairs exhibit higher ISC compared to the "certainly DZ" pairs (see Supplementary Figure E4). We note that given the uncertain status of the "likely MZ" group, this analysis should be interpreted with caution, yet it certainly underscores the main point made in the paper, i.e. that the pre-existing similarities among twins (vs. others, MZ twins vs. DZ twins, or even across audiences) underpin the shared brain responses to the same movie studied here.

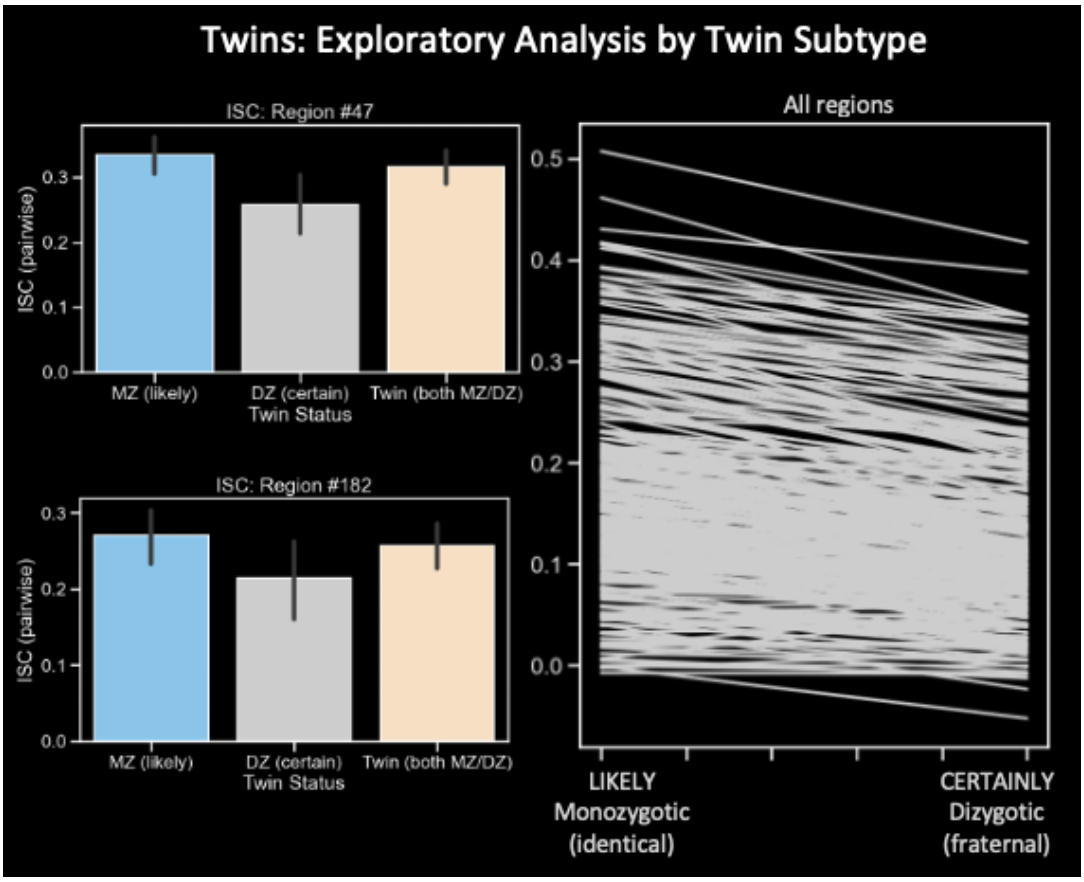

Figure E4. ISC among twins, likely MZ twins, and certainly DZ twins. See text for details.

Supplementary References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux., G. 2014. Machine Learning for Neuroimaging with Scikit-Learn. *Frontiers in Neuroinformatics 8.* https://doi.org/10.3389/fninf.2014.00014.

Andersson, J. L. R., Skare, S, and Ashburner, J. 2003. How to Correct Susceptibility Distortions in Spin-Echo Echo-Planar Images: Application to Diffusion Tensor Imaging. *NeuroImage 20 (2):* 870–88. https://doi.org/10.1016/S1053-8119(03)00336-7.

Avants, B. B., C. L. Epstein, M. Grossman, and J. C. Gee. 2008. Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and

Neurodegenerative Brain. *Medical Image Analysis 12 (1):* 26–41. https://doi.org/10.1016/j.media.2007.06.004.

Behzadi, Y, Restom, K., Liau, J and Liu., T.W., 2007. A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI. *NeuroImage 37 (1)*: 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042.

