



HUMAN
Connectome
PROJECT



HUMAN
Connectome
PROJECT

WU-Minn HCP 1200 Subjects Data Release Reference Manual

1 March 2017

updated 11 April 2017 (*Family Structure variables update*)

Table of Contents

| | |
|---|----|
| Introduction and overview | 7 |
| <i>What's in the HCP 1200 Subjects release?</i> | 10 |
| <i>What's different from the 900 Subjects (Dec 2015) + initial 7T (June 2016) releases?</i> | 12 |
| <i>Multiple levels of data processing</i> | 14 |
| <i>How do I access HCP data?</i> | 16 |
| <i>How can I analyze and visualize HCP data?</i> | 17 |
| <i>What's in store for future releases?</i> | 17 |
| <i>How were the HCP datasets acquired and processed?</i> | 18 |
| <i>What if I want to adapt HCP protocols to my own scanner and research projects?</i> | 18 |
| Open Access and Restricted Data | 18 |
| What about technical support, bug reports, and feature requests? | 19 |
| HCP Data Release Updates Wiki and Data Patches | 19 |
| Gradient nonlinearities in unprocessed datasets | 20 |
| Chapter 1: Getting HCP Data and using ConnectomeDB | 21 |
| HCP Data Sizes | 21 |
| How to access HCP data | 23 |
| How to explore HCP data in ConnectomeDB | 24 |
| How to download HCP data | 27 |
| First time Aspera Download Setup | 28 |
| Handling downloaded HCP data | 30 |
| Connectome in a Box | 30 |
| Accessing HCP Data on the Cloud | 31 |
| Chapter 2: HCP Hardware and Protocols | 33 |
| Standard Operating Procedures (SOPs) | 33 |
| Standard schedule for subject visits | 33 |
| 3T MR scanner and other hardware | 34 |
| Summary of 3T MR imaging protocols | 35 |
| Structural session | 35 |
| Resting-state fMRI (rfMRI) | 35 |

| | |
|--|----|
| Task-evoked fMRI (tfMRI) | 36 |
| Diffusion imaging (dMRI) | 37 |
| Full scanning protocols (PDF) | 38 |
| Standard 3T MR session structure | 38 |
| MR Physiological Data | 38 |
| MR Image Reconstruction and DICOM to NIFTI conversion..... | 39 |
| Initial Structural MR QC and File Transfer to ConnectomeDB | 40 |
| Subjects with Issues noted in the HCP quality control process | 41 |
| 3T Task fMRI scripts and protocol details..... | 46 |
| tfMRI scripts and data files | 46 |
| Details of tfMRI tasks | 49 |
| 7T MR scanner and other hardware | 54 |
| Summary of 7T MR imaging protocols | 54 |
| 7T Retinotopy and Movie Watching stimuli and protocol details | 54 |
| MEG hardware specifications and sensor locations | 55 |
| Scanner hardware and parameters | 55 |
| Sensor locations..... | 55 |
| Stimulus presentation information | 57 |
| Summary of MEG acquisition protocols..... | 57 |
| Resting-state MEG (rMEG) | 58 |
| Task MEG (tMEG)..... | 58 |
| Full scanning protocols (PDF) | 59 |
| Standard MEG session structure..... | 59 |
| MEG Physiological Data..... | 60 |
| Determining Alpha and Beta Peak Frequencies for Individual Subjects..... | 60 |
| MEG Data Conversion: 4D import to FieldTrip..... | 60 |
| MEG Quality control and artifact identification | 60 |
| tMEG scripts and protocol details..... | 61 |
| tMEG scripts and data files..... | 61 |
| MEG vs. MRI Naming Conventions | 62 |
| Details of the tMEG Working Memory task | 62 |

| | |
|---|-----|
| Details of the tMEG Language Processing (Story-Math) task | 65 |
| Details of the tMEG Motor task..... | 69 |
| Summary of Behavioral and Individual Subject measures | 73 |
| Chapter 3: 1200 Subjects Release Data | 76 |
| S1200 HCP Subject Family Structure..... | 78 |
| HCP Users and MEG Data FAQs..... | 81 |
| Directory structure for unprocessed MR data | 81 |
| Directory structure for preprocessed MR data | 84 |
| ICA-FIX denoised rfMRI data | 87 |
| 3T Fix and Fix_extended datasets..... | 89 |
| Individual (within-subject) tfMRI analysis data | 90 |
| Group-average functional MR data..... | 91 |
| Task fMRI group analysis | 93 |
| Resting-state group analysis and functional connectivity | 94 |
| File Names and Directory Structure for Group-average fMRI Data | 95 |
| Parcellation-Timeseries-Netmats (PTN) extensively processed rfMRI data | 96 |
| Data pre-processing and Group-ICA | 97 |
| Node timeseries (individual subjects) | 97 |
| Network matrices (individual subjects and group-averaged) | 98 |
| Example simple MATLAB/FSLNets code | 99 |
| Directory structure for unprocessed MEG data..... | 100 |
| Directory structure for MEG anatomy data (anatomical models for source reconstruction) .. | 102 |
| Directory structure for Channel- (preprocessed) and Source-level processed MEG data | 103 |
| Datacheck results..... | 104 |
| Baddata results | 105 |
| Icaclass and Icaclass_qc results | 106 |
| Rmegpreproc results | 106 |
| Powavg results..... | 107 |
| Tmegpreproc results | 107 |
| Eravg results | 107 |
| Tfavg results..... | 108 |

| | |
|---|-----|
| Icamne results | 109 |
| Icablpenv results | 109 |
| Icablpcorr results | 110 |
| Icaimagcoh results | 111 |
| Bfblpenv results | 111 |
| Bfblpcorr results | 112 |
| Srcavglcmv results | 112 |
| Srcavgdics results | 113 |
| Chapter 4: HCP Processing Pipelines | 116 |
| MR Preprocessing pipelines | 116 |
| Overview of MR Minimal Processing Pipelines | 117 |
| MR Structural Pipelines | 118 |
| MR Functional Pipelines | 121 |
| MR Diffusion Pipeline | 123 |
| bedpostX pipeline for MR Diffusion Analysis | 123 |
| Overview of MEG Processing pipelines | 124 |
| MEG Data Processing Output File Formats | 126 |
| MEG Anatomy processing pipeline | 129 |
| MEG Preprocessing Pipelines | 131 |
| "datacheck" pipeline | 131 |
| "baddata" pipeline | 132 |
| "icaclass" pipeline | 136 |
| "icaclass_qc" pipeline | 138 |
| Manual improvements/corrections to Anatomy and Preprocessing pipeline results | 139 |
| rMEG Channel-level Processing Pipelines | 140 |
| "rmegpreproc" pipeline | 140 |
| "powavg" pipeline | 141 |
| rMEG Source-level Processing Pipelines | 141 |
| "icamne" pipeline | 141 |
| "icablpenv" pipeline | 142 |
| "icablpcorr" pipeline | 143 |

| | |
|---|-----|
| "icaimagcoh" pipeline | 144 |
| "bfbldenv" pipeline | 145 |
| "bfbldcorr" pipeline | 146 |
| "icablddyn" pipeline | 146 |
| tMEG Channel-level Processing Pipelines | 146 |
| "tmegpreproc" pipeline | 149 |
| "eravg" pipeline | 153 |
| "tfavg" pipeline | 159 |
| Alpha Beta peak frequency processing | 167 |
| tMEG Source-level Processing Pipelines | 168 |
| "srcavgdics" pipeline | 168 |
| "srcavgldmv" pipeline | 171 |
| "tmegconnebasic" pipeline | 173 |
| Chapter 5: Behavioral and Individual Subject measures details | 177 |
| FreeSurfer measures | 179 |
| Descriptions for HCP Behavioral and Demographic Measures | 179 |
| Subject Information | 179 |
| Study Completion and MR Sessions | 182 |
| Alertness measures..... | 182 |
| Cognition measures | 183 |
| Emotion measures | 187 |
| Motor measures | 187 |
| Personality measures..... | 188 |
| Health and Family History | 188 |
| MEG Subjects | 189 |
| Psychiatric and Life Function..... | 189 |
| Sensory measures | 190 |
| References..... | 193 |
| Appendices | 202 |

Introduction and overview

This document provides information and guidance on how to use data released by the WU-Minn Human Connectome Project (HCP) consortium in the 1200 Subjects release (S1200 Release, February 2017) from the original young adult HCP study.

The HCP aimed to study and freely share data from 1200 young adult (ages 22-35) subjects from families with twins and non-twin siblings, using a protocol that includes structural and functional magnetic resonance imaging (MRI, fMRI), diffusion tensor imaging (dMRI) at 3 Tesla (3T) and behavioral and genetic testing. A subset of ~50 same-sex twin pairs were additionally studied by resting-state/task magnetoencephalography (MEG) and another subset of ~100 same-sex twin pairs were studied using rfMRI, tfMRI, and dMRI at 7 Tesla (7T) at University of Minnesota (UMinn, see [Van Essen et al., 2013](#)). In all parts of the HCP, participants were scanned on the same equipment using the same protocol for every subject.

Given the richness of the HCP datasets and their utility for a wide range of research purposes, it is important that potential users understand what data are currently released, how the datasets are organized, how they can be accessed, what has changed since the last releases, what's in store for future releases, and what to anticipate with the advent of the HCP Lifespan studies of children and adolescents (HCP-Development, HCP-D) and older adults (HCP-Aging, HCP-A).

Should I be interested in the HCP 1200 Subjects Data release? The S1200 release is the final release of new subjects that were scanned as part of the HCP young adult study, adding to the number of high quality unprocessed, processed, and analysis 3T and 7T MRI, plus sensor- and source-level MEG datasets available through HCP. This includes the addition of 3T MRI data for 200+ new subjects and functional and diffusion 7T MRI data on 100+ new subjects since the HCP 900 Subjects and initial 7T data releases. For the first time, 3T MRI and behavioral Retest data for 46 subjects (all monozygotic twins) is also available.

Final HCP Study (Young Adult) Overall Totals:

1206 Subjects with Behavioral and Demographic Measures

1113 Subjects with 3T MRI (at least structural MRI)

45 Subjects with 3T Retest MRI and Behavioral Measures (46 subjects)

184 Subjects with 7T MRI*

95 Subjects with MEG*

1142 Subjects with Genotyping data (at least 1 gene chip)

**All 7T and MEG subjects were also scanned at 3T.*

HCP Subjects include:

168 Monozygotic Twin pairs** (336 Subjects), including:

105 MZ twin pairs (210 Subjects) with complete 3T image data

103 Dizygotic Twin Pairs** (224 Subjects), including

58 DZ twin pairs (116 Subjects) with complete 3T image data

***determinations of MZ and DZ twin zygosity by genotyping, if available, otherwise by self-report, see [S1200 HCP Subject Family Structure](#) for more details on family structure in the HCP dataset.*

If you are interested in further parsing how many subjects have complete data of a particular modality, or set of modalities, or have a certain range of values for an individual subject measure there are many possible ways to filter the individual subject data in [ConnectomeDB](#), our database platform for sharing HCP imaging and individual subject data. See [How to explore HCP data in ConnectomeDB](#) to learn more about navigating the ConnectomeDB Subject Dashboard and setting filters to search for subjects that meet the the desired criteria. Note: Some subject measures such as Family_ID and ZygosityGT (twin zygosity verified by genotyping) are restricted and require users to be approved for restricted access before they can be viewed or filtered on in ConnectomeDB, see [Open Access and Restricted Data](#) for more details.

The S1200 release includes **re-preprocessed dMRI data (3T and 7T) for all subjects** with an updated HCP diffusion pipeline that incorporates a major improvement in FSL's EDDY that adds image slice outlier detection to remove noise caused by subject movement. This release is also the first time we are releasing 3T dMRI data further processed with FSL's BEDPOSTX (Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques, modeling crossing X fibers) to model white matter fiber orientations and crossing fibers for probabilistic tractography.

With the S1200 release, we now have genotyping data (full genome-wide SNP genotyping data to be released in spring 2017 on dbGaP) available on 1142 of the 1206 total released subjects. Among sibling subjects for which we have genotyping data, we have now genetically confirmed the self reported Mother_ID, Father_ID, and zygosity (for twins) [restricted data](#) demographics measures. Therefore, we have updated the Mother_ID and Father_ID measure values in ConnectomeDB based on the available genotyping data, and for a limited number of subjects (5 subjects with changes to Mother_ID and 28 subjects with changes to Father_ID), these values have changed from what they were in previous releases. Mother_IDs and Father_IDs that have been changed based on genotyping have a 999xx ID number to facilitate filtering for affected subjects.

We have added a “HasGT” measure to indicate which subjects have genotyping data available for family relationship verification and genetic analysis (once the gene chip data is released).

We have also added a “Family_ID” measure that lists all Mother_IDs and Father_IDs associated with a family to facilitate finding related subjects, even for families containing half siblings.

For zygosity, we have added the measure ZygosityGT, which reflects twin zygosity verified by genotyping. We have also kept the original self-reported zygosity information in the database as ZygositySR so that the zygosity data can be tracked data from previous HCP releases that only contained self-reported zygosity under the “Zygosity” measure. See [Zygosity, Genotyping, Family, and Parent IDs](#) in Chapter 5 for more details.

Important! Users who are interested in incorporating family structure into their analyses, especially for heritability analyses, **should be aware that a substantial number (36 twin pairs) of HCP twins who self-reported to be dizygotic (DZ) twins** were found by the genotyping to be monozygotic (MZ) twins. Other twin studies have found that it is common for some MZ twins to believe they are DZ twins based on what hospital staff told parents at child birth without genetic verification. See [S1200 HCP Subject Family Structure](#) for more details.

Another measure added to ConnectomeDB for the S1200 release, the **QC_Issues measure**, indicates subjects who were found in the quality control process to have notable brain anatomical, processing, or data issues that we would like users to be aware of for their analyses. Codes for each issue included are detailed in [Subjects with Issues noted in the HCP quality control process](#) and more specific information on these subjects is available on the [HCP QC Issues wiki page](#).

The current 1200 Subjects release includes behavioral measures, MRI, and MEG data on subjects collected in the HCP production data phase (2012-2015), including unprocessed, preprocessed, and additionally processed 3T MRI, 7T MRI, and MEG data. As stated above, the S1200 release is the final HCP release of new imaging data at 3T and 7T. Previous HCP data releases included three quarterly releases (March, June, and September 2013), a MEG1 release (March 2014), the 500 Subjects Release (June 2014), the 500 Subjects + MEG2 release (November 2014), the 900 Subjects Release (December 2015), and initial 7T Subjects functional imaging data release (June 2016).

Although the S1200 release is the last data release of new subjects that participated in young adult HCP, continued data releases (genetics, group average data, individual-subject cortical parcellation datasets, etc.) of the S1200 subjects are planned as they become available in 2017 and beyond.

For the 500 Subjects release, important changes were implemented in the preprocessing pipelines for MRI data (HCP pipelines v3.4), including the MSM-Sulc (cortical folding-based) registration algorithm and software updates that affected all preprocessed MRI data. For the

900 Subjects Release, the [MSM-All pipeline](#) was added, which performed multimodal surface matching (MSM) registration based on areal features including information from sulcal depth, myelin, and functional connectivity maps to substantially improve intersubject registration of functional cortical areas; it results in markedly sharper group average data. Also for the S900 release, a few other incremental changes were added to the v3.x HCP pipelines that produced other data useful to users, for example, resting state stats and post-FIX results data that allow one to evaluate signal and noise components classified by FIX.

For the S1200 Subjects Release, the pipeline changes are the addition of new subpipelines (also v3.x) that support the update from EDDY 5.0.9 to 5.0.10 in the diffusion pipeline (v3.19.0), addition of the [bedpostX pipeline](#), updates to the bias field correction in the functional preprocessing pipeline (PreFreeSurfer pipeline, but applied only to 7T data, v3.15.1), and updates to the ICA-FIX (v3.16.1), DeDriftAndResample (v3.17.0), Resting State Stats pipelines for 7T data (v3.21.0). See the code and notes on [Github](#) for more details.

When processing new subjects as part of the S900 and S1200 releases, care was taken to process the data with the same pipeline version as was used for the S500 release (for a given generated output). Thus, in general, data from the S500 release can be directly compared in analyses with the S900 and S1200 releases for the same output file. The notable exception is that with the S1200 release, the diffusion data for ALL subjects has been reprocessed with the updated diffusion pipeline (v3.19.0) that implemented the use of EDDY outlier replacement, and thus produces cleaner diffusion data than the previous (S500 and S900) releases. Due to this upgrade, we recommend that users only use S1200 release data to compare preprocessed diffusion data between subjects.

As a reminder, HCP **strongly advises** against mixing data processed with different versions of the processing pipelines. Check the release notes in the {SubjectID}/releasenotes directory distributed with each subject dataset to ensure all data used in your analyses have been processed with the same versions. For example, data released at the 500, 900, and 1200 Subjects Releases were processed using the v3.x branch of the HCP pipelines and should not be used in analyses with HCP data processed with v1 or v2 versions of the HCP pipelines (from the initial “Q1”, “Q2” and “Q3” releases).

What's in the HCP 1200 Subjects release? The 1200 Subjects release includes behavioral and 3T MR imaging data from 1206 healthy young adult participants collected August 2012–October 2015, including:

- 3T MR structural scans available for 1113 subjects.
- 889 subjects have fully complete data for all the four 3T MRI modalities in the HCP protocol: structural images (T1w and T2w), resting-state fMRI (rfMRI), task fMRI (tfMRI), and high angular resolution diffusion imaging (dMRI).

- 46 subjects (all monozygotic twins, 21 twin pairs + 4 MZ twins without retest of co-twin) have 3T HCP protocol Retest data available.
 - 36 Retest subjects have fully complete retest data for all 3T modalities: structural, rfMRI, tfMRI, and dfMRI.
- 184 subjects also have 7T MR scan data available (in addition to 3T MR scans).
 - 175 of the 7T subjects have fully complete data for all the 7T MRI modalities in the HCP protocol: rfMRI, retinotopy fMRI, movie-watching fMRI, and dMRI.
- 95 subjects also have at least some resting-state MEG (rMEG) and/or task MEG (tMEG) data available (in addition to 3T MR scans).
- Final family structure summary information reflecting all released subjects, see [S1200 HCP Subject Family Structure](#) for more details.

More specifically, the 1200 Subjects release includes:

- Unprocessed T1w and T2w (3T only), rfMRI, tfMRI, dMRI 3T and 7T imaging data in NIFTI format.
- 3T and 7T MR imaging data preprocessed using v3 pipelines; including outputs based on MSM-Sulc (cortical folding only-based registration) and MSM-All registration (multi-modal registration based on areal features derived from myelin, Resting State Network (RSN), and rfMRI visuotopic maps). More details in [MR Preprocessing pipelines](#).
- 3T and 7T rfMRI 4D volume and grayordinate (surface vertices + subcortical voxels) data 'cleaned' of structured artifacts using FSL's FIX (FMRIB's ICA-based Xnoiseifier). Resting State Statistics are also available. See [ICA-FIX denoised rfMRI datasets](#) for details.
- 7T Movie-watching task fMRI data (4D volume and grayordinate) similarly 'cleaned' of structured artifacts using FSL's FIX (FMRIB's ICA-based Xnoiseifier).
- Individual 3T task fMRI grayordinate-based analysis results (available at 2mm and 4mm total smoothing levels). See [Individual \(within-subject\) tfMRI analysis data](#) for details.
- 3T and 7T diffusion MRI preprocessed with a new v3.19.0 version of the diffusion pipeline that supports the use of a new version of FSL's EDDY that adds image slice outlier detection and replacement to remove noise caused by subject movement. More details in [MR Diffusion Pipeline](#).
- 3T diffusion MRI further processed with FSL's BEDPOSTX. See [bedpostX pipeline for MR Diffusion Analysis](#) for details.
- Structural MRI-based MEG anatomical models and MEG scans (unprocessed, channel-level processed, and source-level processed) from 95 subjects, including 50 individual

complete datasets with resting state (rMEG) and each of the three task protocols (tMEG: motor, working memory, and language processing).

- Behavioral and demographic data on 1206 subjects, including genetically verified zygosity for most twin pairs, available for viewing in the ConnectomeDB dashboard or for download as comma separated variable (.csv spreadsheet format) files. See [How to explore HCP data in ConnectomeDB](#) for details.

Also, up to date HCP software is available:

- HCP Pipelines for processing 3T and 7T MR data and associated documentation/FAQs are available at: <https://github.com/Washington-University/Pipelines>
- Connectome Workbench v1.2.3 open-source visualization and analysis software designed as part of HCP (also a prerequisite for running HCP pipelines) is available at: <http://humanconnectome.org/software/get-connectome-workbench.html>. Source code at: <https://github.com/Washington-University/workbench>
- megconnectome software v3.0 scripts used to process the MEG data (no changes from the S900 release). See <https://www.humanconnectome.org/documentation/HCP-pipelines/meg-pipeline.html> for more information.

If you are interested in accessing the HCP data, please continue reading the following important information and see more details in [Chapter 3: 1200 Subjects Release Data](#).

What's different from the 900 Subjects (Dec 2015) + initial 7T (June 2016) releases?

- ***All released 3T Diffusion data on all HCP subjects re-processed with updated diffusion preprocessing pipeline.*** All 3T diffusion data was re-processed using a new v3.19.0 version of the diffusion pipeline that incorporates an updated version of FSL's EDDY tools that adds slice outlier detection and replacement to remove noise caused by subject movement. Therefore, diffusion data from previous releases should not be mixed in analyses with the "cleaner" dMRI data now available in the S1200 release.
- ***Addition of 3T Retest data.*** Subjects (all monozygotic twins, 21 twin pairs + 4 MZ twins without retest of co-twin) were recruited to undergo the full 3T HCP imaging and behavioral protocol for a second time. Behavioral and 3T Retest imaging data is available on 46 and 45 of these subjects, respectively. Retest datasets are available in the WU-Minn HCP Retest Data project on the ConnectomeDB splashpage. 7T data is available on 22 Retest subjects. See [Chapter 3: 1200 Subjects Release Data](#) for more information on accessing and using the Retest data.
- ***Addition of preprocessed 7T Diffusion data.*** Only unprocessed Diffusion data was released as part of the Initial 7T Release (June 2016). For all 7T subjects with dMRI

scans, 7T dMRI data preprocessed v3.19.0 version of the diffusion pipeline is now available.

- **Resting State Stats added to 7T rfMRI ICA-FIX cleaned datasets.** Resting state stats were not available as part of the 7T rfMRI ICA-FIX Extended packages of the Initial 7T Release (June 2016). These files provide information about different types of 'noise' and 'signal' in HCP resting state data, gleaned by partitioning the variance according to different processing stages in the FIX denoising pipeline. For more information about the types of variance displayed in these files, please review the RestingStateStats script in the Github Pipelines: <https://github.com/Washington-University/Pipelines/tree/master/RestingStateStats>
- **Genetically verified family structure measures.** We have added restricted data measures HasGT, Family_ID and ZygosityGT, and updated Mother_ID and Father_ID based on genotyping data available from blood and saliva samples from HCP subjects. See [Zygosity, Genotyping, Family, and Parent IDs](#) for more details. *For some subjects, these genetically verified values for these measures have changed from what they were in previous releases.* Notably, **36 HCP twin pairs who self reported (now the ZygositySR measure) as dizygotic twins were found to be genetically monozygotic.**
- **QC_Issue measure added.** A subject data measure has been added to ConnectomeDB to flag subjects with notable brain anatomical, processing, or data noise issues found in the HCP Quality Control process. The issues are notable, but were not considered severe enough to exclude the subject's imaging data from release. Codes for each issue included are detailed in [Subjects with Issues noted in the HCP quality control process](#) and more specific information on these subjects is available on the [HCP QC Issues wiki page](#).
- **New, cleaner FIX-denoised rfMRI data available for 11 previously released subjects.** An analysis of resting state data noise statistics after automated FIX-ICA processing revealed 11 subjects that could benefit from reapplying FIX with some of the originally identified signal components reclassified to noise. The FIX-ICA denoised datasets have been updated in these cases and these subjects have been noted with QC_Issue code E. See [Subjects with Issues noted in the HCP quality control process](#).

Soon to be available:

- **Genetic data on all HCP subjects.** Genome-wide SNP genotyping data for all HCP subjects with useable blood or saliva-based genetic material will be deposited and available on NIH's dbGaP in Spring 2017. We processed samples using one custom microarray chip consisting of the Illumina Mega Chip (has 2

million multiethnic single nucleotide polymorphisms [SNPs]), immune-related content from the ImmunoArray, and psychiatric-related content from the PsychArray. Samples were also processed with a second Illumina Neuro Consortium chip to look for SNPs particularly relevant to neuroimaging studies.

- **Updated MSM-All-registered group-average rfMRI dense connectivity data.** Group-average rfMRI dense connectivity data for a 1000+ group of S1200 subjects with complete rfMRI data is planned to be available in late spring 2017.
- **Updated MSM-Sulc and MSM-All-registered group-average tfMRI data.** For all S1200 subjects with complete tfMRI data, we will release a Connectome Workbench-compatible dataset in late spring 2017 that will include group-averages for both MSM-Sulc and MSM-All registered tfMRI data.
- **Updated parcellation, timecourse, and netmap (PTN) data for all S1200 subjects with complete rfMRI data.** Updated PTN data for a 1000+ group of S1200 subjects is planned to be available in late spring 2017.
- In the meantime, the R820 dense connectome data, the R787 tfMRI data, and the PTN data based on MSM-All registered S900 data is available on the HCP 1200 Release project page. See [Group-average functional data](#) for details.

Multiple levels of data processing. The 1200 Subjects 3T MR imaging data includes “unprocessed” data (in NIFTI format), “minimally preprocessed”, and “analysis” datasets. We encourage investigators to use the preprocessed datasets, in which spatial distortions have been minimized and data have been aligned across modalities and across subjects using appropriate volume-based and surface-based registration methods. Methods used for MR preprocessing, (see [Preprocessing Pipelines](#) section below, were implemented by the HCP consortium through an intensive two-year refinement and optimization process. If your specific research interests make it important to use the unprocessed MR data, it is critical to be aware of certain characteristics that require special preprocessing (see [Gradient Nonlinearities in Unprocessed Datasets](#)).

The 3T and 7T preprocessed MR structural datasets include cortical surfaces, myelin maps, and other data that are suitable for a variety of morphometric analyses.

Denoised versions of preprocessed individual resting-state fMRI scans (by ICA-FIX) are available in two datasets for 3T: the “FIX-Denoised (Extended)” dataset that includes resting state stats results on the un-FIXed timeseries, both the 4D volume and grayordinate timeseries, the ICA components used in the denoising, and a report on the FIX run; the compact “FIX-Denoised (Compact)” dataset that includes only the grayordinate timeseries for each scan. Due to data sizes for 7T, separate FIX-Denoised datasets are packaged for grayordinate timeseries at 1.6mm (as collected at 7T) and 2mm (compatible with 3T data) mesh. Denoised 7T

volumetric timeseries data is packaged separately as the Resting State fMRI Volume FIX-Denoised package. For 7T, the Fix-Denoised (Extended) package does not contain the volume or grayordinate timeseries, but contains Resting State Stats results, FIX ICA-based intermediate outputs, and Workbench scenes showing ICA components classified as noise and signal for the REST scans.

For 3T MR, more extensively processed “**analysis**” datasets for individual tfMRI data analyzed across runs for each of 7 tasks are available for subjects with one or more complete task datasets. In addition, the HCP has more extensively processed group-average resting-state fMRI and task-fMRI data (currently on groups of S900 subjects data registered by MSM-Sulc and MSM-All) for 820 ‘related’ subjects (R820, mixture of related and unrelated subjects) for rfMRI and R787 for tfMRI available for download and as a Connectome Workbench-compatible dataset from the [ConnectomeDB 1200 Subjects project page](#). As noted above, more extensively processed group-average resting-state fMRI and task-fMRI data on a larger group of ~1000 S1200 subjects are being created and will be released in late spring/summer 2017.

The MEG data released (95 subjects) includes the raw data in 4D imaging format and the outputs of a number of data processing pipelines. The utility pipelines include quality control and artifact identification to remove “bad” channels, segments, and physiological artifacts, and an anatomy/source reconstruction pipeline to coregister the MEG to structural MRI (sMRI) and dMRI. The resting-state and task Channel-level (preprocessing) pipelines result in sets of components obtained from ICA together with the related classification into artifacts and non-artifacts, cleaned and segmented data with channel-level Event-Related fields (ERFs), Time-Frequency responses (TFRs) and power-spectra. The Source-level pipelines project the sensor maps from rMEG and tMEG channel-level processing into source space, produce rMEG ICA-based frequency specific connectomes, and produce rMEG Band Limited Power dense timeseries and connectomes (see [Overview of MEG Processing pipelines](#) below).