Chen, G., Shin, Y. W., Taylor, P. A., Glen, D. R., Reynolds, R. C., Israel, R. B., & Cox, R. W. (2017). Corrigendum to "Untangling the relatedness among correlations, Part I: Nonparametric approaches to inter-subject correlation analysis at the group level. *NeuroImage, 145,* 130.

Cox, R W., and Hyde, J.S. 1997. Software Tools for Analysis and Visualization of fMRI Data. N*MR in Biomedicine 10* (4-5): 171–78. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5

De Angelis, L., Gazzola, V., & Keysers, C. (2020). Parametric tests for Leave-One-Out Inter-Subject Correlations in fMRI provide adequate Type I error control while providing high sensitivity. *BioRxiv,* 2020-07.

Esteban, O., Blair, R., Markiewicz, C.J., Berleant, S.L., Moodie, C., Ma, F., Isik, A.I., et al. 2018. fMRIPrep. *Software.* https://doi.org/10.5281/zenodo.852659.

Esteban, O., Markiewicz, C., Blair, R.W., Moodie, C., Isik, A.I., Aliaga, A.E., Kent, A., et al. 2018. fMRIPrep: A Robust Preprocessing Pipeline for Functional MRI. *Nature Methods.* https://doi.org/10.1038/s41592-018-0235-4.

Fonov, VS, AC Evans, RC McKinstry, CR Almli, and DL Collins. 2009. Unbiased Nonlinear Average Age-Appropriate Brain Templates from Birth to Adulthood. *NeuroImage 47,* Supplement 1: S102. https://doi.org/10.1016/S1053-8119(09)70884-5.

Gorgolewski, K., Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. Ghosh. 2011. Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics 5:* 13. https://doi.org/10.3389/fninf.2011.00013.

Gorgolewski, K. J., Esteban, O., Markiewicz, C.J., Ziegler, E., Ellis, D.G., Notter, M.P., Jarecka, D., 2018. Nipype. Software. https://doi.org/10.5281/zenodo.596855.

Greve, D.N, and Fischl,B. 2009. Accurate and Robust Brain Image Alignment Using Boundary-Based Registration. *NeuroImage 48 (1):* 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060.

Haufe, S., DeGuzman, P., Henin, S., Arcaro, M., Honey, C. J., Hasson, U., & Parra, L. C. (2017). Reliability and correlation of fMRI, ECOG and EEG during natural stimulus processing. BioRxiv, 207456.

Jenkinson, M., Bannister, P., Brady, M. and Smith, S. 2002. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage 17 (2):* 825–41. https://doi.org/10.1006/nimg.2002.1132.

Jenkinson, M, and Smith, S. 2001. A Global Optimisation Method for Robust Affine Registration of Brain Images. *Medical Image Analysis 5 (2):* 143–56. https://doi.org/10.1016/S1361-8415(01)00036-6.

Lanczos, C. 1964. Evaluation of Noisy Data. J*ournal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis 1 (1):* 76–85. https://doi.org/10.1137/0701007.

Lindquist, M. A., & Mejia, A. (2015). Zen and the art of multiple comparisons. Psychosomatic medicine, 77(2), 114.

Power, J. D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B,.L. and Petersen. S.E. (2014). Methods to Detect, Characterize, and Remove Motion Artifact in Resting State fMRI. *NeuroImage 84*: 320–41. https://doi.org/10.1016/j.neuroimage.2013.08.048.

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications, 9(1),* 1027. https://doi.org/10.1038/s41467-018-03399-2

Strike, L. T., Hansell, N. K., Chuang, K. H., Miller, J. L., de Zubicaray, G. I., Thompson, P. M., & Wright, M. J. (2023). The Queensland Twin Adolescent Brain Project, a longitudinal study of adolescent brain development. *Scientific Data, 10(1),* 195. https://doi.org/10.1038/s41597-023-02038-w

Satterthwaite, T. D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J. , Calkins, M.E., Eickhoff, S.B. 2013. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage 64 (1):* 240–56. https://doi.org/10.1016/j.neuroimage.2012.08.052.

Shen, X., Tokoglu, F., Papademetris, X., & Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage*, 82, 403–415. https://doi.org/10.1016/j.neuroimage.2013.05.081

Tustison, N. J., B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. 2010. N4itk: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging 29 (6):* 1310–20. https://doi.org/10.1109/TMI.2010.2046908.

Zhang, Y., M. Brady, and S. Smith. 2001. Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. *IEEE Transactions on Medical Imaging 20 (1): 45–57.* https://doi.org/10.1109/42.906424.