Due to the refinement of HCP processing pipelines over the course of the project, HCP data obtained from different releases are likely to have been processed with different versions of the pipelines. Remember to be sure to keep track of version numbers when analyzing HCP data. We do not recommend mixing data processed with different versions of the pipelines. Version numbers of the relevant HCP pipelines used to process each subject’s dataset are available in the release notes directory packaged in each dataset archive. For example, the archive 100307_3T_tfMRI_EMOTION_preproc.zip contains the file release-notes-tfMRI_EMOTION_preprocess.txt:

100307_3T_tfMRI_EMOTION_preproc.zip

Wed Apr 2 11:17:20 CDT 2014
Structural Pipeline v3.1
Functional Pipeline v3.1

Execution 1

How do I access HCP data? Many HCP imaging files are large, and they must be maintained in a well-organized directory structure in order to prevent serious confusion or corruption when analyzing the data. For the 1200 Subjects release, the total amount of imaging data added since S900 for all modalities and scans is about 25 terabytes, bringing the total amount of released HCP imaging data to ~76 terabytes.

HCP provides several ways to access part or all of these data:

(i) **“Connectome in a Box”.** Investigators wanting access to data from many or all 1200 Subjects and most or all imaging modalities are encouraged to use the Connectome in a Box option. Beginning in May 2017, we will offer hard drives containing all of the available imaging data released since the S900 release to be shipped to users for the hardware and delivery costs (see [Connectome in a Box](#) below). Institutions or imaging centers having many investigators who want access to the data should consider this option to facilitate sharing of HCP open access data (see [HCP Data Sizes](#) below); however, subject privacy provisions must be put in place in this case (see [cautionary note about data sharing](#) below).

(ii) **Download preset packages from ConnectomeDB.** The HCP provides an easy way to select standard packages based on modality of interest and a standard set of subjects. This includes selection of datasets by modality for one subject, 10, and 100 unrelated subjects. Downloading is mediated by software that enables high-speed file transfer. Download times will depend on bandwidth and overall user demand, and there is a 500 GB limit on archive size that can be requested in a session (see [How to download HCP data](#)).

(iii) **Download individual subject or user-defined group datasets from ConnectomeDB.** The ConnectomeDB dashboard user interface allows filtering of subjects based on a range of selection criteria, including scores on behavioral and individual difference measures (see [How to access HCP data](#)). The user can then download packages for just subjects selected by the filtering or by individual subjects.

(iv) **Amazon S3.** HCP data from the 1200 Subjects release will be populated in coming weeks in data storage on the cloud through Amazon S3. Software tools (e.g. FSL, Connectome Workbench, FieldTrip) for data analysis are also being added to a virtual machine on the site for on-the-cloud data analysis. This option offers a way to access and analyze HCP data without having to buy Connectome in a Box or store the data locally (Amazon Web Services does charge for computing cycles and download using Amazon tools). See [Accessing HCP Data on the Cloud](#) for details.

Why is access to some data restricted? Since the subjects include twins and their non-twin siblings, it is important for many analyses to know the family status of the subjects under study.

Family status and other potentially sensitive information are part of the Restricted Data that is available only to qualified investigators after signing the Restricted Data Use Terms (see [below](#)). **To protect the privacy of our HCP subjects, anyone wanting to use the Restricted Data MUST understand and agree to comply with the important constraints that are imposed.** To enable investigators to analyze data without being concerned about family structure issues, we have selectable groups of 10 and 100 unrelated subjects for download (see [How to download HCP data](#)).

How can I analyze and visualize HCP data? Most of the preprocessed MR data files are in standard formats supported by major neuroimaging platforms, including NIFTI (Neuroimaging Informatics Technology Initiative, see <http://nifti.nimh.nih.gov>) format for standard 3D or 4D volume data and GIFTI (Geometry Informatics Technology Initiative) format for surface data. Importantly, the preprocessing pipelines for fMRI data (resting-state and task-fMRI) also include files that are in a 'grayordinate' representation that combines cortical surface vertices and subcortical gray matter voxels into a single file using the [CIFTI data format](#) (see [Glasser et al. 2013](#)), that has recently been updated to [CIFTI-2](#). These grayordinate CIFTI files offer major advantages in terms of efficiency in handling very large and complex datasets.

The Connectome Workbench platform can be used to visualize the outputs of HCP preprocessing and additional processing pipelines, including CIFTI, NIFTI and GIFTI files. Connectome Workbench is freely available (current version available at <http://humanconnectome.org/connectome/get-connectome-workbench.html>). A tutorial and associated dataset is available (requires ConnectomeDB login and signing of data use terms). Connectome Workbench also includes many command-line options (wb_command) that support a variety of analysis steps, including processing of CIFTI files.

What's in store for future releases? Although the S1200 release is the final release of new HCP young adult subjects, we are planning to continue to release more data on these subjects, including more extensively processed imaging data.

Tractography for dMRI. We plan to run all HCP 3T diffusion [BEDPOSTX](#) fiber orientation analyzed data through FSL's [PROBTRACKX](#) for probabilistic tractography estimation. This algorithm repetitively samples from the distributions of voxel-wise principal diffusion directions produced by BEDPOSTX, each time computing a streamline through these local samples to generate a probabilistic streamline or a sample from the distribution on the location of the true streamline. By taking many such samples, PROBTRACKX can generate connectivity distributions from user-specified seed voxels or surface vertices (or both).

Individual-Subject Parcellations. HCP's high quality multimodal data was used to subdivide the brain into neurobiologically accurate and functionally meaningful areas generating an improved parcellation of cerebral cortex, the 'HCP_MMP1.0' multimodal cortical parcellation, with 180 areas in each hemisphere and a striking degree of left-right symmetry ([Glasser et al.](#)

[2016](#)). The parcellation was initially generated by a semi-automated approach using data averaged across 210 HCP subjects. Importantly, an automated areal classifier algorithm enables accurate cortical parcellation in individual subjects, even those with atypical topological arrangement of specific areas or who were not used to make the original map. We plan to apply this classifier algorithm to all HCP subjects and release the resultant individual parcellations later in 2017.

How were the HCP datasets acquired and processed? Besides the documentation provided below, there are published descriptions of various aspects of the HCP ([Van Essen et al., 2012](#); [Marcus et al., 2011](#)). Detailed descriptions of many HCP methods are available in eight papers ([Van Essen et al., 2013](#), [Ügurbil et al., 2013](#), [Glasser et al., 2013](#), [Smith et al., 2013](#), [Barch et al., 2013](#), [Sotiropoulos et al., 2013](#), [Marcus et al., 2013](#), and [Larson-Prior et al., 2013](#)) published in a special issue of NeuroImage in 2013. See <http://humanconnectome.org/about/consortium-publications.php> and <http://www.humanconnectome.org/documentation/citations.html> for updated lists of HCP publications.

What if I want to adapt HCP protocols to my own scanner and research projects? [Appendix I: Protocol Guidance and HCP Session Protocols](#) provides some advice and suggestions on this issue, which depends on the particular scanner system you are using as well as your research objectives. Adapting the HCP protocol to your particular scanner and project is also discussed in the Nature Neuroscience article “The Human Connectome Project’s neuroimaging approach” ([Glasser et al. 2016](#)).

Open Access and Restricted Data

The HCP provides imaging, behavioral, and demographic data from a large population of healthy adults. This poses special challenges for protecting the privacy of participants, especially because it is a family study including twins and their siblings. Unless these data are properly managed, there is a risk that some participants might be recognizable to family members and others (restricted data). In addition, some of the data elements (sensitive data) collected might harm or embarrass participants if they were to be inadvertently identified.

To protect the privacy of our participants, the HCP has implemented a two-tiered plan for data sharing, with different provisions for handling Open Access data and Restricted Data.

Open access data (all imaging data and most of the behavioral data) is available to those who [register](#) and agree to the [Open Access Data Use Terms](#). This includes agreement to comply with institutional rules and regulations (see below), **which may require you to get your IRB’s approval to use the data.**

Restricted data elements that could be potentially used to identify subjects include family structure (twin or non-twin status and number of siblings); birth order; age by year; handedness;

ethnicity and race; body height, weight, and BMI; and a number of other categories (see [Behavioral and individual difference measures details](#) below). Restricted data also includes sensitive data (also called tier 2 restricted data) that could harm or embarrass participants if they were to be inadvertently identified, such as psychiatric information from the SSAGA telephone diagnostic interview, drug test results, and physiological data. We have built the distinction between the restricted and sensitive classes of restricted data elements into the user interface of the [ConnectomeDB dashboard](#) used to explore and filter on the behavioral data (see below).

Each qualified investigator wanting to use restricted data must [apply for access](#) and agree to the [Restricted Data Use Terms](#). These terms explain how restricted data may and may not be used and shared, and they reiterate the need for compliance with institutional requirements. **They include major limitations on how Restricted Data can be incorporated into publications and public presentations.** Click [here](#) for additional information about required investigator qualifications and exemplar use case scenarios.

You must comply with your institutional rules and regulations regarding research on human subjects. Different institutions operate under different national and state laws. Before you use any HCP data (either *Open Access* or *Restricted Data*) for research, you should ask your institution's Institutional Review Board (IRB), Ethics Committee, or office that oversees research on human subjects to determine whether you need institutional approval in order to use the data. On request, HCP can provide a certificate stating that you have accepted the Data Use Terms (Open Access or Restricted Access, as appropriate).

What about technical support, bug reports, and feature requests?

We anticipate a wide range of questions, suggestions, and discussion points as HCP data and software become freely available to the community. Users are strongly encouraged to join the HCP Users mailing list (hcp-users@humanconnectome.org) by signing up at <http://www.humanconnectome.org/contact/> or by checking the appropriate box when registering to download HCP data.

Contributions to the hcp-users mailing list will be monitored and responded to by investigators and staff in the WU-Minn HCP consortium. Often this will entail prompt responses to answer questions or suggest solutions to technical problems. As with mailing lists for other brain mapping platforms (e.g., FSL, FreeSurfer), investigators outside the HCP consortium are encouraged to respond as well. Bug reports and feature requests will be entered into the issue tracking system used by HCP software developers by trained HCP staff.

HCP Data Release Updates Wiki and Data Patches

As mentioned above, a [wiki page](#) has been created to document Known Bugs/Missing files and Planned fixes in the HCP data released to date. This wiki will be updated frequently as we discover new bugs and files missing in the datasets. We recommend checking the wiki page before you do any major analyses with HCP data so that you are aware of any issues with files you may want to use. On the wiki, we also describe our plans to fix bugs/redistribute missing files and provide links to patches the HCP is releasing periodically to replace the files affected. For some fixes, HCP patch scripts are written to automatically search through your downloaded or Connectome in a Box directory structure for files that need to be fixed and replaces them without modifying the structure. But remember, the 1200 Subjects release data (and S900 and S500 release data) currently available for download/order was processed with a different version of the HCP pipelines and should not be mixed with the Q1-Q3 Release data, even if patched.

Gradient nonlinearities in unprocessed datasets

If you intend to download the unprocessed NIFTI data, read on! Unusually large gradient nonlinearities are present in the raw images for ALL modalities, and it is important to correct for the spatial distortions they cause.

All HCP imaging data were acquired on a Siemens Skyra 3T scanner with a customized SC72 gradient insert that greatly improves the quality of diffusion imaging scans (the 'Connectome Skyra'). Higher performing gradients require compromises in bore diameter and gradient nonlinearities. Further, in custom-fitting the higher performing gradient set into a standard clinical system, technical limitations prevent centering of the subjects' heads in the bore isocenter. Consequently, the gradient nonlinearities associated with all Connectome Skyra scans exceed those of a conventional clinical 3T scanner. In the HCP processed datasets for all scan modalities (structural, fMRI, and dMRI), these distortions have been corrected for by spatially warping the images using gradient field information specific to the Connectome Skyra. The gradient unwarping code is available at <https://github.com/ksubramz/gradunwarp/blob/master/Readme.md> (Jovicich *et al.*, 2006). The gradient field nonlinearity coefficients for the Connectome Skyra are considered by Siemens to be proprietary information. To request access to these coefficients, please contact your Siemens collaboration manager or email Dingxin Wang at dingxin.wang@siemens.com.

For diffusion MRI, the gradient nonlinearities also cause voxel-by-voxel changes in the strength and orientation of the diffusion encoding gradients. Consequently, the effective b-values and b-vectors in all the primary data that you can download exhibit small variations from voxel to voxel. When analyzing the primary (unprocessed) datasets or the minimally preprocessed dMRI data you may use the code provided in [Appendix 2: Matlab code for voxel-wise correction of dMRI gradients](#) in your fitting routine in order to account for gradient spatial nonlinearities during fiber orientation estimation.

Chapter 1: Getting HCP Data and using ConnectomeDB

HCP Data Sizes

The MRI images collected for each HCP subject are notably high in spatial resolution and (for fMRI and MEG) in temporal resolution as well. Coupled with the long scan duration for each modality and some of the processing results we are making available, the resultant image data files and datasets are very large. This “big data” is generally a good thing, enabling researchers to conduct types of data analyses that have never before been possible. However, the sheer size of the imaging data generated by the HCP protocol requires some special considerations:

If you choose to obtain your data via download, prepare for long download times. The HCP using an optimized download client (Aspera) to make downloading as fast as possible, even when multiple users are downloading at the same time. However, depending on the amount of data you are downloading and your internet bandwidth, you may experience lengthy download times.

Housing the data will require a large amount of disk space. Even data on a few subjects (total data/subject: ~10 GB unprocessed, ~40 GB preprocessed, which includes ICA-FIX denoised rfMRI and MSM-All registered versions of all modalities) could fill up your hard drive quickly! Consider just how much data you need to “try out” the data, or do the analysis you want to do, and download just what you need. If you want all the data, consider ordering “[Connectome in a Box](#)” ([see below](#)) or accessing data through Amazon S3.

Consider organizing investigators at your institution to obtain a single copy of the HCP data for local distribution. In many cases, several groups at a research institution will want access to HCP data. We encourage users to organize efforts to obtain a local archive of the data via joint ordering of Connectome in a Box (recommended) or a single download that can be copied and used by many within your institution. With a modest amount of planning, groups that chose to do this will enjoy faster distribution and save valuable research time.

Important cautionary note about distributing HCP data: If you are going to distribute HCP data in any way, remember that all data users, even those who access HCP data through a local archive and not through ConnectomeDB, must [register](#) and sign the [Open Access Data Use Terms](#) on the HCP site before HCP data is used for research purposes. This step is required to be in compliance with regulations on human subjects research (other approval procedures may also be required by your institution – see [above](#)). Those who distribute HCP data or use HCP data without signing the Data Use Terms are at risk for violation of their institution’s human research policies and possible legal repercussions.

With these realities in mind, the HCP distributes image data in three ways:

1. Via download through

ConnectomeDB. HCP has made convenient data packages in 1 subject, 10, or 100 unrelated subject groups (by modality type), or all MEG Subjects to allow the user to “try out” the data without incurring a large download or data storage burden. The subjects in these set groups are unrelated to facilitate its use to investigators who want to analyze data without being concerned about family structure issues. The ConnectomeDB interface also allows the user to create their own subject groups of interest and download only the modalities necessary for analysis of that group.

2. “Connectome in a Box” (CinaB).

Users can choose a complete 1200 Subjects Release dataset, a starter dataset (~100 unrelated subjects), or MEG only dataset preloaded onto 8TB hard drives, that can be ordered and shipped to you for the cost of the hard drives + shipping. S1200 Connectome in a Box can be ordered now and should be available for shipment by late spring 2017.

3. Amazon S3. Complete 1200 Release data will be accessible on the cloud in the same organization as Connectome in a Box by late spring 2017. See [Accessing HCP Data on the Cloud](#) for details.

In addition to the download and Connectome in a Box option, because of their large size, we provide links to ~33 GB group-average functional connectivity

| HCP Data Sizes (per Subject) | | |
|--|------------------------|----------------|
| Session | Format | .zip File Size |
| Structural | Unprocessed | 71 MB |
| | Preprocessed | 1.1 GB |
| Resting State fMRI (all 4 runs) | Unprocessed | 4 GB |
| | Preprocessed | 11.6 GB |
| | FIX (compact) | 3.9 GB |
| | FIX_extended | 7.7 GB |
| Task fMRI (avg per Task) All 7 Tasks | Unprocessed | 490 MB |
| | Preprocessed | 1.9 GB |
| | Analyzed | 400 MB |
| | Unprocessed | 3.4 GB |
| | Preprocessed | 13.1 GB |
| | Analyzed | 2.8 GB |
| Diffusion | Unprocessed | 2.6 GB |
| | Preprocessed | 1.2 GB |
| | BedpostX | 1.3 GB |
| S900 Group-Average on 881 Subjects | Additionally Processed | 1.7 GB |
| Group-Average “dense” connectomes (each of 2) | Additionally Processed | 33 GB |
| MEG | Unprocessed | 9 GB |
| | Preprocessed | 14 GB |
| | Source-level | 35 GB |
| Total MEG (95 Subjects) | Unprocessed | 780 GB |
| | Preprocessed | 1.1 TB |
| | Both + Source | 4.7 TB |
| Total 3T MR (per Subject) | Unprocessed | 10 GB |
| | Preprocessed | 40 GB |
| | Both | 50 GB |
| | Both+Analyzed | 53 GB |
| Total 7T MR (per Subject) | Unprocessed | 12 GB |
| | Preprocessed | 42 GB |
| | Both+Analyzed | 57 GB |
| Total (All HCP MR scan datasets from 1113 Subjects) | Unprocessed | 15.2 TB |
| | Preprocessed | 57.6 TB |
| | Both+Analyzed | 76.0 TB |

data for visualization through the latest version of Connectome Workbench (currently available for R820 group, soon to be available for 1000+ group). The group average data with associated Workbench files, includes a scenes file that has links to the full-correlation R820 fcMRI dense connectomes. An associated Connectome Workbench Tutorial for the group average data is packaged with the Group-Average data available for download on the ConnectomeDB splash page.

How to access HCP data

Go to the HCP data home page:

<http://humanconnectome.org/data/>

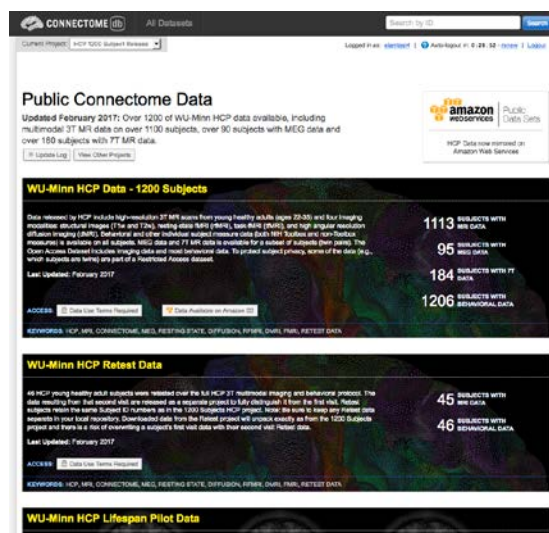
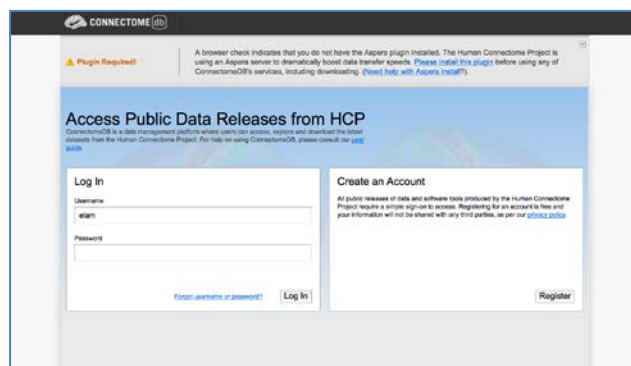
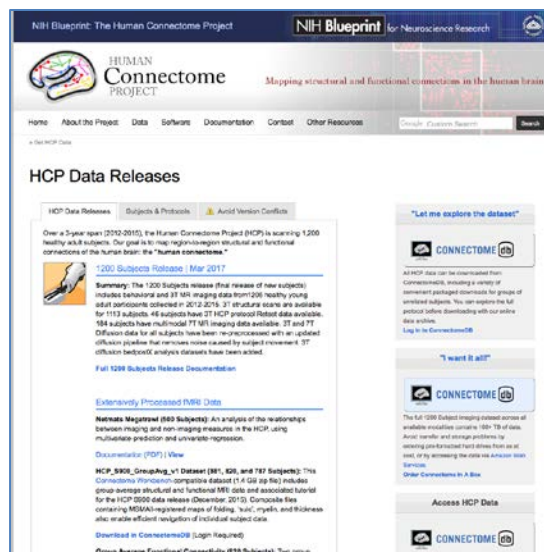
Here, you can get an overview of the 1200 Subjects release, explore links to details of the HCP protocols, and access more useful resources for using HCP data.

To the right are links to:

- Explore and/or download data in our database platform, ConnectomeDB.
- Order all the data via the “Connectome in a Box” option.
- Read more about the HCP Data Use Terms.

Clicking on the ConnectomeDB option takes you to a dialog to register for an account. When you register, you will be sent an email to activate your account within 24 hours.

After you have verified your account and are logged in, you will launch the Public Connectome Data page, with data on multiple connectome projects, including HCP. If you see a “Data Use Terms Required” button below the WU-Minn HCP, click that button to agree to the



HCP Open Access Data Use Terms (required for access).

Once the terms are accepted, buttons and selectors will appear to the right that you will launch the HCP allow you to explore and download HCP data by selecting options for groups of subjects (see [How to explore HCP data in ConnectomeDB](#) and [How to download HCP data](#) below).

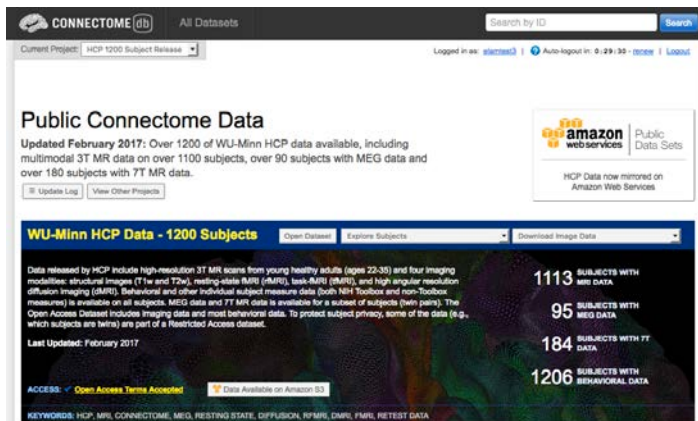
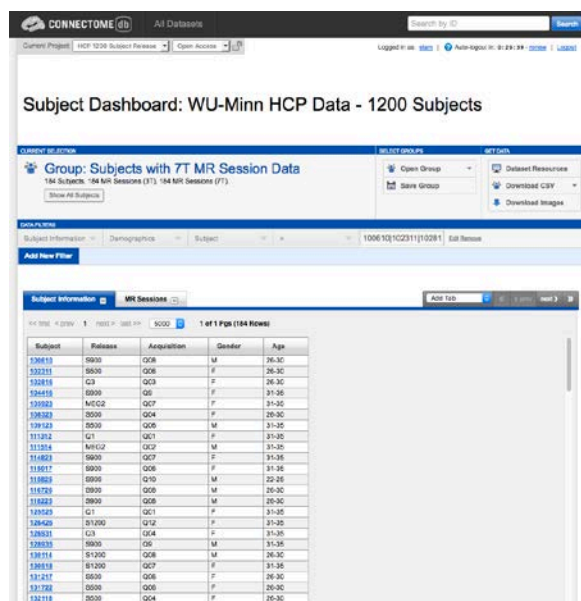
Note: Your login to ConnectomeDB times out every 30 minutes. Click the “renew” link to reset the time. (Auto-logout resets ConnectomeDB to the HCP Data Release page, not the page you were on at logout).

How to explore HCP data in ConnectomeDB

Once your ConnectomeDB account is set up and you have signed the HCP data use agreement, the header for the 1200 Subjects release will appear blue and on the right a “Open Dataset” button, “Explore Subjects” and “Download Image Data” pulldowns will be active.

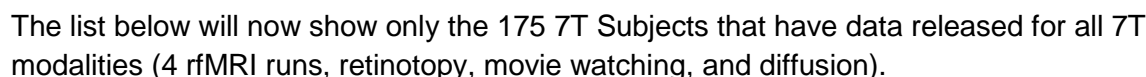
- “Open Dataset” takes you to the HCP 1200 Subjects project page, which has links to documentation processing pipeline software, group average data, stimulus resources for fMRI, the Netmats Megatrawl dataset and Connectome Workbench datasets.
- “Explore Subjects” takes you to the ConnectomeDB subject Dashboard for viewing behavioral and individual difference measures for groups of subjects or exploring the scan level data for individual subjects. You may choose to browse all subjects or groups of subjects, including all 7T, Retest, or MEG subjects.
- “Download Image Data” takes you straight to the Download Packages page to queue downloads for the group of subjects that you select.

Click on “Explore Subjects” and select “Subjects

The “Add Tab” pulldown at the far right allows selection of categories of individual/behavioral data for viewing. Click on “Study Completion” in the pulldown to open a tab that lists which scan and behavioral data in the HCP protocol were collected for each subject. Use the scroll bar at the bottom to find completeness measures for 7T, starting with 7T_Full_MR_Compl.

Select Category “Study Completion”,
Assessment “Study Completion: 7T MR”,
Attribute “Full Protocol Complete”,
Operator “=”, Value “True”. Click “Apply”.



If you click the “Download Images” or “Download CSV” buttons, only the subjects for which you filtered will be part of the data queued for download (see below).

You can also explore all the scan details and behavioral/individual difference data for any subject individually by clicking a Subject ID link from the list.

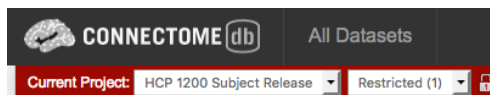
Now click the “Dataset Resources” button at the top right. This takes you to the WU-Minn HCP Dataset page (also accessible through the “Open Dataset” button on the Public Connectome Data splash page that you see when you first log in).

Here you see:

- Overview description of the 1200 Subjects dataset

- Buttons to Browse all Subjects, Open a set Group of Subjects in the dashboard (including any a user saves themselves), or Download Image data for the same set groups of subjects.
- Links to Data Reference documents, online FAQs and individual difference/behavioral Data Dictionary.
- Downloads of individual difference/behavioral data and expanded FreeSurfer anatomical measures
- Links to HCP data processing software for MEG and MRI
- Link to Order the complete dataset via “Connectome in a Box”
- Downloads of 3T, 7T, and MEG tfMRI protocol E-Prime scripts
- Links to Group Average rfMRI and tfMRI datasets
- Link to the Netmats MegaTrawl analysis of correlations between imaging and non-imaging measures in the HCP data
- Datasets designed for viewing in ConnectomeWorkbench

If you have Restricted data access, you will see at the top left of the page is a pulldown to the right “Current Project” that allows you to choose the level of data



you want to view: Open Access, Restricted (contains fields that are potentially identifying), or Sensitive (contains fields potentially harmful or embarrassing to subjects). This pulldown persists throughout ConnectomeDB and, therefore, can be set from the Dashboard where data fields are viewed as well. More information on Access Levels is available in [Open Access and Restricted Data](#).

The default setting is Open Access. Restricted data view contains restricted data (Dashboard column headers in red) and all open access data. Sensitive data view contains sensitive data (Dashboard column headers in black) and restricted and open access data. When you select Restricted or Sensitive the area around the Current Project box is colored red to remind you that you are set to view restricted data.

The HCP built this access level setting into the user interface to allow investigators to control what kinds of data appear for different use case scenarios. For example, one can show only Open Access data when giving a lecture to a class, but view Sensitive data when filtering subjects by positive drug test results for a research study on drug use.

How to download HCP data

In order to facilitate access to the very large data files generated on every subject (see [HCP Data Sizes](#) above), we have prepackaged data into convenient group data packages. Datasets can be downloaded from ConnectomeDB in set packages or users may order all the image data to be sent to them on hard drives (Connectome in a Box, [see below](#)).

Due to the large file sizes of the data, the HCP uses an Aspera server to dramatically boost data transfer speeds. Therefore, downloading data from the HCP requires you to have the Aspera plugin installed on your browser.

If you have not installed the plugin yet, a warning message will appear at the top of the browser window reminding you that you must [install the Aspera plugin](#) before using any of ConnectomeDB's services, including downloading (see the second panel on page 21). Note: Installing the Aspera plugin to your browser requires you to restart your browser, therefore, we recommend installing the plugin *before* you explore/select HCP data in ConnectomeDB.

An [Aspera install help guide](#) is available on the HCP website. Note: One common problem with installing Aspera on the Mac platform for non-US users is that you must add US English to the Languages list (using the checkbox) to your System Preferences>Personal>Language & Text settings.

Upon login to ConnectomeDB, you are brought to the Public Connectome Data splash page. If you choose one of the “Download Image Data” options (e.g. 7T Subjects) in the pulldown at the top right of the WU-Minn HCP Data-1200 Subjects section, you will launch the Download Packages page.

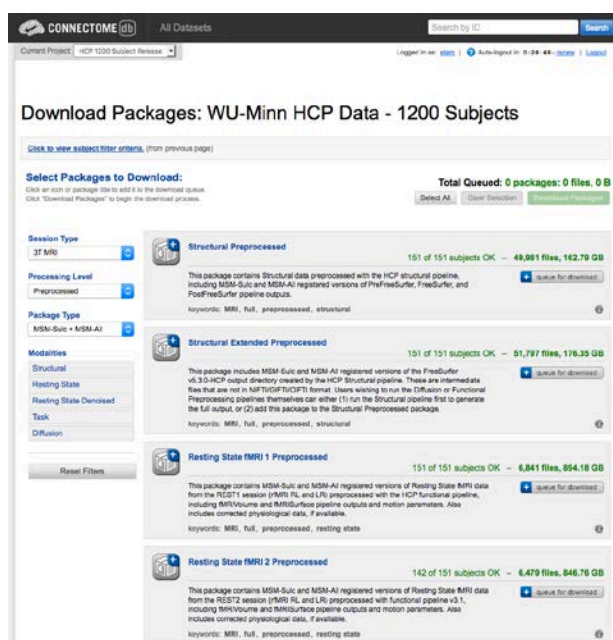
On the left, choose the:

Session Type (3T MRI, 7T MRI or MEG),

Processing Level (unprocessed, preprocessed [includes channel-level processed MEG], Analysis [fMRI only], and Source-level processed [MEG only])

and, for Preprocessed data only,

Package Type (MSM-Sulc + MSM-All or S500 to S900 Extension Only [MR Only]) of the data you wish to download.






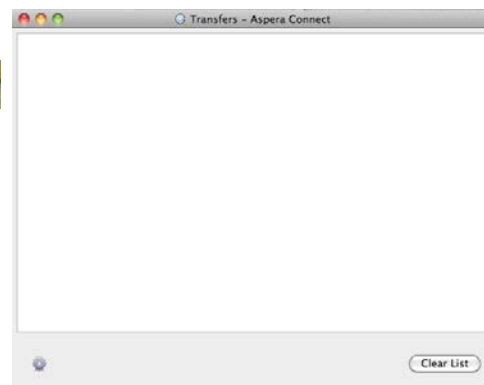
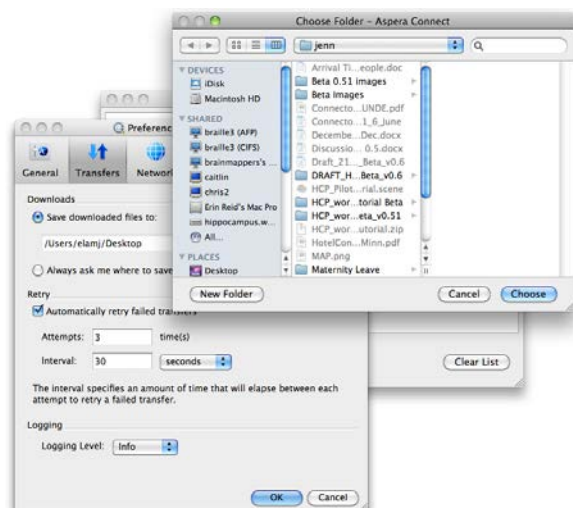
First time Aspera Download Setup

If this is your first time downloading data from ConnectomeDB, **before you start selecting packages to download**, we highly recommend setting up the preferences in your Aspera Connect plugin to download to the appropriate place on your local or network hard drive.

Note: the default Aspera Connect download location is your Desktop, which is likely not where you want your HCP data to go! (Unfortunately HCP has no control over the Aspera defaults).

To set your Aspera Connect preferences:

- Search your computer for Aspera Connect using the Spotlight or Search functions in your operating system. Click on “Aspera Connect” to open the application.
- Open the Aspera Connect: Transfers window:
 - On Mac, in the Menu bar click on Window > Transfers, or use the shortcut **⌘ + T** (control + T)
 - On PC, in the Notification area at the very  bottom right of your Desktop, click the button, then click the Aspera Connect icon: 
 - On Linux, extract the .gzip Aspera Connect plugin file you downloaded. Run the extracted script in a terminal window to launch Aspera Connect. The Aspera Connect icon should appear in your program notification area (often at the top or bottom right of your desktop), click on this icon and choose Transfers in the dropdown.
- Click the gear icon  at the bottom left of the Aspera Connect — Transfers window to launch the Aspera preferences window.
- Click on the “Transfers” tab to setup where in your hard drive or network you would like your HCP data to download.
- Now, each time you download HCP data it will go to the directory you selected (unless you change the preference again).



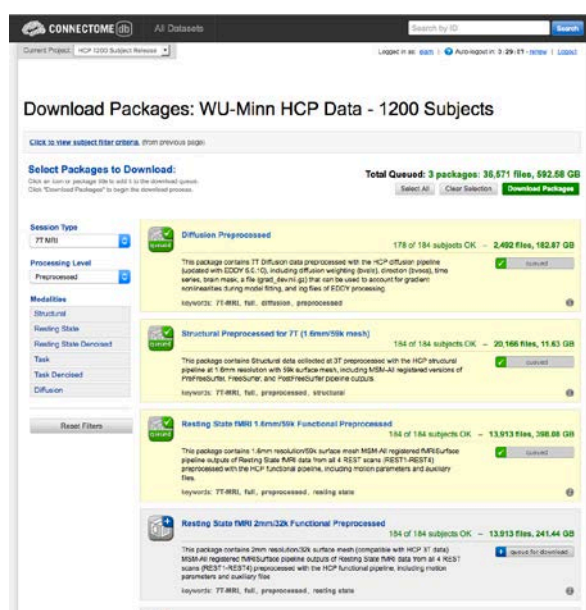
Downloading HCP data packages

Now that you have Aspera setup to download to the correct location, go back to the “Downloading Packages” page in your browser.

Here, as noted above, you can filter the image data available for download by session type (3T MRI, 7T MRI or MEG), processing level and modality by making selections on the left. To make selections for download, click the “add to queue” icons to the left of each dataset package.

The total size of the data you have queued for download is at the upper and bottom right. Use this calculation to check against your available hard drive space.

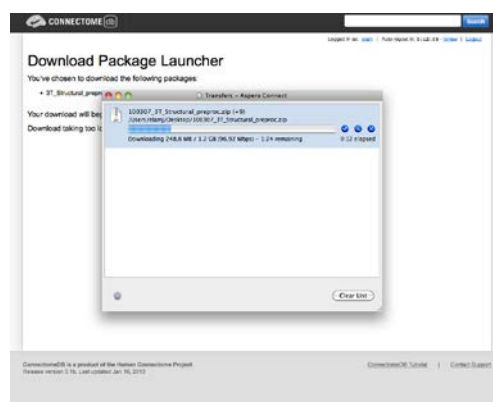
When you are done making your selections, click the Download Packages button at the upper or lower right. A popup will remind you again what packages you selected and the total size of the files. Clicking “Download Now” will automatically open Aspera Connect and launch the download.



Note: If you have not setup Aspera Connect, as described [above](#), by default the data will start downloading to your Desktop (!!). If this not where you want the data to go, we recommend stopping the download by clicking the “X” button to the right of the download progress bar. Click the gear icon at the bottom left of the Aspera Connect: Transfers window to launch the Aspera preferences window (see above). Once you set the preferred download location, you will need to relaunch your download in ConnectomeDB, using your browser’s back button to return to the “Downloading Packages” page.

If the download does not start automatically, use the link on the Download Package Launcher page in your browser to restart. Once started, the blue bar in the Aspera Connect: Transfers window will show your download progress (per subject file, not overall progress).

Of course, the more subjects and packages you selected for download, the longer your download will take (remember, this is big data!). This is why we offer the [“Connectome in a Box”](#) and [Amazon S3](#) options for



those who want data on many subjects (see below).

Handling downloaded HCP data

Now that you have downloaded the data you want, here's some tips for handling the data:

- Each download package consists of a .zip archive, and an md5 checksum. After you download the data you want, [you can use the md5 file to verify the integrity of your downloaded file](#).
- To unzip the downloaded .zip archives, for Windows users we recommend using a utility such as [7-zip](#) (available free) so that the directory structure of the unzipped files remains intact. The built-in unzip capabilities of Linux and MacOSX can be used, as they do not affect the directory structure of the unzipped files.
- Once the archives are downloaded and unzipped, many of the component files are gzipped (.gz files). You will need to unzip all .gz files in order to use them. For this, you need an application that is compatible with gzip. For Windows users, we again recommend 7-zip. Linux has support for gzip built in, and Mac users can use the Mac Gzip utility.
- Downloaded HCP data unzips to a set directory structure that is detailed below in [Directory structure for unprocessed MR data](#), [Directory structure for preprocessed MR data](#), and in [Appendix 3: File Names and Directory Structure for 1200 Subjects Data](#). The directory structure for downloaded HCP data is identical to what is provided in Connectome in a Box (see below).
- If you have downloaded a large amount of HCP data, we recommend sharing the data locally at your research institution. This will save download and organization time for your colleagues and relieve some of the burden on the ConnectomeDB download bandwidth available to other HCP users. Again, if you distribute HCP data in any way, **all users** (those who access HCP data) must [register](#) and sign the [Open Access Data Use Terms](#) in order to comply with regulations on human subjects research (see [above](#)).

Connectome in a Box

An attractive option for getting HCP data is “Connectome in a Box”. This option allows users to order hard drives containing the complete imaging data from chosen HCP data releases “at cost” (approximately \$200/100 Subjects of data, plus shipping costs) and have them shipped to their address.

Because of expanding datasizes and the fact that many groups have already purchased S900 Connectome in a Box, as of the 1200 Subjects Release we are planning to offer:

- S900 Connectome in a Box on six 8TB hard drives
- A 900 Subjects “upgrade”, including 200+ new S1200 subjects, plus updated diffusion MPP, bedpost, and other extension data for S900 subjects
- All released 3T structural + 7T MR data for all 7T subjects
- All released MR + MEG data for all MEG subjects on two 8TB hard drives

All versions of Connectome in a Box can be securely ordered from the [HCP website](http://www.humanconnectome.org/).



Connectome in a Box data is loaded onto Linux-formatted SATA hard drives.

Once the user (or group of users) receives the drives, we recommend that the data be transferred to a network storage unit and made available to all on the user's local network (provided **each user** [registers](#) and signs the [Open Access Data Use Terms](#), see [above](#)).

We recommend signing up for the HCP-Users mailing list at <http://www.humanconnectome.org/contact/> to be apprised of data updates as they are released.

Accessing HCP Data on the Cloud

The HCP is sharing all released MR and MEG data (but not the Restricted Data or any behavioral data) as a participant in the Amazon Web Services (AWS) Public Data Sets program (<http://aws.amazon.com/publicdatasets/>). Via this program, HCP data is hosted on the cloud in an Amazon Simple Storage Service (S3) bucket.

This option allows users to access the full HCP image dataset on the cloud without needing to perform large downloads. The cloud-based platform also enables users who do not currently have the local computational infrastructure to perform large-scale computations on the data through the Elastic Compute Cloud service.

Because of the time it takes to upload the amount of data we are sharing, we expect the 1200 Subjects Release to be available on AWS S3 by mid-late spring 2017.

Details on how to obtain access to HCP data through Amazon S3 and guidance for working with cloud-hosted data, will be available soon on the [HCP website](#).

Chapter 2: HCP Hardware and Protocols

Standard Operating Procedures (SOPs)

A major effort has been made to establish consistent procedures for all aspects of data acquisition and data processing. These are described in a set of Standard Operating Procedures (SOPs) that are included in [Appendix 4](#). These SOPs provide a useful reference for investigators wanting to know more about exactly what was done. Any outdated versions of SOPs are retained and are available on request.

Standard schedule for subject visits

Most subjects complete the full 3T HCP protocol during a two-day visit. If a subject is one of the MEG protocol subjects, the full protocol is typically completed in a three-day visit. Typical 2-day and 3-day schedules are shown below. In the subset of participants who receive MEG sessions, a MEG day will be scheduled before HCP MRI Day 1 because the superconducting sensors in the MEG scanner can be affected by normally imperceptible residual magnetization from the MR scanner's static field. In many cases, the participant's schedule necessitates condensing the MEG, behavioral, and MRI sessions into 2 long acquisition days, but MEG acquisitions always precede MRI sessions.

Breaks are included in the schedule to provide the researchers the opportunity to adjust experimental equipment, and the participant a chance to stretch and relax their eyes, as well as relieve any stiffness or pressure from lying so completely still in the scanner.

Schedule for 7T Subjects at UMinn is coming soon in an S1200 Reference Manual update.

MEG Subjects WU-Minn HCP Schedule

| MEG Day | |
|---------|---|
| 7:30 | Set-up and MEG scanner QC |
| 8:30 | Subject consent at WU and travel to SLU |
| | SLU Subject consent and Task practice |
| 9:30 | Empty Room and Participant Noise Scan |
| | Head Surface Digitization |
| 10:30 | rMEG scans (25 min) |
| | tMEG WM scans (25 min) |
| 11:30 | tMEG Story/Math scans (20 min) |
| | Break and application of EMG electrodes |
| 12:30 | tMEG Motor scans (30 min) |
| | Satisfaction survey |
| 1:30 | Return to WU |

Typical 2-day WU-Minn HCP Subject Schedule

| Day 1 | Day 2 |
|-------|---------------------------------------|
| 7:30 | Set-up and scanner QC |
| 8:30 | Consent and Intake tests* |
| | Mock scanner practice |
| 9:30 | MR check and Diffusion session |
| 10:30 | NIH Toolbox Behavioral Tests |
| 11:30 | Lunch break |
| | Task practice and MR check |
| 12:30 | R-fMRI session #2 (30 min) |
| | T-fMRI session #2 (30 min) |
| 1:30 | Drug/alcohol tests, tobacco/alcohol |
| 2:30 | 7d retrospective, satisfaction survey |
| | Recognition task (ex-scanner) |

*Drug/alcohol tests, surveys: mini mental state, sleep quality, menstrual info, HbA1c, TSH, hematocrit, blood for genotyping.

3T MR scanner and other hardware

3T Scanner hardware. All HCP subjects are scanned on a customized Siemens 3T “Connectome Skyra” housed at Washington University in St. Louis, using a standard 32-channel Siemens receive head coil and a “body” transmission coil designed by Siemens specifically for the smaller space available using the special gradients of the WU-Minn and MGH-UCLA Connectome scanners.

The scanner has a customized SC72 gradient insert and a customized body transmitter coil with 56 cm bore size (*diffusion*: $G_{\max} = 100$ mT/m, max slew rate = 91 mT/m/ms; *readout/imaging*: $G_{\max} = 42$ mT/m, max slew rate = 200 mT/m/ms). The HCP Skyra has the standard set of Siemens shim coils (up to 2nd order).

Relative to a standard commercial Skyra, the customized hardware includes a gradient coil and gradient power amplifiers that together increase the maximum gradient strength from 40 mT/m to 100 mT/m on the WU-Minn 3T. This specifically benefits diffusion imaging, and on theoretical grounds ([Ügurbil et al. 2013](#)) it should provide significant gains over the standard 40 mT/m though not as much as the 300 mT/m customized gradients used by the MGH/UCLA HCP consortium.

Head motion and physiological monitoring. To address head motion, in most scan sessions we acquire dynamic head position information using an optical motion tracking camera system (Moire Phase Tracker, KinetiCor). This system monitors head position precisely and in real-time using an infrared camera mounted in the scanner bore. Images of Moire interference fringes on a target affixed by clay to the bridge of the subject’s nose are streamed in real time to a computer that displays the current position of the sensor and stores the positional information in a data file linked to the associated MRI scan. The stored file of head position and head movement can be used for post-hoc analyses. We also use it as a feedback trigger in dMRI scans to interrupt the movie being viewed whenever suprathreshold displacement and/or rapid head movement occur. Positional information can also be routed to the MRI scanner computer and can in principle be used prospectively to update the MRI slice prescription in real time ([Zaitsev et al., 2006](#)). However, prospective motion correction is not part of our 3 T HCP acquisition protocol because the technology became available only late in the HCP method development phase and was not sufficiently tested and developed before the data collection protocol was finalized ([Van Essen et al., 2013](#)).

We also acquire cardiac and respiratory signals associated with each scan, using a standard Siemens pulse oximeter placed on a digit and a respiratory belt placed on the abdomen. These signals are linked to scan onset using a trigger pulse generated by the pulse sequence. They are written to text files and assigned a unique file name that enables matching to the corresponding scan.

3T Visual projection and E-Prime computer. Visual stimuli are presented and participant responses are collected using a Dell Optiplex 790 computer, running an Intel Core i3-2100 with 8GB of RAM and 64-bit Windows 7 Enterprise SP1. The E-Prime version was E-Prime 2.0 Professional Production Release (2.0.10.242). Visual stimuli are projected with a NEC V260X projector onto a lucite screen at 1024x768 resolution, and viewed by the participant using a mirror mounted on the top of the head coil. Participant responses are registered on a customized fiber-optic button box.

Summary of 3T MR imaging protocols

Structural, fMRI, and dMRI acquisitions were collected over 4 total imaging sessions, each approximately 1 hour in duration. Resting-state and task-fMRI data were collected in two sessions. Each session consisted of two resting-state acquisitions of approximately 15 minutes each, followed by task-fMRI acquisitions of varying durations (see below).

Vitamin E capsule on right side. A capsule of vitamin E was taped to the subject's right temple in every scan session, to enable definitive determination of the right side in the image data.

The following provides basic parameters for the main scan types in each session, and pertinent details about each session. A more complete set of imaging parameters can be found in the protocol exports from the scanner, available in [Appendix 1](#). FOV positioning in all runs was handled in an automated manner using Siemens AutoAlign feature.

Structural session

| Type | Series Description | Description | TR (ms) | TE (ms) | TI (ms) | Flip Angle | FOV (mm) | Voxel Size | BW (Hz/Px) | iPAT | Acquisition Time (min:sec) |
|------------|--------------------|-------------|---------|---------|---------|------------|----------|------------------|------------|------|----------------------------|
| T1w | T1w_MPR1 | 3D MPRAGE | 2400 | 2.14 | 1000 | 8 deg | 224x224 | 0.7 mm isotropic | 210 | 2 | 7:40 |
| T2w | T2w_SPC1 | 3D T2-SPACE | 3200 | 565 | | variable | 224x224 | 0.7 mm isotropic | 744 | 2 | 8:24 |

Resting-state fMRI (rfMRI)

rfMRI data were acquired in four runs of approximately 15 minutes each, two runs in one session and two in another session, with eyes open with relaxed fixation on a projected bright cross-hair on a dark background (and presented in a darkened room). Within each session, oblique axial acquisitions alternated between phase encoding in a right-to-left (RL) direction in one run and phase encoding in a left-to-right (LR) direction in the other run.

Resting state images were collected with the following parameters:

| Parameter | Value |
|------------------|--|
| Sequence | Gradient-echo EPI |
| TR | 720 ms |
| TE | 33.1 ms |
| flip angle | 52 deg |
| FOV | 208x180 mm (RO x PE) |
| Matrix | 104x90 (RO x PE) |
| Slice thickness | 2.0 mm; 72 slices; 2.0 mm isotropic voxels |
| Multiband factor | 8 |
| Echo spacing | 0.58 ms |
| BW | 2290 Hz/Px |

| Condition | Runs | Frames per run | Run Duration (min:sec) |
|----------------------|------|----------------|------------------------|
| REST (Resting-state) | 4 | 1200 | 14:33 |

Task-evoked fMRI (tfMRI)

Following completion of rfMRI in each of the two fMRI scanning sessions, subjects were asked to complete tasks that were designed to activate a variety of cortical and subcortical networks, details in [Task fMRI scripts and protocol details](#). The following table provides a listing of the fMRI scans collected. For each scan type, one run was acquired with right-to-left phase encoding, and a second run with left-to-right phase encoding (in-plane FOV [field of view] rotation obtained by inverting both the RO (readout) and PE [phase encoding] gradient polarity).

tfMRI data were acquired with the same EPI pulse sequence parameters as R-fMRI, except for the run duration information listed below. There are seven tasks (14 tfMRI runs) totaling one hour of total tfMRI scan time, with 3 tasks collected in one session and the remaining 4 tasks collected in another session.

| Task | Runs | Frames per run | Run Duration (min:sec) |
|------------------|------|----------------|------------------------|
| Working Memory | 2 | 405 | 5:01 |
| Gambling | 2 | 253 | 3:12 |
| Motor | 2 | 284 | 3:34 |
| Language | 2 | 316 | 3:57 |
| Social Cognition | 2 | 274 | 3:27 |

| | | | |
|-----------------------|---|-----|------|
| Relational Processing | 2 | 232 | 2:56 |
| Emotion Processing | 2 | 176 | 2:16 |

Scans necessary for field mapping correction were also acquired in each of the fMRI scan sessions.

EV (Explanatory Variable) files included in the dataset provide a set of conditions (and their associated timing) that can be used in the analysis of each task. For each scan, an EPRIME TAB.txt (tab-delimited text spreadsheet) describes task parameters that can be used in various fMRI analysis packages; see [Task fMRI files and protocol details](#).

Diffusion imaging (dMRI)

| Parameter | Value |
|-----------------------|---|
| Sequence | Spin-echo EPI |
| TR | 5520 ms |
| TE | 89.5 ms |
| flip angle | 78 deg |
| refocusing flip angle | 160 deg |
| FOV | 210x180 (RO x PE) |
| matrix | 168x144 (RO x PE) |
| slice thickness | 1.25 mm, 111 slices, 1.25 mm isotropic voxels |
| Multiband factor | 3 |
| Echo spacing | 0.78 ms |
| BW | 1488 Hz/Px |
| Phase partial Fourier | 6/8 |
| b-values | 1000, 2000, and 3000 s/mm ² |

Other details: Diffusion gradients are monopolar. Oblique axial acquisitions alternate between right-to-left and left-to-right phase encoding directions in consecutive runs. Image reconstruction uses SENSE1 multi-channel ([Sotiropoulos et al., 2013](#)).

A full dMRI session includes 6 runs (each approximately 9 minutes and 50 seconds), representing 3 different gradient tables, with each table acquired once with right-to-left and left-to-right phase encoding polarities, respectively. Each gradient table includes approximately 90 diffusion weighting directions plus 6 b=0 acquisitions interspersed throughout each run. Diffusion weighting consisted of 3 shells of b=1000, 2000, and 3000 s/mm² interspersed with an approximately equal number of acquisitions on each shell within each run. The diffusion directions were obtained using a toolbox available from Emmanuel Caruyer that returns uniformly distributed directions in multiple q-space shells. The directions are optimized so that

every subset of the first M directions is also isotropic. References and the toolbox can be found at: <http://www.emmanuelcaruyer.com/q-space-sampling.php>.

Full scanning protocols (PDF)

Complete scanning protocols for each imaging modality can be found in [Appendix 1](#).

Standard 3T MR session structure

Each subject was scanned by MRI in four regular sessions (~4 hours total), with the following shorthand labels:

Day 1:

- Session 1: Structural scans (two T1w and two T2w scans)
- Session 2: Two 15-min rfMRI scans (RL and LR phase encoding) and three tfMRI tasks (one RL and one LR scan for each task)

Day 2:

- Session 3: dMRI scan
- Session 4: Two 15-min rfMRI scans (RL and LR phase encoding) and the four remaining tfMRI tasks (one RL and one LR scan for each task)

Note: Counterbalancing the ordering of the different phase-encoding acquisitions for the resting-state fMRI scans (RL followed by LR in the first fMRI session; LR followed by RL in the second session) was adopted on October 1, 2012. Prior to that, both sets of rfMRI scans were acquired using the RL followed by LR order. The temporal order of actual scan acquisitions for each subject is reflected in the scan number in ConnectomeDB (and also as an explicit ScanOrder variable).

In some cases, scans were acquired in an extra session (coded session 5 or higher, as appropriate). This was done if the data quality was inadequate in the regular scan or if technical problems prevented a full set of acquisitions in the regular session.

MR Physiological Data

We also acquire cardiac and respiratory signals associated with each functional MR scan, using a standard Siemens pulse oximeter placed on a digit and a respiratory belt placed on the abdomen. These signals are linked to scan onset using a trigger pulse generated by the pulse sequence. They are written to text files and assigned unique file names that enable matching to the corresponding scan. These raw files were processed to create a file format expected by FSL's PNM. They are combined into a 3-column physio text file starting from the onset of the

first trigger pulse through to the end of the last acquired frame. They are all sampled equally at 400 Hz, meaning there are roughly 288 samples per frame for functional images. In the aggregate file the first column is used for the trigger pulse information, the second for pulse oximeter (i.e., cardiac) information, and the third column for respiratory information.

The physiology datasets are included for most of the functional datasets. In some collected functional sessions for some subjects, physiological data was not collected or is otherwise unrecoverable.

Before the HCP 900 subjects data release, we identified previously unknown timing errors in the in the processed physiological log files (e.g., `rfMRI_REST1_RL_Physio_log.txt`). These errors resulted in variable offsets in timing, sometimes amounting to hundreds of milliseconds. Therefore, we recommend that the physio log files in all released datasets prior to the S900 release, not be used.

As of the S900 Release (December 2015), we corrected the known timing issues for all scans, including pulse oximetry data that was collected during diffusion MRI scans and included the corrected files in the S900 data packages and S900+ Connectome in a Box. Newly released physio files distributed as of the S1200 release are all corrected and can be used.

Ongoing HCP analyses will compare resting-state and task-fMRI data with vs. without regression of physiological signals. If warranted by these analyses, additional data files reflecting such corrective steps may be included in future data releases.

MR Image Reconstruction and DICOM to NIFTI conversion

Image reconstruction of multiband images was implemented in ICE (the image reconstruction environment of the Siemens scanners) using a customized set of image reconstruction algorithms (see [Ügurbil et al. 2013](#)). While a fMRI or dMRI scan is ongoing, data are streamed to an offline (“remote”) reconstruction computer for unaliasing of the multiband data, which streams the processed data back to the scanner for final processing and viewing. In some cases of technical problems, reconstruction was done at a later time using a “retro” reconstruction.

In the spring of 2013, several months after the commencement of HCP Phase II scanning in August 2012, an improved reconstruction algorithm was developed that reduced spatial blurring in fMRI and dMRI acquisitions. This improvement was deemed worthy of upgrading the reconstruction algorithm (from version r177 to r227), which occurred in late April 2013. In brief, the original reconstruction algorithm (version r177) performed the separation of the multi-band multi-slice images after transforming the acquired fully sampled data to frequency space along the read-out direction, and now the multi-band multi-slice separation is performed in k-space.

The new reconstruction method (version r227) was applied retrospectively to dMRI scans acquired prior to the April 2013 transition that were saved in a compact but unreconstructed data format. For fMRI scans prior to April 2013, unreconstructed scans were not preserved owing to technical considerations related to the time required to complete reconstruction at the scanner. However, concurrent with the change in reconstruction algorithm in April 2013, the reconstruction switched to using graphics-processing-unit (GPU) accelerated reconstruction, which reduces reconstruction times considerably, thus permitting saving of unreconstructed fMRI data from April 2013 onward.

Image reconstruction algorithm version information by subject, for both dMRI and fMRI, is noted in the Study Completion category in the ConnectomeDB dashboard (measures dMRI_3T_ReconVrs and fMRI_3T_ReconVrs). For users wanting to use only data with the same reconstruction algorithm applied in their analyses, one can now filter by reconstruction version (r177 or r227) in the Study Completion category on the ConnectomeDB dashboard (see [How to explore HCP data in ConnectomeDB](#)).

The reconstructed DICOM files are uploaded to the IntraDB database. Image defacing is carried out ([Milchenko and Marcus, 2013](#)) on all DICOM files containing potentially recognizable facial features. This includes structural (T1w, T2w), Bias receive and Bias transmit scans.

Conversion of DICOM files to NIFTI format was carried out using the dcm2nii utility. This utility is a component of the MRICron suite of tools developed by Chris Rorden: (http://www.nitrc.org/frs/?group_id=152).

Initial Structural MR QC and File Transfer to ConnectomeDB

Structural scans (Image QC) and Structural pipeline QC. An experienced rater has evaluated each structural scan and designated it as excellent, good, fair, poor, or unusable based on criteria related to tissue contrast, blurriness, or banding artifacts. To be included in the data release and in the structural pipeline processing, at least one T1w and one T2w rated as good or excellent must have been acquired in the same session (without the subject having exited the scanner). If these criteria are not met using the scans from the initial structural session, an extra scan session is acquired during the subject's visit or (if necessary) during a subsequent visit.

Images are also reviewed by a radiologist for overall scan quality and for neuroanatomical anomalies. If a subject is found to have abnormal, clinically significant findings, the subject's data is not released. If findings in a subject are benign, normal variations, the findings are noted (see [Subjects with Issues noted in the HCP quality control process](#) below), but the data is processed for release.

We also evaluate the segmentation, surface, and map outputs of the HCP Structural pipeline for errors. Details of Image, Radiological, and Structural pipeline QC are available in [Appendix 4. HCP Protocol Standard Operating Procedures \(SOPs\)](#) pp. 17-29.

Usability of other scans. Scans from other modalities (fMRI and Diffusion) are included in the data release only if they meet specified criteria that make them useable in conjunction with other scans of the same modality acquired from the subject. Many aspects of MR QC are described in [Marcus *et al.*, 2013](#).

All scans are permanently stored in an internal database (IntraDB). Scans that are considered usable for preprocessing are transferred to the public-facing ConnectomeDB based on information encoded in a set of scan-specific fields in IntraDB. This process also entails a reorganization of data from session-specific directories (reflecting the fact that data are acquired in multiple scan sessions) to a composite subject-specific directory in ConnectomeDB, with subdirectories appropriate for the scan modality. Files of a given modality are grouped with ancillary files (e.g., bias fields and field maps) from the same session as needed in order to facilitate preprocessing using standardized scripts and pipelines. Many file names are modified and standardized as part of this transfer and reorganization process, again to facilitate preprocessing.

A subject-specific spreadsheet is included with the structural download packages, containing information that may be useful when analyzing the data. [Appendix 5: Data Acquisition Information for an Exemplar Subject](#) shows a representative example.

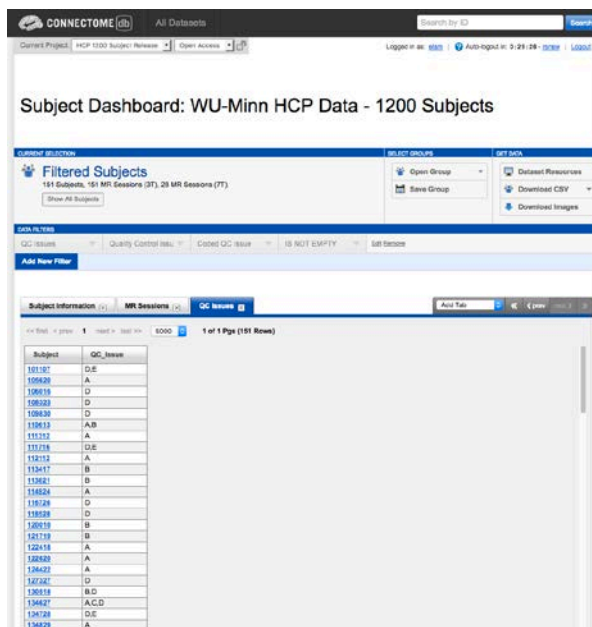
Subjects with Issues noted in the HCP quality control process

As of the S1200 release, a subject data measure (QC_Issue) has been added to ConnectomeDB to flag subjects with notable issues found in the HCP Quality Control process as described for each issue code below. The issues are notable, but were not considered severe enough to exclude the subject's imaging data from release. We are sharing these findings of issues to alert users and allow them to decide whether to include or exclude these subjects in their own analyses.

The QC_Issue measure allows users to filter the data for subjects with or without these issues in the ConnectomeDB Subject Dashboard or in the CSV spreadsheet containing individual subject data that is downloadable from the site.

To view the issue codes in the ConnectomeDB Subject Dashboard, click the “Add Tab” dropdown at the top right of the section listing subjects and choose “QC Issues”. To filter on subjects with issue codes, click the “Add New Filter” button, select the “QC Issues” category, “Quality Control Issue Noted” assessment, “Coded QC Issue” attribute, and “IS NOT EMPTY” operator (see image). At the time of the S1200 Release, there are 157 subjects with imaging data available that have one or more QC_issue codes.

For some issues (e.g., scan and surface QC issues), more subject specific information, including affected run information for issue D, is available on our [QC Issues wiki page](#).



The screenshot shows the ConnectomeDB Subject Dashboard for WU-Minn HCP Data - 1200 Subjects. The 'Filtered Subjects' section shows 151 Subjects, 151 MR Sessions (37), 25 MR Sessions (77). The 'QC Issues' section shows a table with columns 'Subject' and 'QC_Issue'.

| Subject | QC_Issue |
|---------|----------|
| 101102 | D |
| 105620 | A |
| 106016 | D |
| 106020 | D |
| 106030 | D |
| 106153 | A,B |
| 111212 | A |
| 111218 | D,E |
| 112113 | A |
| 113417 | B |
| 116821 | B |
| 116824 | A |
| 118128 | D |
| 118128 | D |
| 120519 | B |
| 121718 | B |
| 122418 | A |
| 126620 | A |
| 126621 | A |
| 127247 | D |
| 128219 | B,D |
| 134627 | A,C,D |
| 134728 | D,E |
| 134828 | A |

Note on the HCP QC process:

We explicitly reviewed the outputs of the HCP Structural Pipelines on all subjects and ascertained that all included subjects had at least reasonable (or better) cortical surface models. (Examples of some of the issue categories are on the QC Issues wiki page). However, a comprehensive QC of all modalities was not feasible, and HCP does not claim to have identified all QC issues for all HCP subjects and modalities. In particular, in the interest of providing data, fMRI and dMRI scans were only very rarely excluded for motion, and we have opted to release fMRI and dMRI scans collected in periods of known (albeit intermittent) problems with the head coil, which resulted in some temporal instabilities in acquisition.

Therefore, the absence of an issues code for a given subject does not imply that there are no issues with the data for this subject. We encourage users to alert us to issues that they discover. We will then update the QC_Issues field and wiki appropriately for the benefit of other users.

Issue Code A: Anatomical anomalies

All HCP subject structural scans and FreeSurfer segmentation and surface outputs from the structural pipeline were subject to a standard quality control process that involved manual viewing and rating of scan quality and anatomical abnormalities by an experienced rater (see [Appendix 4: Post Session 1 Quality Control SOP](#), pp. 17-24). Brain anomalies evident in T1w and T2w scans were noted and further reviewed by a radiologist. Subjects with major radiologic anomalies that were expected to substantially impact brain connectivity were removed from the study, and their imaging and behavioral data are not included in the released data.

Some subjects had focal anomalies that are considered as normal variants or benign findings. We have released their data, but because of their altered anatomy, using data from these subjects may affect some analyses. These subjects are flagged with the issue code **A** in the QC_Issue measure. Additional specifics on the location and type of anomaly for each of these subjects is available for reference on the [QC Issues wiki page](#).

Issue Code B: Segmentation and/or surface errors

As noted above, all brain surfaces and brain structure segmentation results generated by FreeSurfer as part of the HCP structural pipeline were inspected for any notable issues (see [Appendix 4: Post Session 1 Quality Control SOP](#), pp. 25-28). If there were major segmentation or surface errors, we attempted to fix the issues in an automated way by rerunning the Structural Pipelines using a different FreeSurfer ‘seed’. Due to the cohort size of the HCP and limited staff resources, manual editing of segmentations and surfaces was not attempted on any subjects.

Subjects with focal segmentation and surface errors in their structural pipeline outputs are flagged with the issue code **B** for the QC_Issue measure. Additionally, more details on the location and kind of error for each of these subjects are available for reference on the [QC Issues wiki page](#).

Issue Code C: Some of subject’s data acquired during a period of instabilities in the head coil

Data for some subjects was collected during discrete periods in which there were intermittent problems with the temporal stability of the head coil. Subjects collected during these periods of temporal instability are flagged with the issue code **C** for the QC_Issue measure and a list of these subjects is also available on the [QC Issues wiki page](#).

Since spatially localized temporal instability is particularly problematic for analysis of resting state data, we evaluated network matrices (“netmats”) derived from both the “minimally preprocessed” (MPP) data and the FIX-ICA cleaned data, for all rfMRI runs of all subjects scanned during these periods. As expected, many of the netmats from the MPP rfMRI data acquired during these periods contained severe artifacts. Indeed, these periods are a primary contributor to the subjects flagged with Issue Code D (see below). However, the resting-state netmats following the FIX cleaning process were reasonable in all cases. Most of the time, FIX automatically removed the temporal instability, as intended. However, in a small proportion of scans manual reclassification of some of the FIX-ICA components from ‘signal’ to ‘noise’ was necessary to appropriately remove artifact from the FIX-denoised rfMRI scans. (Subjects for which manual reclassification was necessary in one or more rfMRI scans are indicated via Issue Code E, see below).

The situation is less clear for the tfMRI and dMRI data of these ‘temporal instability’ subjects. We have not explicitly evaluated the tfMRI and dMRI data of these subjects, but both are very likely to contain artifacts arising from the temporal instability problems with the head coil. Currently, FIX-denoising has not been applied to the tfMRI data. It is possible that the impact of the temporal instability may be ameliorated to some degree in the released HCP Task Analyses by the model-constrained nature of those analyses. (However, any analysis of the MPP tfMRI data from those subjects using a “resting-state” like approach is almost certainly problematic). As for the dMRI data, it is possible that the “outlier correction” feature in ‘eddy’ will yield appropriately cleaned data for subjects scanned during these periods, but this remains to be investigated.

In summary, (1) the FIX-cleaned rfMRI scans from these subjects is fine to use; (2) the MPP rfMRI data from most of these subjects is particularly problematic, and is provided mostly with the intent that it will serve as a useful test-bed for testing denoising algorithms; and (3) use caution and extra review if you decide to use the tfMRI and dMRI data from these subjects.

Issue Code D: rfMRI scans with prominent artifact in MPP data (but cleaned appropriately by FIX)

An evaluation of “grayplots” (Power 2016) and/or network matrices generated from each subject’s minimally preprocessed rfMRI scan data revealed subjects across the entire HCP study for whom one or more rfMRI scan showed prominent artifacts in their “minimally preprocessed” (MPP) data, due either to coil problems (the majority of cases), or excessive subject motion. Such subjects are flagged with Issue Code **D** and a spreadsheet listing the specific rfMRI scans exhibiting these artifacts is available on the [QC Issues wiki page](#). We believe this process identified most of the rfMRI scans with the most severe artifacts in their MPP data; nonetheless, it should not be considered a complete, comprehensive list of all of the rfMRI scans with severe artifacts. The list of scans with prominent artifacts will hopefully provide a useful test-bed for evaluating denoising algorithms.

As expected, nearly all of the subjects flagged with Issue Code C are also flagged with Issue Code D. However, Issue Code D includes additional subjects for whom prominent artifacts were empirically observed in their MPP rfMRI data, outside of known windows of head coil instability problems.

As noted above (for Issue Code C) we evaluated network matrices (“netmats”) derived from both the MPP and the FIX-ICA cleaned data, for all rfMRI runs associated with Issue Code D. Again, the resting-state netmats following the FIX cleaning process were reasonable in all cases. (As above, if necessary, manual reclassification of some FIX components was performed to ensure appropriate denoising).

In summary, (1) the FIX-cleaned rfMRI scans from subjects with Issue Code D is fine to use; and (2) the MPP rfMRI data from these scans is particularly problematic, and is provided mostly with the intent that it will serve as a useful test-bed for evaluating denoising algorithms.

Issue Code E: Manual reclassification of some FIX-ICA components

As noted above for Issue Codes C and D, in a small number of instances the automatic labeling of ICA components as either ‘signal’ or ‘noise’ by FIX was incorrect, and as a result, the initial denoising by the automatic FIX-ICA process was insufficient, and prominent artifacts remained in the rfMRI scans after FIX. (This probably occurred because the set of scans used to train FIX early in the HCP project did not include some of the more rare/unique types of artifacts that we encountered over the course of the study). For these rfMRI scans, we manually reclassified one or more ‘signal’ components as ‘noise’ components and ran the ReApplyFix pipeline to appropriately clean these data. Note that for the manually reclassified datasets, we used the original MSMAll registration rather than re-running MSMAll after the reclassification. We do not expect a substantial difference in registration because MSMAll relies on a multi-variate approach to resting state fMRI modeling (weighted regression) that is more robust than the univariate approaches (e.g. correlation) typically used to generate functional connectivity maps. In addition, for these subjects, the ReApplyFIX (for both MSM-Sulc and MSM-All registrations) and RestingStateStats pipelines were run on the Washington University Center for High Performance Computer (WU-CHPC) Cluster #2, rather than the WU-CHPC Cluster #1, on which all other 3T processing pipelines were run for all S1200 subjects. We note this in full disclosure of how these data were processed differently, but we do not expect that the difference in processing environments will have an effect of the rfMRI FIX results.

Subjects for whom one or more rfMRI scans had some of their FIX components manually reclassified are flagged with the Issue Code **E**. A list of the specific scans affected is available on the [QC Issues wiki page](#).

Note that the final FIX-cleaned data from rfMRI scans with Issue Code E is perfectly valid to include in analyses. We provide this Issue Code simply so that interested users can know which rfMRI scans required some manual intervention in the FIX-cleaning process.

As expected, all of the subjects with Issue Code E are also flagged with Issue Code D.

3T Task fMRI scripts and protocol details

Task-Evoked Functional Brain Activity

Our primary goals in including task-evoked functional MRI (tfMRI) in the HCP are to: 1) help identify as many “nodes” as possible that can guide, validate and interpret the results of the connectivity analyses that will be conducted on resting state fMRI (rfMRI), resting state MEG (rMEG) and diffusion data; 2) to allow a comparison of network connectivity in a task context to connectivity results generated using rfMRI; and 3) to relate signatures of activation magnitude or location in key network nodes to individual differences in performance, psychometric measures, or other phenotypic traits. To accomplish these goals, we developed a battery of tasks that can identify node locations in as wide a range of neural systems as is feasible within realistic time constraints (see [Barch et al. 2013](#) for more detail).

We assess seven major domains that sample the diversity of neural systems that will be of interest to a wide range of individuals in the field, including: 1) visual, motion, somatosensory, and motor systems; 2) category specific representations; 3) working memory/cognitive control systems; 4) language processing (semantic and phonological processing); 5) social cognition (Theory of Mind); 6) relational processing; and 7) emotion processing. These tasks are described in more detail below and in [Barch et al. 2013](#). Stimuli were projected onto a computer screen behind the subject's head within the imaging chamber. The screen was viewed by a mirror positioned approximately 8 cm above the subject's face.

tfMRI scripts and data files

Script files are run in E-Prime 2.0 Professional to present task fMRI stimuli and collect behavioral responses in the scanner. If you would like to run HCP tasks in your own research project, or examine the stimuli used in HCP tasks, the task stimulus script archives can be obtained through the “3T E-Prime fMRI Scripts” link under “Quick Downloads” on the WU-Minn HCP Data Project page (click “Open Dataset” on the ConnectomeDB splash page, or “Dataset Resources” from the Subject Dashboard).

Tab-delimited versions of the E-Prime data files (TAB.txt) are included in the released data. TAB.txt files are named according to the task condition that they describe and are contained within the directories for each of the two runs within each task (each phase encoding direction). A brief description of the key variables in those files can be found in [Appendix 6: Task fMRI and tMEG E-Prime Key Variables](#). The original edat files will not be available, because they may contain identifying information.

In addition, we include EV .txt files derived from the TAB.txt files (defined above) in the released data. EV files are explanatory variables (predictors) in FSL format (3-columns: onset, duration,

and amplitude). There is a separate EV directory for each of the two runs within each task. Examples of the EV files for each task are detailed below.

The preprocessed data also includes .fsf files for each task. The fsf file is the setup or configuration file for running GLM-based fMRI analyses in 'FEAT' (FMRIB's Expert Analysis Tool: <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FEAT>). The Lev1 .fsf files contain setup information necessary to run GLM analyses on the timeseries data for an individual scan run. Lev2 .fsf files contain setup information to run GLM analyses combining multiple scan runs for an individual participant. Lev3 .fsf files (not included in the release) can be created to setup GLM analyses across multiple participants.

HCP's results of individual (within-subject, level 2) tfMRI analysis for each task are available in download packages separate from the unprocessed and preprocessed datasets. These tfMRI datasets include a design.fsf file showing the setup information that was used in the FEAT GLM analysis for each task. See [Individual \(within-subject\) tfMRI analysis results data](#) and [Appendix 3D](#) for more details.

Here are examples of the EV files for each task and phase encoding direction in the appropriate unprocessed {Subject_ID}/3T/tfMRI_[task_phaseencodingdirection]/LINKED_DATA/EPRIME directory (e.g. 100307/3T/tfMRI_WM_LR/LINKED_DATA/EPRIME) and preprocessed {Subject_ID}/MNINONLinear/Results/tfMRI_[task_phaseencodingdirection]/EVs directory (e.g. 100307/MNINONLinear/Results tfMRI_WM_LR/EVs):

Working Memory

| | | |
|---------------------|---------|--|
| EVs/0bk_body.txt | BLOCKED | onset of 0Back body block condition |
| EVs/0bk_faces.txt | BLOCKED | onset of 0Back faces block condition |
| EVs/0bk_places.txt | BLOCKED | onset of 0Back places block condition |
| EVs/0bk_tools.txt | BLOCKED | onset of 0Back tools block condition |
| EVs/2bk_body.txt | BLOCKED | onset of 2Back body block condition |
| EVs/2bk_faces.txt | BLOCKED | onset of 2Back faces block condition |
| EVs/2bk_places.txt | BLOCKED | onset of 2Back places block condition |
| EVs/2bk_tools.txt | BLOCKED | onset of 2Back tools block condition |
| EVs /0bk_cor.txt | EVENT | onset of correct trials in 0Back blocks |
| EVs /0bk_err.txt | EVENT | onset of error trials in 0Back blocks |
| EVs /0bk_nlr.txt | EVENT | onset of trials in 0Back blocks with no response |
| EVs /2bk_cor.txt | EVENT | onset of correct trials in 2Back blocks |
| EVs /2bk_err.txt | EVENT | onset of error trials in 2Back blocks |
| EVs /2bk_nlr.txt | EVENT | onset of trials in 0Back blocks with no response |
| EVs /all_bk_cor.txt | EVENT | onset of correct trials in 0- and 2Back blocks |



| | | |
|--------------------|-------|---|
| EVs/all_bk_err.txt | EVENT | onset of error trials in both 0- and 2Back blocks |
|--------------------|-------|---|

Gambling

| | | |
|---------------------|---------|-------------------------------|
| EVs/win.txt | BLOCKED | Onset of mostly reward blocks |
| EVs /loss.txt | BLOCKED | Onset of mostly loss blocks |
| EVs /win_event.txt | EVENT | Onset of reward trials |
| EVs /loss_event.txt | EVENT | Onset of loss trials |
| EVs /neutral.txt | EVENT | Onset of neutral trials |

Motor

| | | |
|--------------|---------|----------------------------|
| EVs /cue.txt | BLOCKED | Onset of task cues |
| EVs /lf.txt | BLOCKED | Onset of left foot blocks |
| EVs /rf.txt | BLOCKED | Onset of right foot blocks |
| EVs /lh.txt | BLOCKED | Onset of left hand blocks |
| EVs /rh.txt | BLOCKED | Onset of right hand blocks |
| EVs /t.txt | BLOCKED | Onset of tongue blocks |

Language

| | | |
|----------------|---------|-----------------------|
| EVs /story.txt | BLOCKED | Onset of story blocks |
| EVs /math.txt | BLOCKED | Onset of math blocks |

Social Cognition

| | | |
|----------------------|---------|---|
| EVs /mental.txt | BLOCKED | Onset of mental interaction blocks |
| EVs /rnd.txt | BLOCKED | Onset of random interaction blocks |
| EVs /mental_resp.txt | EVENT | Onset of trials rated as mental interaction |
| EVs /other_resp.txt | EVENT | Onset of trials not rated as mental interaction |

Relational Processing

| | | |
|-------------------|---------|----------------------------|
| EVs /relation.txt | BLOCKED | Onset of relational blocks |
| EVs /match.txt | BLOCKED | Onset of match blocks |

Emotion Processing

| | | |
|---------------|---------|--------------------------------|
| EVs /fear.txt | BLOCKED | Onset of emotional face blocks |
| EVs /neut.txt | BLOCKED | Onset of shape blocks |

Details of fMRI tasks

Working Memory

The category specific representation task and the working memory task are combined into a single task paradigm. Participants were presented with blocks of trials that consisted of pictures of places, tools, faces and body parts (non-mutilated parts of bodies with no “nudity”). Within each run, the 4 different stimulus types were presented in separate blocks. Also, within each run, ½ of the blocks use a 2-back working memory task and ½ use a 0-back working memory task (as a working memory comparison). A 2.5 second cue indicates the task type (and target for 0-back) at the start of the block. Each of the two runs contains 8 task blocks (10 trials of 2.5 seconds each, for 25 seconds) and 4 fixation blocks (15 seconds). On each trial, the stimulus is presented for 2 seconds, followed by a 500 ms inter-task interval (ITI).

| Conditions (Blocked) | |
|----------------------|-------------------|
| 0-back faces | 2-back faces |
| 0-back places | 2-back places |
| 0-back tools | 2-back tools |
| 0-back body parts | 2-back body parts |

| Conditions (Event-Related) | |
|----------------------------|---------------------------|
| 0-back correct trials | 2-back correct trials |
| 0-back error trials | 2-back error trials |
| 0-back no response trials | 2-back no response trials |

Additional Contrasts. These event types can be combined to create two categories of contrasts.

| Working Memory Contrasts |
|---|
| 0-back contrast (activity combined across conditions 1-4) |
| 2-back contrast (activity combined across conditions 5-8) |
| 2-back versus 0-back contrast (2-back contrast minus 0-back contrast) |

| Category Contrasts |
|--------------------|
|--------------------|

| | |
|-----------------|------------------------------------|
| Faces contrast | (0-back faces plus 2-back faces) |
| Places contrast | (0-back places plus 2-back places) |
| Tools contrast | (0-back tools plus 2-back tools) |
| Body contrast | (0-back body plus 2-back body) |

Potential Additional Event Related Contrasts: Researchers can also use the TAB.txt E-Prime data files to generate the following potential event-related contrasts:

1. Targets
 - a. For 2-back tasks, targets are 2-back repeats
 - b. For 0-back tasks, targets match the cue stimulus
2. Non-targets
 - a. For 2-back tasks, non-targets are novel items
 - b. For 0-back tasks, non-targets do not match the cue stimulus
3. Lures
 - a. For 2-back tasks, lures are 1-back or 3-back repeats
 - b. For 0-back tasks, lures are repeated stimuli that do not match the cue stimulus

Recognition Memory

After participants exit the scanner from the session that includes the working memory tasks, they are given a Remember, Know, New item recognition test for the faces and places that were presented during the working memory task. Responses to this recognition memory test can be used to create events to analyze the working memory trials as a function of whether the item was subsequently recognized (remember or know) or not (new). This is referred to as a subsequent memory analysis. The REC_run1_TAB.txt and REC_run2_TAB.txt files needed for this analysis are in the */fMRI_WM_LR/ and */fMRI_WM_RL/ data directories.

References for Working Memory: Localizer (Drobyshevsky *et al.* 2006); Reliable across subjects (Drobyshevsky *et al.* 2006) and time (Caceres *et al.* 2009).

References for Category-Specific Representations: Faces, Places, Tools and Body Parts: Localizer (Downing *et al.* 2001; Peelen and Downing 2005; Taylor *et al.* 2007; Fox *et al.* 2009); Reliable across subjects (Downing *et al.* 2001; Fox *et al.* 2009) and time (Peelen and Downing 2005; Kung *et al.* 2007).

Gambling

This task was adapted from the one developed by Delgado and Fiez (Delgado *et al.* 2000). Participants play a card guessing game where they are asked to guess the number on a mystery card (represented by a "?") in order to win or lose money. Participants are told that potential card numbers range from 1-9 and to indicate if they think the mystery card number is

more or less than 5 by pressing one of two buttons on the response box. Feedback is the number on the card (generated by the program as a function of whether the trial was a reward, loss or neutral trial) and either: 1) a green up arrow with “\$1” for reward trials, 2) a red down arrow next to -\$0.50 for loss trials; or 3) the number 5 and a gray double headed arrow for neutral trials. The “?” is presented for up to 1500 ms (if the participant responds before 1500 ms, a fixation cross is displayed for the remaining time), following by feedback for 1000 ms. There is a 1000 ms ITI with a “+” presented on the screen. The task is presented in blocks of 8 trials that are either mostly reward (6 reward trials pseudo randomly interleaved with either 1 neutral and 1 loss trial, 2 neutral trials, or 2 loss trials) or mostly loss (6 loss trials pseudo-randomly interleaved with either 1 neutral and 1 reward trial, 2 neutral trials, or 2 reward trials). In each of the two runs, there are 2 mostly reward and 2 mostly loss blocks, interleaved with 4 fixation blocks (15 seconds each).

Conditions (Blocked)

Mostly reward blocks

Mostly loss blocks

Conditions (Event-Related)

Reward trials

Loss trials

Neutral trials

References for Gambling Task: Reliable across subjects and robust activation in fMRI (Delgado *et al.* 2000; May *et al.* 2004; Tricomi *et al.* 2004; Forbes *et al.* 2009)

Motor

This task was adapted from the one developed by Buckner and colleagues (Buckner *et al.* 2011; Yeo *et al.* 2011). Participants are presented with visual cues that ask them to either tap their left or right fingers, or squeeze their left or right toes, or move their tongue to map motor areas. Each block of a movement type lasted 12 seconds (10 movements), and is preceded by a 3 second cue. In each of the two runs, there are 13 blocks, with 2 of tongue movements, 4 of hand movements (2 right and 2 left), and 4 of foot movements (2 right and 2 left). In addition, there are 3 15-second fixation blocks per run. This task contains the following events, each of which is computed against the fixation baseline.

Conditions (Blocked)

Left finger blocks

Right finger blocks

Left toe blocks

Right toe blocks

Tongue movement

References for Motor Task: Localizer (Morioka *et al.* 1995; Bizzi *et al.* 2008; Buckner *et al.* 2011; Yeo *et al.* 2011).

Language Processing

This task was developed by Binder and colleagues (Binder *et al.* 2011) and uses the E-prime scripts provided by these investigators. The task consists of two runs that each interleave 4 blocks of a story task and 4 blocks of a math task. The lengths of the blocks vary (average of approximately 30 seconds), but the task was designed so that the math task blocks match the length of the story task blocks, with some additional math trials at the end of the task to complete the 3.8 minute run as needed. The story blocks present participants with brief auditory stories (5-9 sentences) adapted from Aesop's fables, followed by a 2-alternative forced-choice question that asks participants about the topic of the story. The example provided in the original Binder paper (p. 1466) is "*For example, after a story about an eagle that saves a man who had done him a favor, participants were asked, "Was that about revenge or reciprocity?"*" The math task also presents trials aurally and requires subjects to complete addition and subtraction problems. The trials present subjects with a series of arithmetic operations (e.g., "fourteen plus twelve"), followed by "equals" and then two choices (e.g., "twenty-nine or twenty-six"). Participants push a button to select either the first or the second answer. The math task is adaptive to try to maintain a similar level of difficulty across participants. For more details on the task, please see (Binder *et al.* 2011).

Conditions (Blocked)

Story

Math

References for Language Task: Reliable across subjects and robust activation (Binder *et al.* 2011).

Social Cognition (Theory of Mind)

Participants were presented with short video clips (20 seconds) of objects (squares, circles, triangles) that either interacted in some way, or moved randomly on the screen. These videos were developed by either Castelli and colleagues (Castelli *et al.* 2000) or Martin and colleagues (Wheatley *et al.* 2007). After each video clip, participants judge whether the objects had a mental interaction (an interaction that appears as if the shapes are taking into account each other's feelings and thoughts), Not Sure, or No interaction (i.e., there is no obvious interaction between the shapes and the movement appears random). Each of the two task runs has 5

video blocks (2 Mental and 3 Random in one run, 3 Mental and 2 Random in the other run) and 5 fixation blocks (15 seconds each).

Conditions (Blocked)

Random interaction

Mental interaction

References for the Social Cognition Task: Reliable across subjects and robust activation (Castelli *et al.* 2000; Castelli *et al.* 2002; Wheatley *et al.* 2007; White *et al.* 2011).

Relational Processing

This task was adapted from the one developed by Christoff and colleagues (Smith *et al.* 2007). The stimuli are 6 different shapes filled with 1 of 6 different textures. In the relational processing condition, participants are presented with 2 pairs of objects, with one pair at the top of the screen and the other pair at the bottom of the screen. They are told that they should first decide what dimension differs across the top pair of objects (differed in shape or differed in texture) and then they should decide whether the bottom pair of objects also differ along that same dimension (e.g., if the top pair differs in shape, does the bottom pair also differ in shape). In the control matching condition, participants are shown two objects at the top of the screen and one object at the bottom of the screen, and a word in the middle of the screen (either “shape” or “texture”). They are told to decide whether the bottom object matches either of the top two objects on that dimension (e.g., if the word is “shape”, is the bottom object the same shape as either of the top two objects). For both conditions, the subject responds yes or no using one button or another. For the relational condition, the stimuli are presented for 3500 ms, with a 500 ms ITI, and there are four trials per block. In the matching condition, stimuli are presented for 2800 ms, with a 400 ms ITI, and there are 5 trials per block. Each type of block (relational or matching) lasts a total of 18 seconds. In each of the two runs of this task, there are 3 relational blocks, 3 matching blocks and 3 16-second fixation blocks.

Conditions (Blocked)

Relational processing

Matching

References for the Relational Processing Task: Localizer (Smith *et al.* 2007).

Emotion Processing

This task was adapted from the one developed by Hariri and colleagues (Hariri *et al.* 2002). Participants are presented with blocks of trials that either ask them to decide which of two faces presented on the bottom of the screen match the face at the top of the screen, or which of two shapes presented at the bottom of the screen match the shape at the top of the screen. The

faces have either an angry or fearful expression. Trials are presented in blocks of 6 trials of the same task (face or shape), with the stimulus presented for 2000 ms and a 1000 ms ITI. Each block is preceded by a 3000 ms task cue (“shape” or “face”), so that each block is 21 seconds including the cue. Each of the two runs includes 3 face blocks and 3 shape blocks, with 8 seconds of fixation at the end of each run.

Conditions (Blocked)

Face

Shape

Note: A bug was written into the E-prime script for the EMOTION task, such that the task stopped short of the last three trials of the last task block in each run. This bug was not discovered until data had been collected on several participants. Consequently, the BOLD images and E-Prime data for the EMOTION task are shorter than our original design described above.

References for the Emotion Processing Task: Localizer (Hariri *et al.* 2002); Moderate reliability across time (Manuck *et al.* 2007).

7T MR scanner and other hardware

Coming soon in an updated version of the S1200 Reference Manual.

Summary of 7T MR imaging protocols

Coming soon in an updated version of the S1200 Reference Manual.

7T Retinotopy and Movie Watching stimuli and protocol details

Coming soon in an updated version of the S1200 Reference Manual.

MEG hardware specifications and sensor locations

Scanner hardware and parameters

All HCP subjects being scanned by MEG are scanned on a whole head MAGNES 3600 (4D Neuroimaging, San Diego, CA) system housed in a magnetically shielded room, located at the Saint Louis University (SLU) medical campus.

The MEG system includes 248 magnetometer channels together with 23 reference channels (18 magnetometer and 5 first-order gradiometer channels). The root mean squared (RMS) noise of the magnetometers is ~ 5 fT/sqrt (Hz) on average in the white noise range (above 2 Hz). Data was recorded at 2034.5101 Hz sampling rate. The electrooculography (EOG, 2 channels), electrocardiography (ECG, 1 channel), and electromyography (EMG, 4 channels) recordings are synchronized with the MEG. All electrode impedances are maintained below 10 kOhms. Other parameter settings that are used for all HCP MEG recordings are: Bandwidth of 400Hz, High pass filter: DC, Acquisition mode: continuous, Data encoding: Delta.

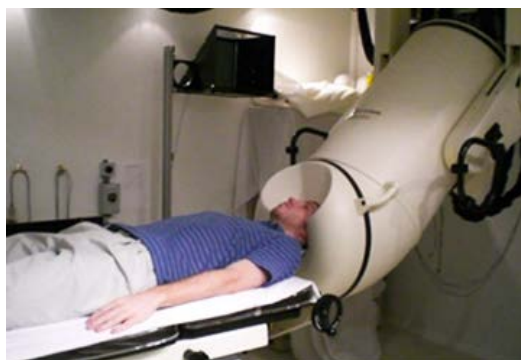
A Polhemus FASTRAK-III is used for spatial digitization of three anatomical landmarks (nasion and L/R periauricular points), five localizer coils, and subjects' head shape (about 2400 points).

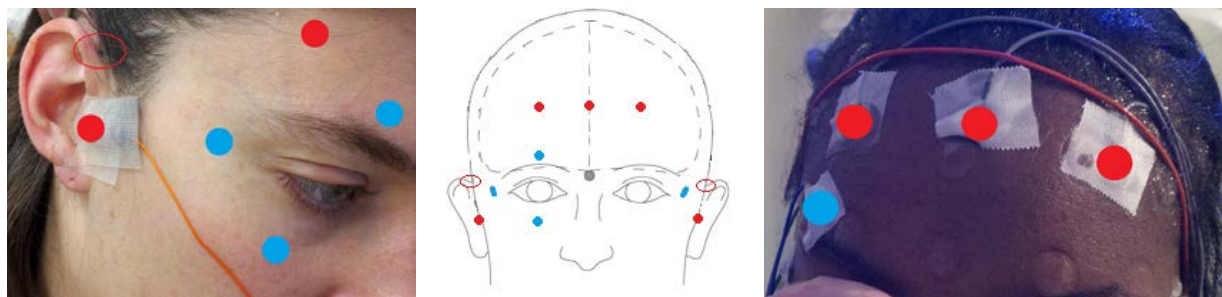
A Sun Blade running 64 bit Solaris 8 Unix and communicating with a dedicated UltraAX-12 system is used for data collection and sensor configuration. Raw data are stored in the native 4D file format. Following data acquisition, the data are uploaded via https to the Washington University in St. Louis (WashU) internal HCP database for quality control and processing steps.

Sensor locations

HCP MEG participants are positioned in the MEG scanner supine, with the crown of the head touching but not pressing into the back of the MEG dewar. This position sets the eyebrows inside the dewar, leaving the rest of the face free of the scanner, as in the picture below.

To co-register MEG data to the MRI structural scans, a subject's reference system and locator coils are needed. The reference system is defined through 3 points, the nasion (gray dot in diagram below) and the peri-auricular points. These are placed where the ear folds to the scalp (red oval in photo below), and they are touched with a stylus perpendicularly to the participant's scalp a few times, to record their position on the head.





Five MEG position-sensor coils are affixed to the head between 1 and 5 cm from the hairline: at each tragus, midline above nasion, and above each eyebrow (as in red dots in photos).

Coil 1 – in front of participant's LEFT tragus (not shown)

Coil 2 – in front of participant's RIGHT tragus (red dot in photo)

Coil 3 – center of forehead near hairline (red dot in center of rightmost photo)

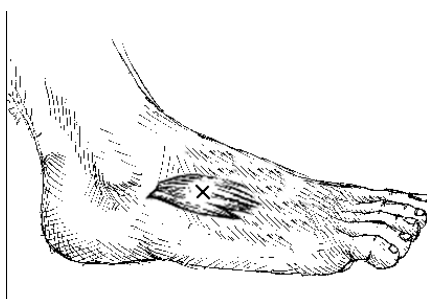
Coil 4 – left side of forehead near hairline in line with Coil 3 (as in rightmost photo)

Coil 5 – right side of forehead near hairline in line with Coil 3 (as in rightmost photo)

To reject cardiac and ocular activity from MEG data (see below for artifact removal pipelines), vertical eye leads (blue dots in photo) are applied above and below participant's right eye, centered on the pupil (channels 3 & 4). Horizontal eye leads (blue dot) are applied about 1cm lateral to the participant's outer canthus of each eye (right as in photo, left not shown, channels 5 & 6). ECG (electrocardiography) sensors are applied bilaterally to the collarbone, or to the lower chest if the collarbone signal is poor.

A ground electrode (GND channel) and a reference electrode (REF1 channel) are applied to the participant's collarbone.

EMG (electromyography) electrode stickers are applied to the participant's skin as in the photos below. The participant is asked to extend the toes and the electrode is placed by palpation over the active muscle.



Sensors are placed on the lateral superior surface on the extensor digitorum brevis muscle (orange ovals) and near the medial malleolus of each foot.

Sensors are placed on the first dorsal interosseus muscle between thumb and index finger, and the styloid process of the ulna at the wrist of each hand.



Stimulus presentation information

Stimuli for task MEG are generated using E-Prime 2.0 Pro on a HP personal computer. The stimulus computer presents visual stimulus through an LCD projector (ImagePro 8935, DUKANE) onto a mirror ~3 feet above the participant at 1024x768 resolution for viewing.

Participant button press responses are recorded via fiber optic as the Response channel in the MEG data. Motor-task motion responses, EOG and ECG are recorded via electrodes as described above.

Audio stimuli are output to an Audiosource amplifier that drives two 300-ohm Etymotic Research (ER-30) transducers. Sound is transferred to the subject through flexible hollow tubes (~7 m), with a measured auditory delay of 20 ms. The delay is compensated in software by offsetting the trigger from the auditory presentation, providing a coincident trigger and stimulus.

For visual stimuli, the TTL triggers and another digital trigger signal from a photodiode installed on the right-top corner of the screen are combined and recorded as the Trigger channel in the MEG data.

All TTL triggers (from the parallel port) of the stimulus computer are recorded as the Trigger channel in the MEG data.

Summary of MEG acquisition protocols

MEG acquisitions were collected for each subject in one 3 hour recording session at Saint Louis University. Our acquisition protocol is included in the [standard operating procedures \(SOP\) Appendix 4](#). Subjects lie supine with body supports to encourage staying still, and are positioned with the top of their head touching but not pressing against the dewar, and the back of their head cushioned by one towel and a thin layer of bubble wrap for comfort.

The MEG recordings for a subject are split over several runs or scans, e.g. three runs for the resting state data and 2 runs for the motor task. At the start and end of each run the head position is recorded. There is no continuous recording of the head position.

The following provides basic parameters for the main scan types in the session. A more complete set of scanning parameters can be found in the protocol exports from the scanner, available in [Appendix 1](#).

Empty room scans

Empty room measurements (5 min) are acquired daily and used to monitor the MEG system for hardware malfunctions or excessive environmental noise. These data are compared to a set of reference scans collected to estimate baseline noise levels and corrective action is taken appropriately. These data are not contained in the S1200 release.

Participant noise scans

Participants are scanned for magnetic artifacts (1 min) and are degaussed and rescanned, if necessary. If multiple participant noise scans were made for a subject, only the final scan is uploaded and shared.

Resting-state MEG (rMEG)

rMEG data were acquired in three runs of approximately 6 minutes each. Subjects are instructed to remain still during scanning, delays between subject instructions and scan start may result in scan duration exceeding six minutes. During the scan the subject is supine, padded for comfort, instructed to relax with eyes open and maintain fixation on a projected red crosshair on a dark background (presented in a darkened room). ECG and EOG electrodes are attached to the subject for cardiac and oculomotor monitoring and off line artifact rejection. Resting-state scans always follow the empty room noise scan and the participant noise scan, and precede the task scans.

| Condition | Runs | Run Duration (min:sec) |
|----------------------|------|------------------------|
| REST (Resting-state) | 3 | ~6:00 |

Task MEG (tMEG)

Following completion of rMEG, subjects were asked to complete tasks that were designed to activate a variety of cortical and subcortical networks. The following table provides a listing of the tMEG scans collected.

tMEG data were acquired with the same parameters as rMEG, except for the run duration information listed below. Duration of each scan is somewhat flexible per run to ensure complete capture of the E-Prime scripts for each of the three tasks.

| Task | Runs | Blocks per run | Run Duration (min:sec) |
|--------------------------------------|------|------------------|------------------------|
| Working Memory | 2 | 16 | 10:00 |
| Story vs. Math (Language Processing) | 2 | 7 story, 15 math | 7:00 |
| Motor | 2 | 32 | 14:00 |

An E-Prime <Subject_ID>_MEG_<Task_run>.tab (tab-delimited text file) and Excel spreadsheet provide output task parameters that can be used in various tMEG analysis packages; see [Task MEG files and protocol details](#).

Full scanning protocols (PDF)

Complete scanning protocols for each MEG acquisition can be found in [Appendix 1](#).

Standard MEG session structure

Each MEG subject was scanned in one 3 hour session, with the following shorthand labels:

MEG Day: MEG Session

- Empty room scan
- Participant Noise Scan
- Head surface digitization
- Three 6-min rMEG scans
- Six tMEG scans (2 scans x 3 tasks: Working Memory, Language Processing [Story-Math], Motor)

The tMEG was collected as six scans, two scans per task type, to acquire enough repetitions to create averages without causing participant fatigue. Task order was not counterbalanced, all were run in the same order: working memory first, then language processing [story-math], then motor.

In some cases, scans were aborted and re-acquired within the same session. This was done if the subject did not comply with task instructions or if technical problems prevented a full set of acquisitions. In these cases, the restart was noted on the acquisition data sheet. Only full length scans are included in the released data.

MEG Physiological Data

In the MEG session, we acquire electrooculography (EOG) to measure eye movement, electrocardiography (ECG), to measure electrical activity of the heart, and electromyography (EMG) to measure electrical activity produced by skeletal muscles movement signals associated with each scan.

Determining Alpha and Beta Peak Frequencies for Individual Subjects

Alpha and beta frequency bands represent two of the strongest oscillations in the human brain. Alpha oscillations are believed to have an inhibitory role and to originate from the thalamus. Beta oscillations are mostly dominant in the sensorimotor cortex and are associated with a "status quo" signaling mechanism. Alpha and beta peak frequencies were calculated for each subject in 2 steps:

The first step is carried out in the [avgspecextract pipeline](#). For each Task, the results of the [tfavg pipeline](#) (time-frequency spectra) are averaged within each sensor group and across time. The same averaging is applied for the results of the [powavg pipeline](#) for Resting state scans, which contain only spectral information (non-time resolved).

The second step is carried out in the [avgspecmanualalphabeta pipeline](#), which calls the GUI function `avgspecmanualalphabetagui`. tMEG and rMEG average spectra are loaded into the GUI where spectral peaks in the alpha and beta band are automatically estimated. The median values for these automatically estimated peak alpha and peak beta across all tMEG and rMEG sensor derivations and conditions are reported as individual subject data in ConnectomeDB under the category "MEG Subjects".

MEG Data Conversion: 4D import to FieldTrip

The raw data are stored on disk in a binary file named 'c,rfdc', which consists of the actual data (16-bits per sample), followed by a 'footer' that contains the metadata needed to interpret the signals. In addition, each data file is accompanied by a file named 'config', which contains additional metadata needed to interpret the signals. The raw binaries can be read with the custom software that comes with the 4D-system, or with FieldTrip. The latter approach requires a MATLAB license. FieldTrip contains the low-level code that is needed to interpret the metadata. More information about using FieldTrip for importing the data into MATLAB can be found on the FieldTrip website <http://fieldtriptoolbox.org>.

MEG Quality control and artifact identification

Quality control and artifact identification steps are included in the pre-processing of raw MEG data using the “datacheck”, “baddata”, and “icaclass” pipelines. These utility pipelines are described in more detail in the [MEG Channel-level processing pipelines](#) section below.

All scans are uploaded from the scanner at SLU and permanently stored in an internal database (IntraDB). Scans that are considered usable for preprocessing are transferred to the public-facing ConnectomeDB based on information encoded in a set of scan-specific fields in IntraDB. This process also entails a reorganization of data from session/modality-specific directories to a composite subject-specific directory in ConnectomeDB, with subdirectories appropriate for the scan modality. Files of a given modality are grouped with ancillary files from the same session as needed in order to facilitate preprocessing using standardized scripts and pipelines. Many file names are modified and standardized as part of this transfer and reorganization process.

tMEG scripts and protocol details

Task-Evoked Brain Activity

Our primary goals in including task MEG (tMEG) in the HCP are to: 1) help identify well-characterized task-related brain areas or “nodes” that can guide, validate and help interpret the results of the connectivity analyses that will be conducted on resting state functional MRI (rfMRI), resting state MEG (rMEG) and diffusion data; 2) to allow a comparison of network connectivity in a task context to connectivity results generated using rMEG and rfMRI; and 3) to relate signatures of activation magnitude or location in key network nodes to individual differences in performance, psychometric measures, or other phenotypic traits. To accomplish these goals, for MEG we use task paradigms testing three domains that are a subset of those used in the HCP for tfMRI: Working Memory, Language processing (Story-Math), and Motor. Significant effort has been made to ensure that MEG and fMRI are performed on the same subjects using the same tasks performed, as closely as possible, with the same timing (see [Larson-Prior et al., 2013](#)). Together with behavioral and genetic information on these subjects, these data will facilitate the study of individual variability across a large population.

These tasks are described in more detail below and in the cited publications. Stimuli are projected onto a mirror ~1 meter above the participant for viewing within the magnetically shielded room (MSR) in which all the MEG recordings are acquired.

tMEG scripts and data files

Script files are run in E-Prime 2.0 Pro to present task MEG stimuli and collect behavioral responses in the scanner. The E-Prime scripts can be made available upon request via email to Tracy at tracyn@npg.wustl.edu for researchers wishing to run HCP tasks in their own research.

Please put "HCP-MEG EPRIME" in the subject line of your email. Tab-delimited text versions of E-Prime data files (in *.tab and *.xlsx formats) will be included in the released data. A brief description of the key variables in those files can be found in [Appendix 6: Task fMRI and tMEG E-Prime Key Variables](#). The original *.edat2 files cannot be shared as they may contain identifying information.

MEG vs. MRI Naming Conventions

Despite the fact that the MEG and fMRI HCP data are collected using the same task paradigms (with slightly different protocol details) and the same eyes-open fixation resting state, differences in the scripting used in the processing pipelines between the MRI data and MEG data have necessitated slightly different naming conventions for each modality. These are shown in the table below.

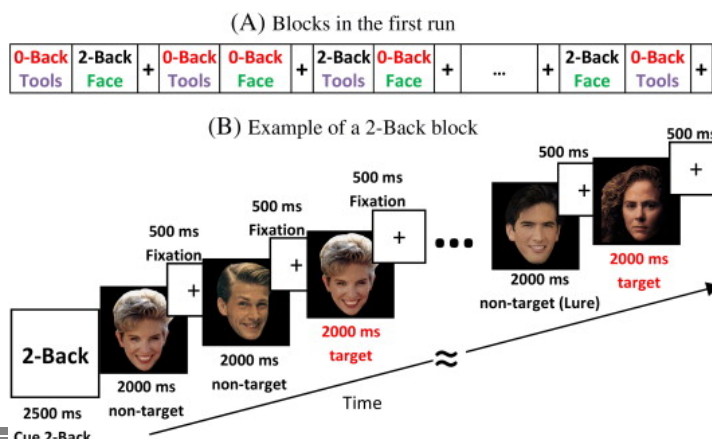
| Paradigm | MEG abbreviation | fMRI abbreviation |
|---------------------|------------------|-------------------|
| Resting-state | Restin | REST1 or REST2 |
| Working Memory | Wrkmem | WM |
| Language Processing | StoryM | LANGUAGE |
| Motor | Motort | MOTOR |

Details of the tMEG Working Memory task

Working memory is assessed using an N-back task in which participants are asked to monitor sequentially presented pictures. Participants are presented with blocks of trials that consisted of pictures of tools or faces. Within each run, the 2 different stimulus types are presented in separate blocks. Also, within each run, $\frac{1}{2}$ of the blocks use a 2-back working memory task and $\frac{1}{2}$ use a 0-back working memory task (as a working memory comparison). Participants are instructed to press a button for every picture. If the currently presented picture matches the cued picture (0-Back) or the same picture that was presented two pictures before (2-Back), subjects press one button with their right index finger. For non-matching pictures, participants press a second button with their right middle finger. Two runs are performed, 16 blocks each, with a bright fixation "rest" on dark background for 15 seconds between blocks.

Stimulus Overview

In the working memory task, participants are instructed to retain images in their visual working memory and compare them with subsequently presented images. There are 2 memory load conditions.



The first is a match-to-sample condition (termed in the following text as 0-Back) during which a cue “Target” image is presented at the beginning of a block and which the subject has been instructed to memorize. Then a sequence of 10 images is presented. Each of these images is displayed for 2000 ms. At the end of this interval, the subject must respond on a button pad with the index or middle finger of the right hand if this image matched the Target or not, respectively. The response must take place within 500 ms, during which a fixation cross is presented on the center of the screen.

The second condition is a 2-Back condition. During such a block, the subject is presented with a sequence of 10 images and must respond if each image is identical to the one 2 positions earlier or not (figure, right). At the beginning of the block there is a cue screen informing the subject that the upcoming stimuli are part of the 2-Back protocol. The timing of the cue screen, the presentation of the 10 stimulus images and of the response interval are identical to that of the 0-Back condition.

There are 2 different categories of images used in this experiment: images of faces and tools. Each block contains images from a single category. Some of the images in the non-matched trials have been characterized as “Lure”. These images have been selected so that they have common features with the target image, but are still different. These trials are flagged as “Lure”.

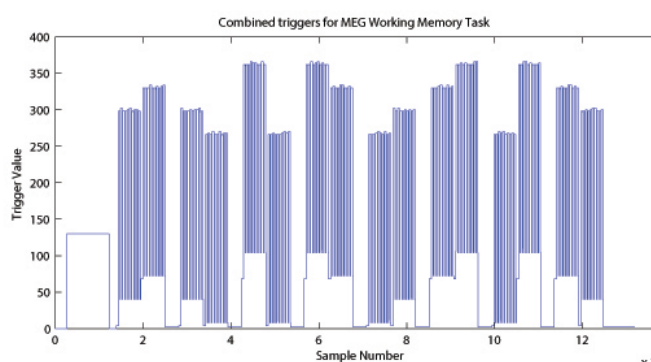
In between the 0-Back and 2-Back blocks are interleaved Fixation Blocks, during which the subject has been instructed to remain fixated on a fixation cross at the center of the screen. The duration of each of these fixation blocks is 15 seconds.

Block/Trial Overview

This experiment is performed in 2 runs with a small break between them. The block/trial breakdown is identical in both runs although the presented stimuli are variable.

Block/Trial breakdown per run:

Total Number of Blocks = 24;
 Number of 0-Back Blocks = 8;
 Number of 2-Back Blocks = 8;
 Number of Faces Blocks = 8;
 Number of Tools Blocks = 8;
 Number of Fixation Blocks = 8;



Original Trigger channel sequence from one run of Working Memory Task. E-prime and Photodiode triggers are superimposed.

Per run trial breakdown with cuing stimuli not included as trials:

Total number of image trials = 160;

Number of 0-Back trials= 80;

Number of 2-Back trials= 80;

Number of Faces trials= 80;

Number of Tools trials= 80;

Trigger Overview

The signal on the trigger channel consists of 2 superimposed trigger sequences. One from the Stimulus PC running the E-prime protocol and one from a photodiode placed on the stimulus presentation screen. The trigger channel for one Working Memory task run is shown, top right.

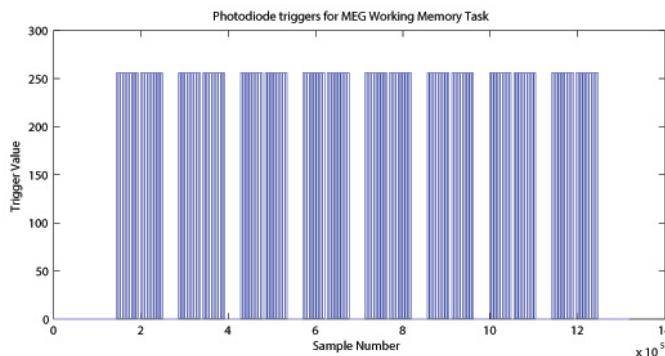
The photodiode is activated on the onset of each of the 10 stimulus images within each block. It is deactivated when each of these stimulus images is removed from the screen. The trigger value for on is 255 and the trigger value for off is 0. The photodiode trigger sequence extracted from the trigger channel of one Working Memory task run is shown, middle right.

The Stim-PC triggers are more detailed. As in the initial trigger channel, these triggers are superimposed on the photodiode triggers, in the following description it is assumed that the photodiode triggers have been subtracted from the trigger channel so that only the triggers from the E-Prime stimulation protocol remain. Such an E-Prime trigger sequence extracted from the trigger channel is shown, previous page, bottom right.

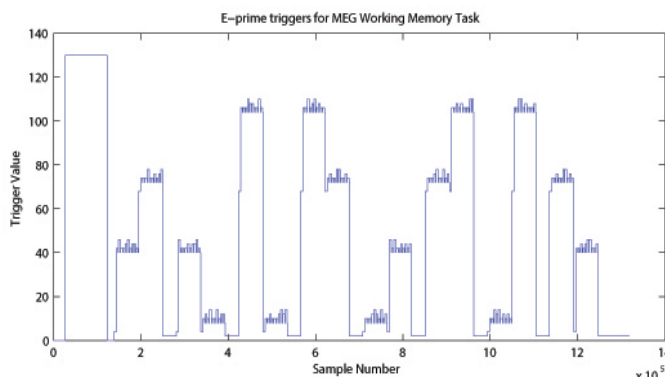
Because of the high number of combinations of Memory Load, Image category, and Match category, the trigger values by E-Prime follow the following algorithmic rules:

Memory Type Trigger Base = { 4 for 0-Back
68 for 2-Back }

Image Category Trigger Pad={ 4 for Faces



Photodiode trigger sequence extracted from the Trigger channel of one Working Memory Task run.



E-Prime trigger sequence extracted from the Trigger channel of one Working Memory Task run.

36 for Tools }

Match Type Trigger Pad= { 2 for Non-target

4 for Lure

6 for Target }

The trigger encoding is based on the following protocol:

Trigger at Onset of Cue Screen = (Memory Type Trigger Base)

Trigger at Onset of Stimulus Image =

(Memory Type Trigger Base)+(Image Category Trigger Pad)+(Match Type Trigger Pad)

Trigger at Offset of Stimulus Image =

(Memory Type Trigger Base)+(Image Category Trigger Pad)

The trigger during the fixation blocks is fixed to: Fixation Trigger = 2;

For descriptions of variables (column headers) to sync the tab-delimited E-Prime output for each run, see [Appendix 6: Task fMRI and tMEG E-Prime Key Variables](#).

Details of the tMEG Language Processing (Story-Math) task

This task was developed by Binder and colleagues (Binder *et al.* 2011). The task consists of two runs that each interleave 7 blocks of a story task and 15 blocks of a math task. The lengths of the blocks vary (average of approximately 30 seconds), but the task was designed so that the math task blocks match the length of the story task blocks, with some additional math trials at the end of the task to complete the ~7 minute run as needed.

The story blocks present participants with brief auditory stories (5-9 sentences) adapted from Aesop's fables, followed by a 2-alternative forced-choice question that asks participants about the topic of the story. The example provided in the original Binder paper (p. 1466) is "*For example, after a story about an eagle that saves a man who had done him a favor, participants were asked, "Was that about revenge or reciprocity?"*" The math task also presents trials aurally and requires subjects to complete addition and subtraction problems. The trials present subjects with a series of arithmetic operations (e.g., "fourteen plus twelve"), followed by "equals" and then two choices (e.g., "twenty-nine or twenty-six"). Participants push a button to select either the first or the second answer. The math task is adaptive to try to maintain a similar level of difficulty across participants.

Stimulus Overview

In the Story-Math task participants listened either to an auditory sequence of sentences comprising Story narratives (Aesop fables) or matched-duration simple arithmetic Math problems (sequential addition/subtraction) (Binder *et al.*, 2011). Each of the stimulus sequences were followed by a 2-alternative forced choice question. In the Story case, the distinction of 2 alternative answers is qualitative (e.g. Was this story about loyalty or betrayal?) In the Math case, the distinction is quantitative (e.g. What is the correct answer: 3 or 7?).

Participants responded with the index and middle finger of the right hand on a button pad for the first or the second alternative, respectively. The 2 alternatives are presented in sequence and NOT at the same time (e.g. screen 1: What is the correct answer?, screen 2: 3, screen 3: or, screen4: 7). Each Story lasted about 30 seconds. Each Individual Math problem is much shorter, so multiple Math problems are put in sequence in order to form a block of matched duration of about 30 seconds. The number of operations within each Math problem was variable and was dynamically adjusted between blocks in order to adapt to the mathematical fluency of the subject. Typically, the number of operations varied between 3 and 5.

Block/Trial Overview

This experiment is performed in 2 runs with a small break between them. Run 1 consisted of 4 Story blocks and 4 Math blocks. Run 2 consisted of 5 Story blocks and 5 Math blocks.

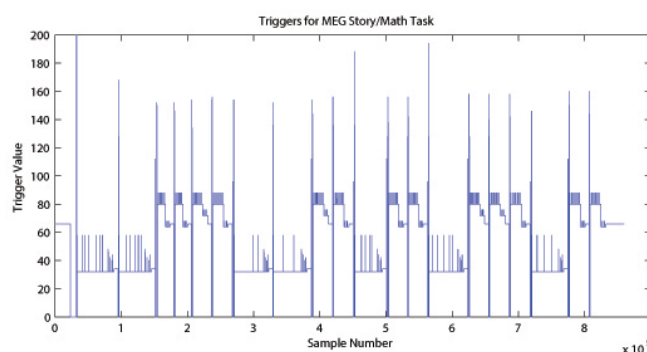
Trigger Overview

The trigger channel contains the trigger sequence from the E-Prime experimental protocol which runs on the stimulus presentation PC. The trigger channel for one Story/Math task run is shown, right.

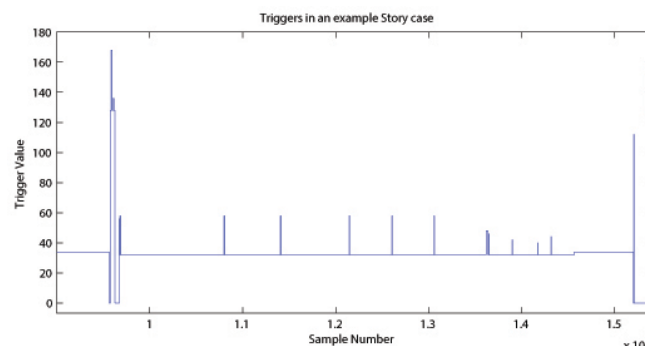
An example trigger sequence during a

Story block is shown, below.

Before the beginning of a Story block, there is a first header trigger block. In this header trigger block, trigger values go from 0 to 128. Then there is a following trigger going from 128 to a higher value and then returns back to



Trigger channel sequence from one entire run of Story Math Task.



Trigger channel sequence during a Story block.

128. This excess value over 128, is the ID number of the current Story, multiplied by 2.

Then another trigger pulse follows within the header block. The excess value of this second trigger pulse over 128 represents the level of the current story multiplied by 2. The 'story level' is a quantitative representation of 4 different levels of difficulty for the participant to distinguish between the two different options presented at the end of each Story. Therefore, each story has 4 different pairs of options, one for each level of difficulty. For example, these 4 different levels can be:

"This story was about:"

Level 1: a bull or a mule?

Level 2: squeezing or kicking?

Level 3: aging or power?

Level 4: inexperience or deliberation?

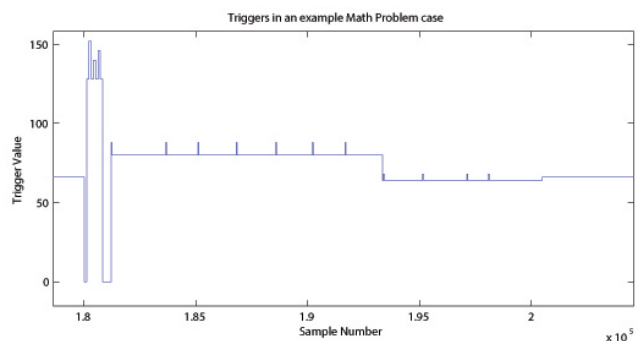
At the end of the header trigger block the trigger returns to 0.

When the actual stimuli in the Story Block are about to start, and for its entire duration, the trigger baseline changes from 0 to 32, upon which all other triggers are superimposed.

| Trigger value | Meaning | Duration |
|---------------|--|---|
| 96 | Start of story block | transient |
| 32 | Story block offset | for the duration of the block |
| 24 + offset | Start of narrative | transient |
| 26 + offset | Start of single story | transient |
| 16 + offset | End of narrative | transient |
| 14 + offset | Start of question ("That story was about ?") | transient |
| 12 + offset | Start of correct option (e.g. "Loyalty") | transient |
| 10 + offset | Start of incorrect options (e.g. "Betrayal") | transient |
| 8 + offset | Start of "Or" between the 2 above options | transient |
| 2 + offset | Start of response period | for the duration of the response period |

An example trigger sequence during a Math block is shown below. Before the beginning of a Math problem block, there is a first header trigger block. In this header trigger block, trigger values go from 0 to 128. Then there is a following trigger going from 128 to a higher value and then returns back to 128. This excess value over 128, is the Difficulty Level of the current Math problem multiplied by 2.

Then there are 2 more trigger pulses within the header block superimposed on the base value of 128. The values of these 2 trigger pulses encode the Number of each math problem. The numerical amount over 128 of the value of the first trigger pulse, represents the value of the 4 Most Significant Bits of the Math problem number divided by 8. The numerical amount over 128 of the value of the second trigger pulse, represents the value of the 4 Least Significant Bits of the Math problem number multiplied by 2.



Trigger channel sequence during a Math problem block.

For example, if these 2 trigger pulses take values 142 and 156, then after the header block base value (128) is subtracted these values become 14 and 28, respectively. By multiplying the first by 8 and dividing the second by 2, these values become 112 and 14, respectively. The Math problem number is then derived by summing these numbers, so in this example the Math ID number is $112 + 14 = 126$.

The same Math problem number can exist for different difficulty levels. So the unique identification of a used problem requires both the difficulty level number and the Math problem number. So if in the example above the difficulty level was 10, then in order to uniquely identify this math problem one would need to know the pair of values [10, 126]. This is also reflected in the naming of the wav files used for the presentation of a specific math problem. So for the above example the wav file used by the E-Prime protocol would be called:

math-level10-126-Q.wav

At the end of the header trigger block the trigger returns to 0.

When the actual Math Problem Block starts, and for its entire duration, the trigger baseline changes from 0 to 64, upon which all other triggers are superimposed.

| Trigger value | Meaning | Duration |
|------------------------|--|-------------------------------------|
| 112 | Start of math block | transient |
| 64 | Math block offset | for the duration of the block |
| 16 + offset | Math block instruction offset | for the duration of the instruction |
| 8 + instruction offset | Presentation of number or arithmetic operand | transient |
| 0 + offset | Math block question period offset in case the first | for the duration of the question |

| Trigger value | Meaning | Duration |
|----------------------------|---|---|
| | option provided is the correct answer | |
| 4 + offset | Math block question period offset in case the second option provided is the correct answer | for the duration of the question |
| 4 + question period offset | Presentation of question elements | transient |
| 2 + offset | Start of response period | for the duration of the response period |

The first part of the Math problem block is the narration of the arithmetic sequence. An example of an instruction would be: "3 + 2 + 6 + 4", where the onset of the presentation of each number and operand is marked by a transient trigger on top of the math block instruction offset. An example of the question would be: "equals 15 or 17", where the onset of the presentation of each of these elements is marked by a transient trigger on top of the question period offset.

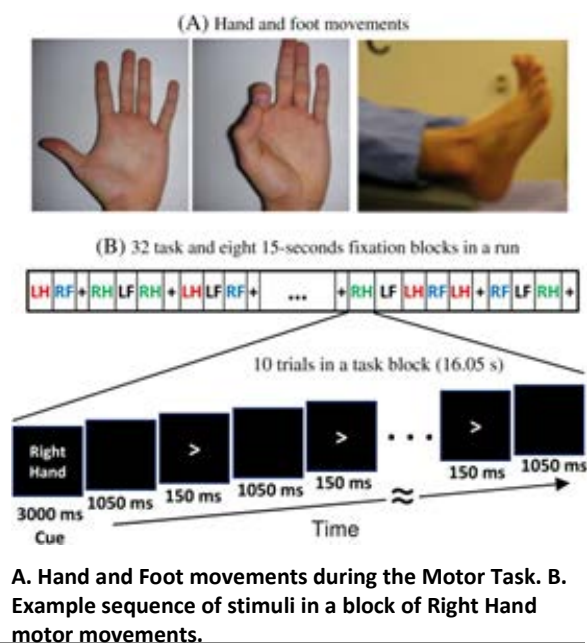
The transient triggers last about 20 msec. At the end of the response period, the Math Block ends and the trigger returns to 0.

For descriptions of variables (column headers) to sync the tab-delimited E-Prime output for each run, see [Appendix 6: Task fMRI and tMEG E-Prime Key Variables](#).

References for Language Processing Task: Reliable across subjects and robust activation (Binder *et al.* 2011).

Details of the tMEG Motor task

Sensory-motor processing is assessed using a task in which participants are presented with visual cues instructing the movement of either the right hand, left hand, right foot, or left foot. Movements are paced with a visual cue, which is presented in a blocked design. This task was adapted from the one developed by Buckner and colleagues (Buckner *et al.* 2011; Yeo *et al.* 2011). Participants are presented with visual cues that ask them to either tap their left or right index and thumb fingers or squeeze their left or right toes. Each block of a movement type lasts 12 seconds (10 movements), and is preceded by a 3 second cue. In each of the two runs, there



are 32 blocks, with 16 of hand movements (8 right and 8 left), and 16 of foot movements (8 right and 8 left). In addition, there are nine 15-second fixation blocks per run. EMG signals were used for onset of event for hand and foot movement. EMG electrode stickers are applied as shown in [MEG hardware specifications and sensor locations](#) to the skin to the lateral superior surface of the foot on the extensor digitorum brevis muscle and near the medial malleolus, also the first dorsal interosseus muscle between thumb and index finger, and the styloid process of the ulna at the wrist.

Stimulus Overview

In the Motor task, participants executed a simple hand or foot movement. The limb and the side were instructed by a visual cue, and the timing of each movement was controlled by a pacing arrow presented on the center of the screen (A, right). The paradigm included movement and rest blocks (B, right).

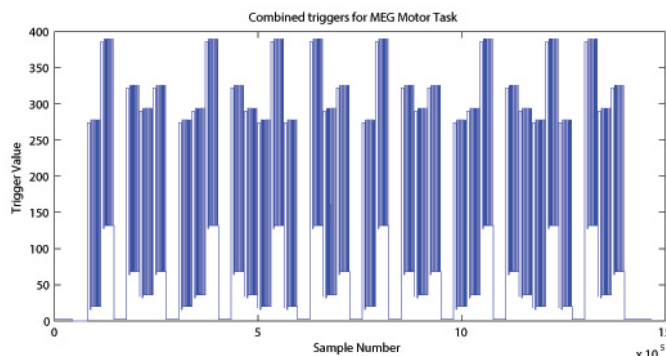
Block/Trial Overview

Each block started with an instruction screen, indicating the side (left, right) and the limb (hand, foot) to be used by the subject in the current block. Then, 10 pacing stimuli were presented in sequence, each one instructing the participant to make a brisk movement. The pacing stimulus consisted of a small arrow in the center of the screen pointing to the side of the limb movement (left or right, above). The interval between consecutive stimuli was fixed to 1200 msec. The arrow stayed on the screen for 150 msec and for the remaining 1050 msec the screen was black.

In addition to the blocks of limb movements there were 10 interleaved resting blocks, each one of 15 sec duration. During these blocks the screen remained black. The last block was always a resting block after the last limb movement block.

The experiment was performed in 2 runs with a small break between them. The block/trial breakdown was identical in both runs. Each of the runs consisted of 42 blocks. 10 of these blocks were resting blocks, and there were 8 blocks of movement per motor effector. This yielded in total 80 movements per motor effector.

In addition to the recorded MEG channels, EMG activity was recorded from each limb. Also ECG and EOG electrodes were used to record heart- and eye movement-related electrophysiological activity.

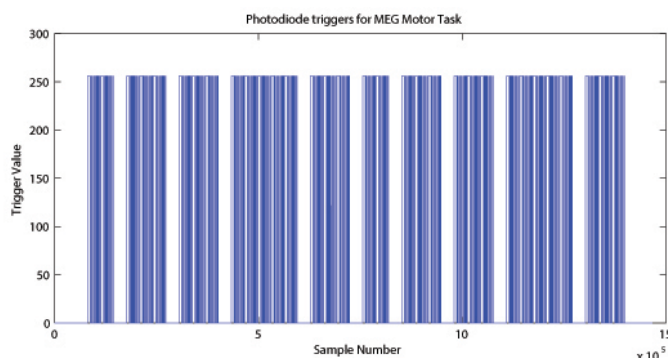


Original Trigger channel sequence from one run of Motor Task. E-Prime and Photodiode triggers are superimposed.

Trigger Overview

The signal on the trigger channel consists of 2 superimposed trigger sequences. One from the Stimulus PC running the E-Prime protocol and one from a photodiode placed on the stimulus presentation screen. The trigger channel for one Motor task run is shown, top right.

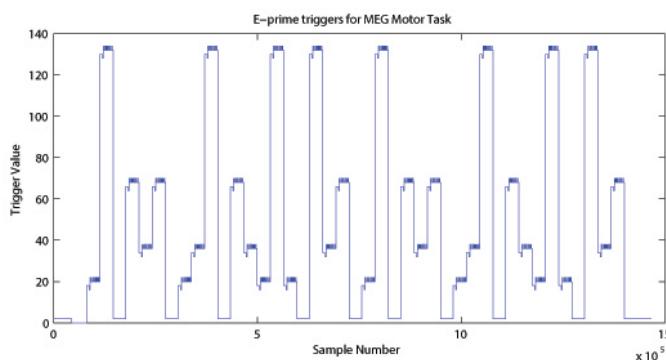
The photodiode was activated whenever a cueing stimulus or pacing arrow was presented on the display. It was deactivated when the display was black. The trigger value for on is 255 and the trigger value for off is 0. The photodiode trigger sequence extracted from the trigger channel of one Motor task run is shown, bottom right.



Photodiode trigger sequence extracted from the Trigger channel of one Motor Task run.

The E-Prime triggers contain the information about the experimental sequence. These triggers are superimposed on the photodiode triggers, in the following description it is assumed that the photodiode triggers have been subtracted from the trigger channel so that only the triggers from the E-Prime stimulation protocol remain. Such an E-Prime trigger sequence, extracted from the trigger channel is shown, right.

For descriptions of variables (column headers) to sync the tab-delimited E-Prime output for each run see [Appendix 6: Task fMRI and tMEG E-Prime Key Variables](#). This task contains the following events, each of which is computed against the fixation baseline.



E-Prime trigger sequence extracted from the Trigger channel of one Motor Task run.

References for Motor Task:

Localizer (Morioka *et al.* 1995; Bizzi *et al.* 2008; Buckner *et al.* 2011; Yeo *et al.* 2011).

| Trigger value | Meaning | Duration |
|---------------|------------------------|--------------------------|
| 2 | resting block | entire duration of epoch |
| 16 | epoch offset left hand | entire duration of epoch |
| 32 | epoch offset left foot | entire duration of epoch |



| | | |
|---------------------------|-------------------------|-----------------------------------|
| 64 | epoch offset right hand | entire duration of epoch |
| 128 | epoch offset right foot | entire duration of epoch |
| 2 + epoch offset | cue stimulus onset | transient |
| 4 + epoch offset | movement epoch offset | entire duration of movement epoch |
| 2 + movement epoch offset | pacing arrow onset | transient |

| Conditions (Blocked) | Trigger codes |
|----------------------|---------------|
| Left finger blocks | 22 |
| Right finger blocks | 70 |
| Left toe blocks | 38 |
| Right toe blocks | 134 |

Summary of Behavioral and Individual Subject measures

HCP Subjects undergo many behavioral tests that are part of the [NIH Toolbox](#) battery and several Non-Toolbox behavioral measures. Behavioral measures are collected for the following domains (detailed descriptions of these measures can be found in [Chapter 5: Behavioral and Individual Subject measures details](#)).

*Available to users with access to Restricted Data only. To apply download the [Restricted Access Data Use Terms Application](#).

Demographics

- Gender
- 5-year Age Range
- Age by Year*
- Has Genotyping data Available (HasGT)*
- ZygositySR (Self-Reported)*
- ZygosityGT (Verified by Genotyping)*
- Mother_ID (Verified by Genotyping, if HasGT=true for sibling pairs, otherwise self-reported)*
- Father_ID (Verified by Genotyping, if HasGT=true for sibling pairs, otherwise self-reported)*
- Family_ID (Verified by Genotyping, if HasGT=true for sibling pairs, otherwise self-reported)*
- Race, Ethnicity*
- Handedness*
- Self-Reported demographics on education, income, relationship status from SSAGA*

Health and Family History*

- Height, Weight, Body Mass Index*
- SSAGA BMI Category (Current and at Heaviest)*
- Hematocrit Levels*
- Blood Pressure*
- Thyroid Function*
- Glucose Levels*
- History of Endocrine disorders*
- Menstrual Cycle Information*
- Parental Psychiatric or Neurological Illnesses*

MEG Subjects

- Alpha Peak
- Beta Peak

Alertness

- Cognitive Status (MMSE)
- Sleep (PSQ)

Cognition

- Episodic memory (Picture sequence and Verbal)
- Executive Function (Cognitive Flexibility and Inhibition)
- Fluid Intelligence
- Language (Reading decoding and Vocabulary comprehension)
- Processing Speed
- Self-regulation/Impulsivity (Delay Discounting)
- Spatial Orientation
- Sustained Attention
- Working Memory

Emotion

- Emotion recognition
- Negative Affect
- Psychological Well-being
- Social Relationships
- Stress and Self-Efficacy

Motor

- Endurance
- Locomotion
- Dexterity
- Strength

Personality

- Five Factor Model (NEO-FFI)

Psychiatric and Life Function

- Achenbach Self-Report of Life function and Psychiatric Clinical Symptoms*
- Self-reported Psychiatric Clinical Symptom measures from SSAGA*

Sensory

- Audition
- Olfaction
- Pain

- Taste
- Contrast Sensitivity
- Color Vision*
- Visual Acuity*

Substance Use

- Breathalyzer (day of scan)*
- Urine Drug Screen*
- Seven-day Alcohol and Tobacco Use Retrospective*
- Self-Reported Substance Use and Abuse measures from SSAGA*

Chapter 3: 1200 Subjects Release Data

Final HCP data collection yielded 214 new 3T subjects with data from all or nearly all modalities and behavioral tests. Data from all 3T MR modalities in the HCP protocol is now available for over 1100 subjects. New for this release, Retest data on the full 3T MR HCP protocol is available for 46 subjects. 7T data (resting state, retinotopy, movie-watching and diffusion) are now available for 184 subjects. Resting state and/or task MEG data are available on a total of 95 subjects. A detailed file-by-file listing of the content of the release is provided in the [Appendix 3: File Names and Directory Structure for 1200 Subjects Data](#).

The S1200 release also includes whole genome genotyping data on 1142 of the 1206 total released HCP subjects. Family structure data, which was originally based on self reports of sibling and parent identities, has been updated to reflect genotyping findings of relationships (see [Zygosity, Genotyping, Family, and Parent IDs](#) for more information). Whole genome SNP data on these subjects will be released in spring 2017 on dbGaP.

Pipeline changes and Data versioning

Since the S900 (Dec 2015) and S500 (June 2014) releases, only incremental improvements and bug fixes have been made to the version 3.x branch of the MR preprocessing and processing pipelines. The only exception is the diffusion preprocessing pipeline, where we have switched to using an updated version of FSL's EDDY tool (v5.0.10) for 3T and 7T diffusion minimal preprocessing. This is why we reprocessed all diffusion data on all subjects for the S1200 release. We then added bedpostX for 3T diffusion modelling of crossing fibers for probabilistic tractography. Otherwise only minor pipeline adjustments were necessary to accommodate the processing of the 7T MR data in 7T-specific processing pipelines. The latest version 3 (v3.21 as of Feb 2017) pipeline scripts to be used with FreeSurfer, FSL, and Connectome Workbench are available on GitHub at <https://github.com/Washington-University/Pipelines>.

Therefore, with the updated diffusion processing, all data currently available on all subjects in ConnectomeDB, regardless of original collection or release dates, can be compared between subjects. Since you might be using previously released versions of the data, check the release notes distributed with each subject dataset you plan to use to ensure all data used in your analyses have been processed with the same major pipeline versions.

S1200 Data available

In addition to the unprocessed, minimally preprocessed, and additionally processed (ICA-FIX denoised rfMRI data, within-subject analyzed individual tfMRI grayordinates data) 3T data of all

MR modalities for more subjects, the 1200 Subjects release contains updated preprocessed and bedpostX-processed 3T diffusion data on all subjects, full protocol 3T retest data on 45 subjects, and 7T HCP protocol (rfMRI, retinotopy, movie-watching, diffusion) data for 184 subjects.

46 Subjects (all monozygotic twins, 21 twin pairs + 4 MZ twins without retest of co-twin) were recruited to undergo the full 3T HCP imaging and behavioral protocol for a second time (a second 2-day visit). 3T Retest imaging data is available on 45 of these subjects. Retest datasets are available in the separate WU-Minn HCP Retest Data project on the ConnectomeDB splashpage. 7T data is available on 22 Retest subjects. The retest data are released as a separate project to fully distinguish it from the first visit. Retest subjects retain the same Subject ID numbers as in the 1200 Subjects HCP project. Most restricted data for retest subjects is only available in the main HCP project because most restricted data are collected prior to the scan and behavioral test visit (in the screening and SAAGA phone interviews) that was part of the retest.

Note: Be sure to keep any Retest data separate in your local repository. Downloaded data from the Retest project will unpack exactly as from the 1200 Subjects project. Be careful not to overwrite a subject's first visit data with their second visit Retest data!

We are in the process of creating group-average 3T rfMRI and tfMRI datasets, and ICA-based group-average parcellation, individual node timeseries, and individual and group-average netmats (parcellated connectomes) data ("PTN" data) for a larger group of 1000+ subjects that are part of the 1200 Subjects Release. This updated data will be released in spring 2017. In the meantime, group-average and PTN datasets will continue to be available for R800+ subjects from the S900 Release.

MEG data in the 1200 Subjects release is unchanged from that of the S900 release and is comprised of excellent quality resting-state MEG (rMEG) and/or one or more of each of 3 task MEG (tMEG) paradigms (working memory, language processing [story-math], and motor) scans from each of 95 individual subjects. Some subjects have all scans in the HCP MEG protocol, but many subjects have quality scans of only a subset of these types. The data is available in unprocessed, preprocessed (channel-level), and source-level processed formats. Task MEG data also include minimally processed E-Prime output information for users. All MEG subjects also have behavioral, MR imaging data, and MEG anatomical head models based on both 3T MR structural scans and head surface digitization available in ConnectomeDB. Anatomical models for MEG source estimation and scans of baseline empty room noise and baseline participant noise are available for all MEG subjects.

S1200 HCP Subject Family Structure

Of the 1206 released HCP subjects, 1142 subjects had usable genetic material from blood or saliva samples processed using one custom microarray chip consisting of the Illumina Mega Chip (has 2 million multiethnic single nucleotide polymorphisms [SNPs]), immune-related content from the ImmunoArray, and psychiatric-related content from the PsychArray. Samples were also processed with a second Illumina Neuro Consortium chip to look for SNPs particularly relevant to neuroimaging studies. Subjects with genotyping data for at least one of the two chips can be filtered with restricted data access in ConnectomeDB or in the downloadable restricted CSV file with the measure HasGT=TRUE.

The available genotyping data has been used to confirm the biological family relationships (and zygosity of twin pair subjects) between HCP Subjects who were all originally recruited as families of siblings that included a twin pair. Family relationships are defined by these restricted data measures (see [Zygosity, Genotyping, Family, and Parent IDs](#) for details):

ZygositySR (NotTwin, NotMZ, and MZ) is the original self-reported zygosity.

ZygosityGT (MZ, DZ) is the twin zygosity verified by genotyping.

Mother_ID and Father_ID are verified by genotyping between siblings, if available, and, if not, are based on self report.

Family_ID is a listing of all Mother_IDs and Father_IDs (linked by underscores) associated with subjects who share at least one parent.

Note: in order to genetically confirm a relationship (shared parents) between siblings, and/or zygosity among twin pairs, BOTH siblings/twin pairs must be HasGT=TRUE.

Family Structure Summary for HCP Subjects (1206 Total):

1. There are 457 unique Family_IDs for the 1206 total subjects in the study. If we compute family structure without regard to the status of HasGT (i.e., some siblings are not genetically verified) these 457 families consist of:

- 48 singleton families
- 138 sibships of 2
- 210 sibships of 3
- 54 sibships of 4
- 6 sibships of 5
- 1 sibship of 6

2. 395 Family_IDs have HasGT=TRUE for all subjects in the family. (i.e., the entire family structure of the participating subjects is genetically verified).

These 395 families consist of:

- 45 singleton families
- 123 sibships of two
- 177 sibships of three
- 44 sibships of four
- 5 sibships of five
- 1 sibship of six

3. 298 subjects (149 pairs) are genetically confirmed MZ twins [ZygosityGT=MZ; HasGT=TRUE] This includes 70 subjects (35 pairs) that self-reported as DZ (ZygositySR='NotMZ'), but which are actually MZ.

4. 188 subjects (94 pairs) of genetically confirmed DZ twins [ZygosityGT=DZ; HasGT=TRUE].

5. 66 subjects with ZygositySR=MZ, but ZygosityGT=Blank

This consists of:

- 28 subjects whose putative MZ twin is not part of the 1206 released study subjects.
- 38 subjects whose putative MZ twin IS part of the 1206, but HasGT=FALSE for one of the pair, thus ZygosityGT=Blank. (i.e., MZ status cannot be genetically confirmed for these 19 pairs).

6. 65 subjects with ZygositySR=NotMZ, but ZygosityGT=Blank.

This consists of:

- 47 subjects whose putative DZ twin is not part of the 1206 released study subjects.
- 18 subjects whose putative DZ twin IS part of the 1206, but HasGT=FALSE for one or both of the pair, thus ZygosityGT=Blank. (i.e., DZ status cannot be genetically confirmed for these 9 pairs).

7. 64 subjects have HasGT=FALSE (leaving 1142 subjects with HasGT=TRUE)

HCP Family Types (1203 Subjects Included)**

The HCP family structure data (SubjectID, Mother_ID, Father_ID, and ZygosityGT [or ZygositySR if ZygosityGT is blank]) were run through a script [hcp2blocks.m](https://github.com/andersonwinkler/hcp2blocks.m) (available on GitHub courtesy of Anderson Winkler) to itemize the numbers of all the different types of families represented in the final cohort of S1200 HCP Subjects that have been released.

Note: The Family types listed here are only based on biological relationships. The family groupings do not imply shared household or upbringing environment.

NS = Non-sibling (unrelated), FS = Full Sibling, HS = Half Sibling, DZ = Dizygotic Twin, MZ = Monozygotic Twin. If subjects have more than one kind of relationship represented among siblings both are indicated with a “/” between them, for example, when a family has one or more HS/NS, this indicates that one subject is the half sibling of one or more other family members *and* is unrelated to one or more other family members.

| Family type | # of Families of type | Sibship Size | Number of Subjects |
|-----------------------|-----------------------|--------------|--------------------|
| 1 NS | 48 | 1 | 48 |
| 2 FS | 85 | 2 | 170 |
| 3 FS | 33 | 3 | 99 |
| 2 FS + 1 HS | 3 | 3 | 9 |
| 3 HS | 1 | 3 | 3 |
| 3 HS/NS | 1 | 3 | 3 |
| 4 FS | 11 | 4 | 44 |
| 3 FS + 1 HS | 1 | 4 | 4 |
| 5 FS | 1 | 5 | 5 |
| 3 FS/HS + 2 FS/HS | 1 | 5 | 5 |
| 2 DZ | 18 | 2 | 36 |
| 2 DZ + 1 FS | 58 | 3 | 174 |
| 2 DZ + 1 HS | 5 | 3 | 15 |
| 2 DZ + 2 FS | 18 | 4 | 72 |
| 2 DZ + 1 FS + 1 HS | 2 | 4 | 8 |
| 2 DZ + 2 HS | 1 | 4 | 4 |
| 2 DZ + 2 HS + 1 HS/NS | 1 | 5 | 5 |
| 2 MZ | 34 | 2 | 68 |
| 2 MZ + 1 FS | 101 | 3 | 303 |
| 2 MZ + 1 HS | 8 | 3 | 24 |
| 2 MZ + 2 FS | 20 | 4 | 80 |
| 2 MZ + 2 HS | 2 | 4 | 8 |
| 2 MZ + 3 FS | 2 | 5 | 10 |
| 2 MZ + 4 FS | 1 | 6 | 6 |
| TOTALS | 456 | | 1203 |

****3** Subjects were not included in this listing due to their missing both ZygosityGT and ZygositySR (these 3 subjects self reported as twins but did not specify their zygosity, so ZygositySR is blank). Should a user want to keep these, they may edit their own copy of the restricted measures CSV, adding in the sibship information by hand, then run the script again.

The total number of families in the table differs by 1 from the 457 mentioned in the Family Structure Summary above because 2 of the 3 excluded subjects are of the same family.

HCP Users and MEG Data FAQs

Updated HCP MR Data FAQ is available on the HCP public wiki at:

<https://wiki.humanconnectome.org/display/PublicData/HCP+Users+FAQ>

Updated MEG Data FAQ is available on the HCP public wiki at:

<https://wiki.humanconnectome.org/display/PublicData/MEG+Data+FAQ>

Users are strongly encouraged to join the HCP Users mailing list and ask HCP acquisition, data, and analysis questions for 3T MR, 7T MR, or MEG data there (post questions to hcp-users@humanconnectome.org). Contributions to the hcp-users mailing list are monitored and responded to by investigators and staff in the WU-Minn HCP consortium. Sign up at <http://www.humanconnectome.org/contact/> or by checking the appropriate box when registering for ConnectomeDB.

Directory structure for unprocessed MR data

Unprocessed datasets downloaded from ConnectomeDB are organized into subject-specific and modality-specific archives. When unpacked, they are moved into a <subject_id>/unprocessed/3T/ subdirectory (by creating it if not present, or placing into existing directories if present). If you have ordered Connectome in a Box, the directory structure is the same as the unpacked downloaded data. A full list of file names in each directory and subdirectory is provided in [Appendix 3: File Names and Directory Structure for 1200 Subjects Data, Section A](#).

Unprocessed data for exemplar subject 100307 unpacks to the following directory structure:

```
100307/unprocessed/3T/  
  100307_3T.csv  
  Diffusion/  
  rfMRI_REST1_LR/  
  rfMRI_REST1_RL/  
  rfMRI_REST2_LR/  
  rfMRI_REST2_RL/  
  T1w_MPR1/  
  T2w_SPC1/  
  tfMRI_EMOTION_LR/  
  tfMRI_EMOTION_RL/  
  tfMRI_GAMBLING_LR/  
  tfMRI_GAMBLING_RL/  
  tfMRI_LANGUAGE_LR/  
  tfMRI_LANGUAGE_RL/
```

tfMRI_MOTOR_LR/
tfMRI_MOTOR_RL/
tfMRI_RELATIONAL_LR/
tfMRI_RELATIONAL_RL/
tfMRI_SOCIAL_LR/
tfMRI_SOCIAL_RL/
tfMRI_WM_LR/
tfMRI_WM_RL/

The 3T/ subdirectory signifies that these data were acquired on the 3T Connectome Skyra at Wash U.

For the subjects that are scanned at 7T (184 of the 1206), the 7T data unpack to a 7T/ subdirectory. The .csv file contains the subject-specific spreadsheet of scan session data, which provides useful metadata about the scans acquired for this particular subject. An example of the spreadsheet is available as [Appendix 5: MR Data Acquisition Information for an Exemplar Subject](#).

Within each subdirectory, files are named as indicated by the exemplars below for several scan types.

T1w_MPR1/
100307_3T_AFI.nii.gz
100307_3T_BIAS_32CH.nii.gz
100307_3T_BIAS_BC.nii.gz
100307_3T_FieldMap_Magnitude.nii.gz
100307_3T_FieldMap_Phase.nii.gz
100307_3T_T1w_MPR1.nii.gz

T2w_SPC1/
100307_3T_AFI.nii.gz
100307_3T_BIAS_32CH.nii.gz
100307_3T_BIAS_BC.nii.gz
100307_3T_FieldMap_Magnitude.nii.gz
100307_3T_FieldMap_Phase.nii.gz
100307_3T_T2w_SPC1.nii.gz

The FieldMap, BIAS, and AFI scans are ancillary files copied into multiple structural subdirectories, in order to facilitate using standardized scripts and pipelines.

If multiple structural scans for a given subject passed the QC criteria and are included in ConnectomeDB, they will be in separate subdirectories T1w_MPR2/ and/or T2w_SPC2, with uniquely named structural files (e.g. 100307_3T_T1w_MPR2.nii.gz and/or 100307_3T_T2w_SPC2.nii.gz). Each subdirectory will include copies of relevant ancillary files needed for preprocessing.

The four 15-min rfMRI scans for each subject are downloaded into separate subdirectories, because the initial preprocessing is carried out separately for each scan.

rfMRI_REST1_LR
100307_3T_BIAS_32CH.nii.gz


```
100307_3T_BIAS_BC.nii.gz
100307_3T_rfMRI_REST1_LR_SBRef.nii.gz
100307_3T_rfMRI_REST1_LR.nii.gz
100307_3T_rfMRI_REST1_LR_Physio_log.txt
100307_3T_SpinEchoFieldMap_LR.nii.gz
100307_3T_SpinEchoFieldMap_RL.nii.gz
```

rfMRI_REST1_RL

```
100307_3T_BIAS_32CH.nii.gz
100307_3T_BIAS_BC.nii.gz
100307_3T_rfMRI_REST1_RL_SBRef.nii.gz
100307_3T_rfMRI_REST1_RL.nii.gz
100307_3T_rfMRI_REST1_RL_Physio_log.txt
100307_3T_SpinEchoFieldMap_LR.nii.gz
100307_3T_SpinEchoFieldMap_RL.nii.gz
```

Here, the ancillary files are SpinEchoFieldMap and BIAS scans acquired at the beginning of the rfMRI session and copied into both the rfMRI_REST1_LR/ and rfMRI_REST1_RL/ directories to simplify preprocessing scripts.

For tfMRI scans, the 14 scans (7 tasks x two phase encoding directions) are downloaded into separate subdirectories, because the initial preprocessing is carried out separately for each scan. For example, the working memory (WM) LR phase encoding imaging data:

tfMRI_WM_LR

```
100307_3T_BIAS_32CH.nii.gz
100307_3T_BIAS_BC.nii.gz
100307_3T_SpinEchoFieldMap_LR.nii.gz
100307_3T_SpinEchoFieldMap_RL.nii.gz
100307_3T_tfMRI_WM_LR.nii.gz
100307_3T_tfMRI_WM_LR_SBRef.nii.gz
```

and the relevant E-Prime data are in appropriate subdirectories:

```
tfMRI_WM_LR/LINKED_DATA/EPRIME
100307_3T_REC_run2_TAB.txt
100307_3T_WM_run2_TAB.txt
```

tfMRI_WM_LR/LINKED_DATA/EPRIME/EVs

```
0bk_body.txt
0bk_cor.txt
0bk_err.txt
0bk_faces.txt
0bk_nlr.txt
etc....
```

and the relevant Physiological data are in the PHYSIO subdirectory:

```
tfMRI_WM_LR/LINKED_DATA/PHYSIO/100307_3T_tfMRI_WM_LR_Physio_log.txt
```

Note that inclusion of an ancillary file in any given directory does not necessarily mean that it is used as part of the current HCP preprocessing pipelines. For example, the BIAS and AFI scans are not used currently (for any modality) for the preprocessed data.

The 6 dMRI scans (3 b-values x two phase encoding directions) are in a single Diffusion/ subdirectory:

Diffusion/

```
100307_3T_BIAS_32CH.nii.gz
100307_3T_BIAS_BC.nii.gz
100307_3T_DWI_dir95_LR.bval
100307_3T_DWI_dir95_LR.bvec
100307_3T_DWI_dir95_LR.nii.gz
100307_3T_DWI_dir95_LR_SBRef.nii.gz
100307_3T_DWI_dir95_RL.bval
100307_3T_DWI_dir95_RL.bvec
100307_3T_DWI_dir95_RL.nii.gz
100307_3T_DWI_dir95_RL_SBRef.nii.gz
100307_3T_DWI_dir96_LR.bval
100307_3T_DWI_dir96_LR.bvec
100307_3T_DWI_dir96_LR.nii.gz
100307_3T_DWI_dir96_LR_SBRef.nii.gz
100307_3T_DWI_dir96_RL.bval
100307_3T_DWI_dir96_RL.bvec
100307_3T_DWI_dir96_RL.nii.gz
100307_3T_DWI_dir96_RL_SBRef.nii.gz
100307_3T_DWI_dir97_LR.bval
100307_3T_DWI_dir97_LR.bvec
100307_3T_DWI_dir97_LR.nii.gz
100307_3T_DWI_dir97_LR_SBRef.nii.gz
100307_3T_DWI_dir97_RL.bval
100307_3T_DWI_dir97_RL.bvec
100307_3T_DWI_dir97_RL.nii.gz
100307_3T_DWI_dir97_RL_SBRef.nii.gz
```

Directory structure for preprocessed MR data

Preprocessed datasets downloaded from ConnectomeDB unpack into subdirectories under the <subject_id>/ parent directory (by creating them if not present, or placing into existing directories if present). If you have ordered Connectome in a Box or are using HCP data in the AWS S3 bucket, the directory structure is the same as the unpacked downloaded data.

Preprocessing generates thousands of files, many of which are of little or no use. Therefore, in the preprocessed dataset package for each subject HCP only releases a subset of files that are

likely to be of general use to investigators. A list of file names in each directory and subdirectory is provided in [Appendix 3B](#).

Structural data (both MSM-Sulc and MSM-All registered)

- **T1w/** contains T1w and T2w volume data
- **T1w/Native/** contains FreeSurfer surfaces in their native mesh and original dimensions after rigid-body rotation to AC-PC alignment.
- **MNINonLinear/** contains cortical surfaces and other data volumetrically registered to MNI152 space (using nonlinear FNIRT) followed by surface registration to Conte69 '164k_fs_LR' mesh ([Van Essen et al. 2012b](#)) (via FreeSurfer fsaverage as an intermediate). A Connectome Workbench-readable *164k_fs_LR.wb.spec file is included for quickly reading and visualizing many of these files in Workbench.
- **MNINonLinear/Native/** replicates some of the files in T1w/Native/ but contains additional files used during surface-based registration.
- **MNINonLinear/xfms/** contains files encoding the transformation between acpc and MNINonLinear volumetric space.
- **MNINonLinear/fsaverage_LR32k/** contains files spatially downsampled to a 32k mesh (average vertex spacing of ~2 mm), which is useful for analyses of rfMRI and dMRI connectivity data. As above, a Connectome Workbench-readable *32k_fs_LR.wb.spec file is included for quickly reading and visualizing many of these files in Workbench.

fMRI data

- **MNINonLinear/Results/** contains volumetric and CIFTI grayordinates data for rfMRI scans (15 min each), motion parameters, and physiological data in four subdirectories,
 - rfMRI_REST1_RL
 - rfMRI_REST1_LR
 - rfMRI_REST2_RL
 - rfMRI_REST2_LR

plus volumetric and CIFTI grayordinates data and physiological data for 7 pairs of tfMRI scans (each task run once with right-to-left and once with left-to-right phase encoding).

- tfMRI_EMOTION_RL
- tfMRI_EMOTION_LR
- tfMRI_GAMBLING_RL
- tfMRI_GAMBLING_LR
- tfMRI_LANGUAGE_RL

- tfMRI_LANGUAGE_LR
- tfMRI_MOTOR_RL
- tfMRI_MOTOR_LR
- tfMRI_RELATIONAL_RL
- tfMRI_RELATIONAL_LR
- tfMRI_SOCIAL_RL
- tfMRI_SOCIAL_LR
- tfMRI_WM_RL
- tfMRI_WM_LR

The *.d[type].nii files in these directories are in CIFTI format (based on NIFTI-2, hence the .nii extension), which can be directly read by the [Connectome Workbench software](#), in which results from other modalities (structural-MR , fMRI, dMRI, MEG) can be also read and overlaid.

Each of the directories above contain an .fsf file and a “EVs” directory containing explanatory variables. These files can be used to run first-level analyses in the FSL 5 program FEAT. The directories above also contain a *TAB.txt file containing the timing of events, so that researchers might create their own explanatory variables of interest.

- **Motion parameters.** Estimates of motion parameters are saved into two different files: Movement_Regressors.txt and Movement_Regressors_dt.txt. The first file (Movement_Regressors.txt) contains 12 variables. The first six variables are the motion parameters estimates from a rigid-body transformation to the SBRef image acquired at the start of each fMRI scan.

- trans_x (mm)
- trans_y (mm)
- trans_z (mm)
- rot_x (deg)
- rot_y (deg)
- rot_z (deg)

The second six variables are temporal derivatives of those motion parameters

- trans_dx
- trans_dy
- trans_dz
- rot_dx
- rot_dy
- rot_dz

The second file (Movement_Regressors_dt.txt) contains 12 variables derived by removing the mean and linear trend from each variable in Movement_Regressors.txt

- The **MNINonLinear/Results/** directory also contains seven other directories, one for each task:
 - tfMRI_EMOTION
 - tfMRI_GAMBLING
 - tfMRI_LANGUAGE
 - tfMRI_MOTOR
 - tfMRI_RELATIONAL
 - tfMRI_SOCIAL
 - tfMRI_WM

In the preprocessed data, these directories contain an .fsf file that can be used to run analysis across the two runs of each task.

Note: HCP's results of individual (within-subject) tfMRI analysis are available in “analysis” download packages [separate](#) from the preprocessed datasets. When unpackaged, these data will integrate into these task-specific directories (MNINonLinear/Results/...). See [Individual \(within-subject\) tfMRI analysis results data](#) and [Appendix 3D](#) for more details.

Diffusion Data

- Diffusion data includes diffusion weighting, direction, time series, brain mask, and gradient nonlinearity data, with the following files and directory structure:
 - **T1w/** contains T1w_acpc_dc_restore_1.25.nii.gz (structural volume sampled at the same resolution as the diffusion data)
 - **T1w/Diffusion** contains the following subdirectories and files:
 - bvals (contains the diffusion weighting (b-value) for each volume)
 - bvecs (contains the diffusion direction (b-vector) for each volume)
 - data.nii.gz (preprocessed diffusion time series file)
 - nodif_brain_mask.nii.gz (brain mask in diffusion space)
 - grad_dev.nii.gz (contains the effects of gradient nonlinearities on the bvals and bvecs for each voxel)

ICA-FIX denoised rfMRI data

Here, we describe the approach we took to generate resting-state fMRI data that has been cleaned of structured noise through a process that pairs independent component analysis (MELODIC) with the FSL tool FIX, that automatically removes artifactual or “bad” components.

The ICA-FIX denoised rfMRI data is available as additionally processed datasets for HCP subjects. For further details on this method see [Smith et al., 2013](#); [Salimi-Khorshidi et al., 2014](#), and [Griffanti et al., 2014](#).

For rfMRI data, the HCP investigated the use of independent component analysis (ICA)-based artifact removal, to remove non-neural spatiotemporal components from each (highpass filtered) 15-minute run of rfMRI data. ICA is a powerful approach for decomposition of fMRI data as a summation of “good” and “bad” components, where each component comprises a weighted set of voxels (the component's spatial map), along with a single timeseries that is common to those voxels identified ([Beckmann and Smith, 2004](#)). Once ICA has identified a number of artifactual components, the data can be “cleaned” by subtracting these components from the data.

In working with the high quality/high temporal quantity HCP data, HCP investigators have developed a new approach that combines ICA with a more complex automated component classifier referred to as FIX (FMRIB's ICA-based X-noisifier) that has been specifically trained on HCP data. The FIX approach and initial results of classification accuracy are described in detail in ([Salimi-Khorshidi et al., 2014](#)), and the effects of the ICA + FIX cleanup (and optimal methods to remove the bad components from the data) are evaluated in detail in ([Griffanti et al., 2014](#)).

For HCP data, ICA is run using FSL's MELODIC with automatic dimensionality estimation (MELODIC estimates how many components the given quality and quantity of data will support being separated from each other); this dimensionality is limited to a maximum of 250. These components are fed into FIX, which classifies components into “good” vs. “bad”. Bad components are then removed from the data. All of this is run using the volumetric data, rather than the grayordinate version of the data, because many artifacts are inherently 3D and do not respect tissue boundaries. The same set of artifactual processes is then removed from the (already created) grayordinates version of the data, by first applying the same highpass temporal filtering, and then regressing the bad components' timeseries out. For both volume and surface cleanup, the cleanup is done in a “non-aggressive” manner — both the good and bad component timeseries are regressed into the data, and then the resulting bad spatial maps are multiplied by the associated timeseries and subtracted from the original dataset. Thus, in this non-aggressive approach, only the unique variance associated with the bad components is removed from the data. Applying “aggressive” cleanup means removing all variance associated with the bad components, and not just the unique part, relative to non-artifact component timeseries. We have taken the more conservative non-aggressive approach to avoid removing variance of interest from the data, with the understanding that our cleanup will be less effective for more global types of noise whose variance is shared across good and bad components (this decision will be revisited in future cleanup investigations). As part of this cleanup, we also used 24 confound timeseries derived from the motion estimation (the 6 rigid-body parameter timeseries, their backwards-looking temporal derivatives, plus all 12 resulting regressors squared — [Satterthwaite et al., 2013](#)). The motion parameters have the temporal highpass

filtering applied to them and are then regressed out of the data aggressively, as they are not expected to contain variance of interest.

Data from 25 HCP subjects (100 rfMRI runs) were hand-labeled, to train FIX. The average number of components per 15-minute run estimated by ICA was 229; of these, on average 24 components were hand-classified as “good” and the remainder as “bad”. Leave-one-subject-out testing of the classifier (i.e., leaving one subjects' 4 runs out, training FIX on the other 19 subjects, and testing on the 4 runs left out) resulted in a mean accuracy of 99.3% in identifying “good” components correctly, and 99% accuracy in identifying “bad” components correctly (with corresponding median values being 100% and 99.3%). The balance between these two accuracies can be adjusted through the setting of a single controlling parameter; one has the option to choose to give greater importance to the accuracy of identification of good or of bad components. It is likely that an important factor in the very high accuracy of FIX classification is the high HCP data quality (and quantity — number of voxels and timepoints in each run), meaning that ICA is able to do a good job of separating multiple signal and noise components. See [Smith et al., 2013](#) for examples of good and bad components and results of the cleanup. The cleanup reduces the resting-state network timeseries amplitude by ~ 30%, but, despite that reduction, does not reduce the effective group-level results (as judged by group-level statistics applied to both RSN spatial maps and network matrices).

3T Fix and Fix_extended datasets

Denoised rfMRI data, registered by both MSM-Sulc (cortical folding-based alignment) and MSM-All (based on alignment of folding, myelin maps, and rsfMRI networks), is being released as two different datasets for the 4 rfMRI runs for each subject. The FIX-Denoised (compact) dataset contains ICA-FIX denoised rfMRI grayordinate surface timeseries only. The FIX-Denoised (Extended) rfMRI dataset contains ICA-FIX denoised rfMRI volumetric timeseries, and ICA component data (for evaluation of FIX), plus a “stats.dscalar.nii” and a “RestingStateStats” folder. These files provide information about different types of ‘noise’ and ‘signal’ in HCP resting state data, gleaned by partitioning the variance according to different processing stages in the FIX denoising pipeline. Resting State Stats results are based on MSM Sulc (cortical folding-based only) registered resting state data. For more information about the types of variance displayed in these files, please review the RestingStateStats script in the Github Pipelines: <https://github.com/Washington-University/Pipelines/tree/master/RestingStateStats>.

We recommend using the fix compact package for general analysis of resting state data across many or all HCP subjects. If you would like to evaluate the FIX cleanup, we recommend downloading fix_extended for a limited number of subjects, in addition to fix (compact).

Each of these datasets unpacks into the <SubjectID>/MNINonLinear/Results/ subdirectories for the 4 rfMRI scans. A list of file names in each directory and subdirectory for both datasets is provided in [Appendix 3C](#). As an example, for the REST1_LR run:

rfMRI_REST1_LR_Atlas_hp2000_clean.dtseries.nii is the grayordinates timeseries file and **rfMRI_REST1_LR_hp2000_clean.nii.gz** is the volumetric timeseries file.

The *.dtseries.nii file is in CIFTI format (based on NIFTI-2, hence the .nii extension), which can be directly read by the [Connectome Workbench software](#), in which results from other modalities (structural-MR , fMRI, dMRI, MEG) can be also read and overlaid.

Individual (within-subject) tfMRI analysis data

For the 1200 Subjects Release, as with the 900 Subjects Release, HCP is only releasing the grayordinate (surface and subcortical volume)-based results of individual (within-subject, level 2) tfMRI analysis for both MSM-Sulc and MSM-All registered datasets. Volume-smoothed tfMRI level 2 tfMRI analysis results are no longer being distributed due to its inherent drawbacks in statistical power and spatial resolution compared to surface-smoothed data (see [Glasser et al. 2013](#); [Barch et al. 2013](#)). Investigators who in previous releases have used the volume-smoothed data should be aware of these limitations and are encouraged instead to use surface smoothing and the CIFTI format whenever feasible (implementation of CIFTI-compatibility for tfMRI analyses in FSL is in progress).

The grayordinate-based tfMRI analysis data are available at two total smoothing levels: 2 mm and 4 mm for both MSM-Sulc (registration based on cortical folding) and MSM-All (registration based on alignment of folding, myelin maps, and rsfMRI networks) are available in download packages for each of the seven tasks. When unpackaged, these data will integrate into the fMRI directory structure (**MNINonLinear/Results/...**). See [Appendix 3D](#) for a list of file names in the directory and subdirectory structure.

Previously (up through the 500 Subjects Release), HCP distributed the grayordinate-based tfMRI analysis data for MSM-Sulc registered data at several total smoothing levels: 2 mm, 4 mm, 8 mm, and 12mm -smoothed versions and volume-smoothed data at a single 4 mm smoothing level. Below, the grayordinates-based analysis pipeline is described. For information on how the volume-smoothed data were created see the [500 Subjects + MEG2 Reference Manual](#).

In the preceeding fMRISurface pipeline, all HCP fMRI grayordinates data are minimally smoothed with surface (cortical) and parcel constrained (subcortical) smoothing of 2mm FWHM to regularize the mapping process to match the 2mm average vertex spacing of the 32k_fs_LR mesh. Therefore, the 2mm smoothed grayordinates dataset is considered “minimally smoothed” and smoothing of the grayordinates data beyond 2mm total smoothing is considered “additional smoothing”.

The within-subject fixed-effects grayordinate-based analyses generate output files that index mean effects for an individual subject averaged across the two scan runs for a specific task. The output files are placed in the .feat subdirectory of the per task (not per run) directories for each task, e.g. **MNINonLinear/Results/tfMRI_EMOTION/tfMRI_EMOTION_hp200_s4_level2.feat**

The grayordinate-based output of a within-subject task fMRI was performed as follows. First, the minimal preprocessing pipeline handled motion correction, distortion correction, registration to standard space, and generation of a grayordinate time series for each individual task run. Then the grayordinate data for individual task runs was processed in a "Level 1" analysis that included high-pass filtering at 200s, and additional smoothing (for the 4mm dataset) to bring the total smoothing levels to 2 mm or 4 mm FWHM. The additional smoothing for the 4mm dataset was conducted on the surface for the surface-based grayordinate data, and was constrained by segmentation (parcel) borders for the subcortical grayordinate data. Model estimation and temporal autocorrelation was handled using a version of FSL's 'film_gls' equipped to handle surface-based data in its internal spatial smoothing. The two runs for each task and subject were then combined in a "Level 2" fixed-effects analysis. The results of this "Level 2" analysis are contained in the task fMRI analysis dataset for each task and subject. These outputs are in standard HCP grayordinate space and are output in CIFTI format (e.g. file name for 2mm grayordinates Emotion task analysis:

<SubjectID>_tfMRI_EMOTION_level2_hp200_s2.dscalar.nii).

The *.d[type].nii files in these directories are in CIFTI format (based on NIFTI-2, hence the .nii extension), which can be directly read by the [Connectome Workbench software](#), in which results from other modalities (structural-MR , fMRI, dMRI, MEG) can be also read and overlaid.

Group-average functional MR data

Coming soon: R1000+ Group-average

A number of processing pipelines are currently being implemented and refined by the HCP that make use of the minimally processed data to carry out further analyses at the individual and group levels. With data from more subjects and improved analysis methods we are updating the group-average functional dataset we are distributing to include analyses on a larger group of 1000+ subjects, including many subjects that are related to at least one other subject in the group (R1000+). The results of this analysis are not currently available for release with the main 1200 Subjects Release, therefore we are continuing to distribute group-average data for the 800+ subjects, including many subjects that are related (R800+). The following describes those datasets, which have been available since just after the 900 subjects release and are based on both MSM-Sulc and MSM-All registered datasets.

820 subjects (R820) with complete resting state fMRI data and 787 subjects (R787) with complete task fMRI data for all 7 tasks, are included in the R800+ group-averages for these modalities. One can view the subjects included in these analyses using the “Open group” function on the ConnectomeDB dashboard.

The group-average data are available for download through the links on the ConnectomeDB splash page. This data includes:

- R881 group-average structural scans (T1w, T2w), cortical surfaces, and myelin maps (R881 data is being provided as a structural atlas due to its larger size)
- R787 group-average task fMRI analysis results for the 7 HCP tasks and many task contrasts
- CIFTI format files compatible for viewing and comparison in Workbench v1.0 and later.
- Group-average functional connectivity matrix (“dense” functional connectome, the grayordinate \times grayordinate full correlation matrix), computed with and without mean gray timecourse regression (MGTR), for the R820 group (available as separate download packages due to their size).

Because of their large size (33 GB each) these two dense functional connectome files are released separately from the rest of the group average data. As an alternative see the information below about viewing the data remotely in Connectome Workbench.

- Group-PCA Eigenmaps (2 GB each):

HCP_S900_R820_rfMRI_MSMSulc_groupPCA_d4500_Eigenmaps.dtseries
HCP_S900_R820_rfMRI_MSMAII_groupPCA_d4500_Eigenmaps.dtseries

The Eigenmaps can be used as input to group-ICA. They can also be used to generate the dense connectome, but to do this optimally is not trivial, and requires following the procedures outlined in Section IV of the [mound-and-moat document](#). The code for performing these steps is included in the [S900 PTN release zip file](#).

- ***If you prefer to view these dense connectome files in Connectome Workbench (recommended), you do not need to download them.*** The data are accessible in Workbench by remote access (requires internet connection and ConnectomeDB login), using the following URLs:

https://db.humanconnectome.org/spring/cifti-average?resource=HCP_Resources:GroupAvg:HCP_S900_820_rfMRI_MSMSulc_groupPCA_d4500ROW_zcorr
https://db.humanconnectome.org/spring/cifti-average?resource=HCP_Resources:GroupAvg:HCP_S900_820_rfMRI_MSMAII_groupPCA_d4500ROW_zcorr

Note: These URLs are preloaded in scenes 4-5 in the HCP_S900_GroupAvg_v1.scene file distributed as part of the **HCP_S900_GroupAvg_v1** archive available for download under “Connectome Workbench Data” on the [ConnectomeDB](#) splash page. A tutorial for quickly viewing the Group-average S900 data in the Workbench GUI, a.k.a. wb_view, is included as a PDF file in the archive.

The dataset is customized for viewing in the latest version of Connectome Workbench (<http://www.humanconnectome.org/connectome/get-connectome-workbench.html>), but most of the files are in standard NIFTI and GIFTI format and can also be viewed using other brain-mapping software platforms. A number of files, including the task fMRI files and the rfMRI functional connectivity files, are in the new CIFTI-2.0 file format (<http://www.nitrc.org/projects/CIFTI/>) that can incorporate surface vertices (both hemispheres) plus subcortical gray-matter voxels into a single file format. The CIFTI format is also used to represent surface maps (e.g., myelin maps) for both the left and right hemispheres in a single *.dscalar.nii CIFTI file type and cortical parcellations for both hemispheres are contained in a single *.dlabel.nii CIFTI file. The scenes file for Workbench distributed as part of this dataset makes use of these CIFTI files, but the downloaded datasets also include conventional hemisphere-specific GIFTI files (*.shape.gii; *.label.gii).

Task fMRI group analysis

For the R787 group-average task analysis, each of the seven HCP tasks was processed using FLAME 1 (FMRIB's Local Analysis of Mixed Effects stage 1) mixed effects applied across subjects. We are currently releasing the grayordinates-based results that incorporate 2mm smoothing that respects the topology of the cortical sheet and the boundaries of subcortical anatomical parcels (see below and [Barch et al., 2013](#) for details).

The grayordinates-based analysis began with outputs of the HCP “fMRISurface” pipeline (see above) in which the data from the cortical gray matter ribbon were projected onto the surface and then onto registered surface meshes with a standard number of vertices. Subcortical data were also projected to a set of subcortical grey matter parcel voxels, and when combined with the surface data formed the standard grayordinates space ([Glasser et al. 2013](#)). Smoothing of the left and right hemisphere time series and autocorrelation estimates (from FILM) were done on the surface using a geodesic Gaussian algorithm. Subcortical gray matter time series were smoothed within defined gray matter parcels to 2mm FWHM. Activity estimates were computed for the preprocessed functional time series from each run using a general linear model (GLM) implemented in FSL's FILM ([Woolrich et al., 2001](#)). Predictors were convolved with a double gamma “canonical” hemodynamic response function ([Glover, 1999](#)) to generate the main model regressors. To compensate for slice-timing differences and variability in the HRF delay across regions, temporal derivative terms derived from each predictor were added to each GLM and were treated as confounds of no interest. Subsequently, both the 4D time series and the GLM design were temporally filtered with a Gaussian-weighted linear highpass filter with a (soft)

cutoff of 200 seconds. Finally, the time series was prewhitened within FILM to correct for autocorrelations in the fMRI data. Surface-based autocorrelation estimate smoothing was incorporated into FSL's FILM at a sigma of 5mm.

The GLM Model Design for each task is described in [Barch et al., 2013](#). Fixed-effects analyses were conducted using FSL's FEAT to estimate the average effects across runs within-subjects. Mixed-effects analyses treating subjects as random effects were conducted using FSL's FLAME to estimate the average effects of interest for the group. The grayordinates-based group-level cross-run and cross-subject statistical comparisons occurred in the standard grayordinates space. NIFTI-1 matrices were processed separately for left and right surface and subcortical volume data, and surface outputs were converted to GIFTI at the conclusion of analysis. Subject-level and group-level z-statistic maps were combined from left and right hemisphere cortical and subcortical gray matter into the CIFTI data format (<http://www.nitrc.org/projects/CIFTI/>; see [\(Glasser et al. 2013\)](#) for visualization using the Connectome Workbench platform [\(Marcus et al., 2013\)](#)).

Important note about the Related 787 (R787) dataset. As noted above, many of the subjects in the R787 dataset are related. However, family structure was not taken into account in the fMRI group average analyses of R787. Therefore, the associated standard errors and p-values are optimistic. Since the p-values are not valid, any false discovery rate (FDR) inferences based on them will not be valid either. We plan for standard errors and p-values to be revised in future related datasets once family structure is incorporated into the analysis.

Resting-state group analysis and functional connectivity

Resting-state fMRI data from the R820 group of subjects were analyzed using steps outlined below that have yielded MSM-Sulc and MSM-All versions of a group-average functional connectivity matrix ("dense" functional connectome). Both registration versions of the dense connectome use the full correlation matrix.

For group average analysis, each of the four resting state runs for each subject in the group was cleaned (denoised) using 24-parameter motion regression followed by ICA+FIX denoising ([Smith et al., 2013](#); [Salimi-Khorshidi et al., 2014](#)). The resulting cleaned runs were combined across the 468 related subjects respectively using variance normalization of the timeseries (using the same approach as MELODIC [\(Beckmann and Smith, 2004\)](#)), followed by a group-PCA approach (MIGP - MELODIC's Incremental Group-PCA) that approximates full temporal concatenation of all subjects' data followed by PCA, outputting the strongest 4500 spatial eigenvectors (PCA components, weighted by the eigenvalues).

Important Note: When using seed-based methods to explore group-average functional connectivity data, some seeds have a "mound" of high correlation surrounded by a "moat" of low correlation. The HCP is currently working on ways to reduce this "mound and moat" or

“ripple” pattern by modifying the interpolation methods used during image reconstruction and preprocessing and other methods; for additional information, see <http://www.humanconnectome.org/documentation/mound-and-moat-effect.html>.

File Names and Directory Structure for Group-average fMRI Data

Download the group-average functional dataset by clicking the “Download Group Average Data” button under “Connectome Workbench Data” on the [ConnectomeDB](#) 1200 Subjects project page. The dataset archive unpacks to **HCP_S900_GroupAvg_v1/**. It includes the following files:

- Group-average R881 surface/volume data and myelin maps generated from the T1w and T2w images.
- Group-average MSM-Sulc and MSM-All Task fMRI grayordinates-based FEAT-level 3 analyzed files (CIFTI format) at smoothing levels 2mm and 4mm:
 - HCP_S900_787_tfMRI_ALLTASKS_level3_zstat1_hp200_s2_MSMAII.dscalar.nii
 - HCP_S900_787_tfMRI_ALLTASKS_level3_zstat1_hp200_s2_MSMSulc.dscalar.nii
 - HCP_S900_787_tfMRI_ALLTASKS_level3_zstat1_hp200_s4_MSMAII.dscalar.nii
 - HCP_S900_787_tfMRI_ALLTASKS_level3_zstat1_hp200_s4_MSMSulc.dscalar.nii

These are CIFTI files in which minimal spatial smoothing (2mm FWHM) was constrained to 'grayordinates' (cortical surface vertices and subcortical voxels). By avoiding spatial blurring across anatomical compartments (gray matter vs white matter and CSF) and across sulcal banks, this provides the highest spatial fidelity attainable using these data.

The *.d[type].nii files in these directories are in CIFTI format (based on NIFTI-2, hence the .nii extension), which can be directly read by the [Connectome Workbench software](#), in which results from other modalities (structural-MR , fMRI, dMRI, MEG) can be also read and overlaid.

- Reference files generated from other studies:
 - Gordon333_Key.txt
 - Gordon333.32k_fs_LR.dlabel.nii
 - Human.Broadmann09.32k_fs_LR. dlabel.nii
 - Human.Composite_VGD11.32k_fs_LR.dlabel.nii
 - Human.MedialWall_Conte69.32k_fs_LR. dlabel.nii
 - RSN-networks.32k_fs_LR.dlabel.nii

The Human.Composite_VGD11.32k_fs_LR.dlabel.nii file is derived from a composite cortical parcellation containing 52 distinct areas accurately mapped to the fs_LR atlas surface and based on architectonic or retinotopic fMRI maps ([Van Essen et al. 2012](#)). Abbreviations

associated with labels of cortical areas (FRB08, OFP03, etc.) refer to the publication that defined a particular cortical area in the composite map (cf. Table 3 of [Van Essen et al. 2012](#)).

The Gordon333.32k_fs_LR.dlabel.nii and Gordon333_Key.txt are derived from a cortical parcellation generated using resting state functional connectivity boundary maps to identify putative borders between cortical areas [Gordon et al. 2014](#).

The RSN-networks.32k_fs_LR.dlabel.nii file displays resting state network cortical parcellation maps from [Yeo et al. 2011](#) (7 and 17 network maps) and the Resting State network consensus communities (with and without gaps in the communities filled) from [Power et al. 2011](#).

- File for easily opening S900 Group Average data in Connectome Workbench:
 - HCP_S900_GroupAvg_v1.scene

Release Notes and Tutorial files:

- ReleaseNotes_HCP_S900_GroupAvg_v1.txt
- ConnectomeWorkbenchTutorial_HCP_S900_GroupAvg_v1.pdf

Parcellation-Timeseries-Netmats (PTN) extensively processed rfMRI data

Coming soon: R1000+ PTN data

With data from more subjects we are updating the rfMRI Parcellation-Timeseries-Netmats (PTN) extensively processed rfMRI dataset we are distributing to include analyses on a larger group of 1000+ subjects, including many subjects that are related to at least one other subject in the group (R1000+). The results of this analysis are not currently available for release with the main 1200 Subjects Release, therefore we are continuing to distribute PTN data for the R820 group (includes many subjects that are related). The following describes those datasets, which have been available since just after (January 2016) the 900 subjects release and are based on MSM-All registered datasets.

Resting-state fMRI (rfMRI) data from the 900 subjects 2015 HCP data release was processed, using data from all 820 subjects having four complete rfMRI runs, yielding the following outputs:

1. Group-average “parcellations”, obtained by means of group-ICA.
2. Subject-specific sets of “node timeseries” – for each subject, a representative time series per ICA component (“parcel”).

3. A subject-specific “parcellated connectome” – for each subject, a nodes x nodes matrix – the functional connectivity between node timeseries.

In addition to the descriptive text and citations below, which may be useful when writing papers that take advantage of these “higher-level” HCP outputs, please remember to include the generic HCP acknowledgements (and core HCP citations) when using HCP data in your research: see <http://www.humanconnectome.org/documentation/citations.html>

Data pre-processing and Group-ICA

Each 15-minute run of each subject's fMRI data was preprocessed according to [Smith 2013]; it was minimally-preprocessed [Glasser 2013], and had artefacts removed using ICA+FIX [Salimi-Khorshidi 2014, Griffanti 2014]. Each dataset was then temporally demeaned and had variance normalisation applied according to [Beckmann 2004].

Group-PCA output was generated by MIGP (MELODIC's Incremental Group-PCA) from 468 subjects. This comprises the top 4500 weighted spatial eigenvectors from a group-averaged PCA (a very close approximation to concatenating all subjects' timeseries and then applying PCA) [Smith NeuroImage 2014].

The MIGP output was fed into group-ICA using FSL's MELODIC tool [Hyvärinen 1999, Beckmann 2004], applying spatial-ICA at several different dimensionalities (15, 25, 50, 100, 200, 300). The dimensionality determines the number of distinct ICA components; a higher number typically means that the significantly modulated areas within the spatial component maps will be smaller. Spatial-ICA was applied in grayordinate space (surface vertices plus subcortical grey matter voxels, [Glasser 2013]). The spatial-ICA maps are released as a separate ‘dscalar’ grayordinates file (CIFTI format) for each distinct ICA decomposition (thus each file contains between 25 and 300 spatial maps). Volumetric MNI152 3D-space versions of these maps were also generated, primarily for display purposes. The sets of ICA maps can be considered as “parcellations”, though they lack some properties often assumed for parcellations - for example, ICA maps are not binary masks but contain a continuous range of “weight” values; they can overlap each other; and a given map can include multiple spatially separated peaks/regions. Group parcellations utilizing other parcellation methods will be released in the future.

Node timeseries (individual subjects)

For a given “parcellation” (group-ICA decomposition), the set of ICA spatial maps was mapped onto each subject's fMRI timeseries data to derive one representative timeseries per ICA component (for these purposes we consider each ICA component as a network “node”). For each subject, these 15 (or 25, 50, 100, 200 or 300) timeseries can then be used in network analyses, as described below. One method was used to estimate the node-timeseries - the

standard "dual-regression stage-1" approach, in which the full set of ICA maps was used as spatial regressors against the full data, estimating one timeseries for each ICA map [Filippini 2009].

Node timeseries – saved as ASCII text files - are estimated for the 820 subjects having complete fMRI data (4800 total timepoints).

Network matrices (individual subjects and group-averaged)

Network-matrices (also referred to as "netmats" or "parcellated connectomes") were derived from the node-timeseries. For each subject, the N (25-300) node-timeseries were fed into network modelling, creating an NxN matrix of connectivity estimates. Network modelling was carried out using the FSLNets toolbox (fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLNets). Netmats were estimated for the 820 subjects (complete timeseries) described above. We applied network modelling in two ways:

1. netmats1: Using "full" normalized temporal correlation between every node timeseries and every other. This is a common approach and is very simple, but it has various practical and interpretational disadvantages [Smith 2012].
2. netmats2: Using partial temporal correlation between nodes' timeseries. This aims to estimate direct connection strengths better than achieved by full correlation. To slightly improve the estimates of partial correlation coefficients, a small amount of L2 regularization is applied (setting $\rho=0.01$ in the Ridge Regression netmats option in FSLNets) [Smith OHBM 2014, FSLNets].

Netmat values were Gaussianised from Pearson correlation scores (r-values) into Z-stats, and are released for individual subjects, as well as averaged over all 820 subjects. The average netmats are provided as "pconn" files, readable by HCP workbench software (and directly viewable in the wb_view workbench display tool). The individual subjects' netmats are saved as raw text files, with one row per subject; each row contains the NxN matrix of connectivity estimates, unwrapped to a long single row of NxN values. The row ordering matches the list of subject IDs saved in the file subjectIDs.txt

Subject-measure predictions/correlations and heritability estimation

The "HCP MegaTrawl" is a large set of analyses based on partial correlation netmats, derived from all subjects with complete fMRI (see [Smith 2013] for an example). Multivariate analyses seek to model (across subjects) a given "subject measure" (e.g., one of the behavioural variables such as fluid intelligence), finding a set of edge weights in the HCP data that can partially explain that subject measure. Results of this analysis for all the HCP subject measures are presented as web pages with multivariate prediction and univariate regression results for each subject measures, and thumbnail volume images showing the edges (node-pairs) whose

connection most strongly correlates with the variable. We also calculate and present heritability estimates for the netmats. The 820-subject MegaTrawl is currently available – this will be updated soon with 1000+ subjects.

Description of released files

subjectIDs.txt Ordered list of the subjects included in this PTN release

groupICA_3T_HCP820_MSMAII.tar.gz :

Group-ICA "parcellations" at several dimensionalities (levels of detail):

| | | |
|---------------------------|--|------------------------|
| melodic_IC.dscalar.nii | ICA spatial maps (unthresholded Zstats); one "timepoint" per map. | Grayordinates |
| melodic_IC_ftb.dlabel.nii | Summary "find the biggest" labels image for all ICA spatial maps. | Grayordinates |
| melodic_IC_sum.nii.gz | ICA spatial maps projected from the surface into the volume, for reporting purposes. | MNI152 space |
| melodic_IC_sum.sum | Summary "thumbnail" PNG images created at the most relevant axial slices(s). | Slices of MNI152 space |

NodeTimeseries_3T_HCP820_MSMAII_ICAd*_ts*.tar.gz :

Node-timeseries, with one tarfile for each choice of group-ICA dimensionality and timeseries estimation approach

Inside each tarfile there is one timeseries text file per subject (concatenated across all 4 runs)

Within each text file there is one column per "node" (ICA component)

ICAd25 (etc) describes the original group-ICA dimensionality

ts2: multiple regression (against the set of ICA spatial maps) is used to estimate node timeseries (same as first stage of "dual regression")

ts3: eigentimeseries is used to estimate node timeseries; first PCA component within a spatial map is used

netmats_3T_HCP820_MSMAII_ICAd*_ts*.tar.gz :

Netmats (parcellated connectomes), with one tarfile for each choice of group-ICA dimensionality and node timeseries estimation approach

| | |
|-------------------|---|
| *_netmat1/* | One netmat file per subject, computed using full correlation, Z-transformed. |
| *_netmat2/* | As above, but using partial correlation with modest Tikhonov regularisation ("ridgep" in FSLNets with parameter 0.01) |
| *_Znet1.pconn.nii | Group-level netmat; a one-group t-test across all subjects' full correlation netmats, Z-transformed |
| *_Znet2.pconn.nii | As above, but for partial correlations |

Example simple MATLAB/FSLNets code

```
BATCH='3T_Q1-Q6related468_MSMSulc'; % main name of this data release
d=25; % ICA dimensionality
```

```
M='ts3'; % node-timeseries method

% load in node timeseries
ts_dir=sprintf('node_timeseries/%s_d%d_%s',BATCH,d,M);
ts=nets_load(ts_dir,0.72,1,1);

% view mean spectra
ts_spectra=nets_spectra(ts);

% estimate subjectwise netmats
netmat1=nets_netmats(ts,1,'corr'); % full correlation
netmat2=nets_netmats(ts,1,'ridgep',0.01); % regularised partial correlation

% estimate groupwise netmats
[Znet1,Mnet1]=nets_groupmean(netmat1,0,1);
[Znet2,Mnet2]=nets_groupmean(netmat2,0,1);

% view hierarchical clustering of nodes
SUMPICS=sprintf('groupICA/groupICA_%s_d%d.ica/melodic_IC_sum',BATCH,d);
[hier,linkages]=nets_hierarchy(Znet1,Znet2,ts.DD,SUMPICS);
set(gcf,'PaperPositionMode','auto','Position',[1 1 46*(ts.Nnodes+1) 1574 ]);
```

Directory structure for unprocessed MEG data

Unprocessed datasets downloaded from ConnectomeDB are organized into subject-specific and scan-specific archives. A full list of file names in each directory and subdirectory is provided in [Appendix 3: File Names and Directory Structure for 1200 Subjects Data, Section E](#).

Unprocessed data for exemplar subject 012345 unpacks to the following directory structure:

```
012345/unprocessed/MEG/
  1-Rnoise/4D
  2-Pnoise/4D
  3-Restin/4D
  4-Restin/4D
  5-Restin/4D
  6-Wrkmem/4D
  6-Wrkmem/EPRIME
  7-Wrkmem/4D
  7-Wrkmem/EPRIME
  8-StoryM/4D
  8-StoryM/EPRIME
  9-StoryM/4D
  9-StoryM/EPRIME
  10-Motort/4D
  10-Motort/EPRIME
```


11-Motort/4D
11-Motort/EPRIME

The MEG/ subdirectory signifies that these data were acquired in the MEG lab at Saint Louis University (SLU). Since all subjects will also be scanned at 3T at Wash U, the 3T data (separately available on ConnectomeDB and in Connectome in a Box) will unpack to a 3T/ subdirectory.

Within each subdirectory, files are named as indicated by the exemplars below for several scan types.

The MEG data was acquired in 11 scans. Scan 1 and 2 are empty room (R) and participant (P) noise measurements, scan 3-5 are resting state measurements, scans 6-11 are task measurements. The 4D subdirectory signifies that the data originates from the 4D Magnes MEG scanner. The EPRIME subdirectory contains the log files from the stimulus presentation software (only for the task measurements).

For the noise scans, the 2 scans are downloaded into separate subdirectories (1 and 2), because the initial preprocessing is carried out separately for each scan. For example, for the first noise scan:

1-Rnoise/4D
 config
 c,rfDC

The config file contains additional header information in a mixture of textual and binary representation and the c,rfDC file contains the raw data in binary format (see [Data Conversion: 4D import to FieldTrip](#) above for more information). Note that the two noise scans (1 and 2) do not have head localization data.

For the empty-room and subject noise datasets, the only applicable pipeline is datacheck. Therefore, the noise datacheck pipeline results do not comprise a separate package but are included in the packages for the unprocessed empty room data.

The three (6 min) rMEG scans for each subject are downloaded into separate subdirectories (3-5), because the initial preprocessing is carried out separately for each scan. For example, for the first rMEG scan:

3-Restin/4D
 config
 c,rfDC
 e,rhfp1.0Hz,COH
 e,rhfp1.0Hz,COH1

As above, the config file contains additional header information and the c,rfDC file contains the raw data. The e,rhfp1.0Hz,COH file contains the head localization data at the start of the scan, and the e,rhfp1.0Hz,COH1 file contains the head localization data at the end of the scan.

For tMEG scans, the 6 scans (3 tasks x 2 runs each in subdirectories 6-11) are downloaded into separate subdirectories, because the initial preprocessing is carried out separately for each scan. EPRIME log files for each scan are available in ASCII tab-delimited format (*.tab) and in Microsoft Excel (*.xlsx) format. For example, for the first working memory task scan:

```
6-Wrkmem/4D
    config
    c,rfdc
    e,rhfp1.0Hz,COH
    e,rhfp1.0Hz,COH1

6-Wrkmem/EPRIME
    012345_MEG_Wrkmem_run1.xlsx
    012345_MEG_Wrkmem_run1.tab
```

Directory structure for MEG anatomy data (anatomical models for source reconstruction)

Anatomical model datasets downloaded from ConnectomeDB are organized into subject-specific archives. When unpacked, they are moved into the <subject_id>/MEG/anatomy subdirectory. A full list of file names in each directory and subdirectory is provided in [Appendix 3: File Names and Directory Structure for 1200 Subjects Data, Section F](#).

Anatomical model data for exemplar subject 012345 unpacks to the following directory structure:

```
012345/MEG/anatomy
    012345_MEG_anatomy_transform.txt
    012345_MEG_anatomy_headmodel.mat
    012345_MEG_anatomy_sourcemodel_2d.mat
    012345_MEG_anatomy_sourcemodel_3d4mm.mat
    012345_MEG_anatomy_sourcemodel_3d6mm.mat
    012345_MEG_anatomy_sourcemodel_3d8mm.mat
    012345.L.inflated.4k_fs_LR.surf.gii
    012345.R.inflated.4k_fs_LR.surf.gii
    012345.L.midthickness.4k_fs_LR.surf.gii
    012345.R.midthickness.4k_fs_LR.surf.gii
    T1w_acpc_dc_restore.nii.gz

    provenance/
        012345_MEG_anatomy_transform.txt.xml
        012345_MEG_anatomy_headmodel.mat.xml
        012345_MEG_anatomy_sourcemodel_2d.mat.xml
        012345_MEG_anatomy_sourcemodel_3d4mm.mat.xml
        012345_MEG_anatomy_sourcemodel_3d6mm.mat.xml
        012345_MEG_anatomy_sourcemodel_3d8mm.mat.xml

    figures/
```

012345_MEG_anatomy_headmodel.png
012345_MEG_anatomy_sourcemodel_3d4mm.png
012345_MEG_anatomy_sourcemodel_3d6mm.png
012345_MEG_anatomy_sourcemodel_3d8mm.png
012345_MEG_anatomy_sourcemodel_2d.png

figures/provenance/

012345_MEG_anatomy_headmodel.png.xml
012345_MEG_anatomy_sourcemodel_3d4mm.png.xml
012345_MEG_anatomy_sourcemodel_3d6mm.png.xml
012345_MEG_anatomy_sourcemodel_3d8mm.png.xml
012345_MEG_anatomy_sourcemodel_2d.png.xml

The *transform.txt and *headmodel.mat files contains the volume conduction model (headmodel) based on the anatomical coregistration information. To protect the privacy of HCP subjects we do not release the unmasked 1mm resolution T1w MRI anatomical image that is used to coregister the MEG with the MRI data as they can be identifying. The 3-dimensional grid based source models at different resolutions are contained in the *sourcemodel_3d*mm files. The *sourcemodel2d file contains the sourcemodel defined on the cortical sheet. The *.surf.gii are individual cortical surface GIFTI files for each hemisphere that can be read by Connectome Workbench. The T1w_acpc_dc_restore.nii.gz file is the preprocessed T1w volume from the individual subject's MR data.

Provenance information (in Extensible Markup Language, i.e. *.xml), quality control figures (in Portable Network Graphics format, i.e. *.png) and provenance information for the figures. are contained in the provenance/, figures/, and figures/provenance/ subdirectories, respectively.

Directory structure for Channel- (preprocessed) and Source-level processed MEG data

All channel- and source-level processed MEG data for each subject should unpack to the MEG/ directory under the <SubjectID> directory in a subdirectory for each type of experiment, shown here for exemplar subject 012345:

012345/MEG
 Rnoise
 Pnoise
 Restin
 Wrkmem
 StoryM
 Motort

A complete list of all file names in each directory and subdirectory is provided in [Appendix 3G](#).

Under each of the experimental conditions, the directory structure represents the analysis pipelines that have been executed on the data.

For the empty-room and subject noise datasets, the only applicable pipeline is datacheck. The noise datacheck pipeline results do not comprise a separate package but are included in the packages for the unprocessed empty room data.

For the resting state MEG dataset, the pipelines start with datacheck->baddata->icaclass. [Channel-level rMEG](#) processing continues with rmegpreproc->powavg. [Source-level rMEG](#) processing continues with icamne -> icablpenv->icablpcorr, icamne-> icaimagcoh, and bfblpenv->bfbldpcorr. In addition to dense connectivity datasets, parcellated datasets can also be created by icablpenv, icabldpcorr, icaimagcoh, bfbldpcorr, and bfbldpcorr.

For the three task MEG datasets, the pipelines start with datacheck->baddata->icaclass ->tmeqpreproc. [Channel-level tMEG](#) processing continues with eravg for the Event-Related fields and tfavg for averaged Time-Frequency representations. [Source-level tMEG](#) processing continues with srcavgdics, which uses input from the Eravg pipeline, and srcavgldmv, which uses input from the Tfavg pipeline. In addition to dense connectivity datasets, parcellated datasets can also be created by srcavgdics and srcavgldmv.

Channel-level preprocessed MEG data for exemplar subject 012345 unpacks to the directory structure that is listed below for each of the pipelines. Most pipeline results are accompanied by a portable network graphics (*.png) bitmap file, which summarizes the results for visual inspection. Each of the *.txt, *.mat and *.png data files that are listed below is accompanied with a similarly named *.xml file in the provenance directory, which details the version of the software used to produce the results. These XML files are not all listed below, but are present in the release packages.

Datacheck results

The results of the [Datacheck](#) pipeline for exemplar subject 012345 unpack to the following directory structure:

MEG/Wrkmem/datacheck/

- 012345_MEG_6-Wrkmem_datacheck_info.txt
- 012345_MEG_7-Wrkmem_datacheck_info.txt

figures/

- 012345_MEG_6-Wrkmem_datacheck_MEGREF_powspectrm.png
- 012345_MEG_6-Wrkmem_datacheck_MEG_lowfreq_power.png
- 012345_MEG_6-Wrkmem_datacheck_MEG_powerline_noise.png
- 012345_MEG_6-Wrkmem_datacheck_MEG_powspectrm.png
- 012345_MEG_6-Wrkmem_datacheck_elecchan_ECG.png
- 012345_MEG_6-Wrkmem_datacheck_elecchan_HEOG.png
- 012345_MEG_6-Wrkmem_datacheck_elecchan_VEOG.png

012345_MEG_6-Wrkmem_datacheck_headshape.png
012345_MEG_6-Wrkmem_datacheck_jumps.png
012345_MEG_6-Wrkmem_datacheck_neighb_correlation.png
012345_MEG_6-Wrkmem_datacheck_triggers.png
012345_MEG_7-Wrkmem_datacheck_MEGREF_powspectrm.png
etc.

There are similar results for the other scans, each with the corresponding scan type and number in the directory and in the file names:

MEG/Restin/datacheck/
MEG/StoryM/datacheck/
MEG/Motort/datacheck/

MEG/Rnoise/datacheck/ and **MEG/Pnoise/datacheck/** are contained in the Unprocessed Noise MEG package

Baddata results

The results of [Baddata](#) pipeline for exemplar subject 012345 unpack to the following directory structure:

MEG/Restin/baddata/
012345_MEG_3-Restin_baddata_badchannels.txt
012345_MEG_3-Restin_baddata_badsegments.txt
012345_MEG_3-Restin_baddata_manual_badchannels.txt
012345_MEG_3-Restin_baddata_manual_badsegments.txt
012345_MEG_4-Restin_baddata_badchannels.txt
etc.
figures/
012345_MEG_3-Restin_baddata_badchan_cor_scatter.png
012345_MEG_3-Restin_baddata_badchan_cor_topo3D.png
012345_MEG_3-Restin_baddata_badchan_cor_topo.png
012345_MEG_3-Restin_baddata_badchan_std_scatter.png
012345_MEG_3-Restin_baddata_badchan_std_topo.png
012345_MEG_3-Restin_baddata_icaqc_badchannel_A88.png
012345_MEG_3-Restin_baddata_icaqc_badchannel_A246.png
etc. (# of icaqc_badchannel files/channels varies with scan)

012345_MEG_3-Restin_baddata_icaqc_badsegment_1.png
012345_MEG_3-Restin_baddata_icaqc_badsegment_2.png
012345_MEG_3-Restin_baddata_icaqc_badsegment_3.png
etc. (# of icaqc_badsegment files varies with scan)

012345_MEG_3-Restin_baddata_icaqc_results_1.png
012345_MEG_3-Restin_baddata_icaqc_results_2.png
012345_MEG_3-Restin_baddata_icaqc_results_3.png
etc. (# of icaqc_results files varies with scan)

There are similar results for the other scans, each with the corresponding scan type and number in the directory and in the file names:

MEG/Wrkmem/baddata/
MEG/StoryM/baddata/
MEG/Motort/baddata/

Icaclass and Icclass_qc results

The results of the [Icclass](#) and [Icclass_qc](#) pipelines for exemplar subject 012345 unpack to the following directory structure:

MEG/Restin/icaclass/
 012345_MEG_3-Restin_icaclass.mat
 012345_MEG_3-Restin_icaclass.txt
 012345_MEG_3-Restin_icaclass_vs.txt
 012345_MEG_3-Restin_icaclass.mat
 012345_MEG_3-Restin_icaclass_vs.txt
 012345_MEG_4-Restin_icaclass.mat
 etc.

 figures/
 012345_MEG_3-Restin_icaclass_refch.png
 012345_MEG_3-Restin_icaclass_1.png
 012345_MEG_3-Restin_icaclass_2.png
 012345_MEG_3-Restin_icaclass_3.png
 etc. (# of icaclass files varies with scan)

 012345_MEG_3-Restin_icaclass_vs_1.png
 012345_MEG_3-Restin_icaclass_vs_2.png
 012345_MEG_3-Restin_icaclass_vs_3.png
 etc. (# of icaclass_vs files varies with scan, but should be same # as icaclass files)

There are similar results for the other scans, each with the corresponding scan type and number in the directory and in the file names:

MEG/Wrkmem/icaclass/
MEG/StoryM/icaclass/
MEG/Motort/icaclass/

Rmegpreproc results

The results of the [Rmegpreproc](#) pipeline (only for Resting state scans) for exemplar subject 012345 unpack to the following directory structure:

MEG/Restin/rmegpreproc/
 012345_MEG_3-Restin_rmegpreproc.mat
 012345_MEG_4-Restin_rmegpreproc.mat

012345_MEG_5-Restin_rmegpreproc.mat

Powavg results

The results of the [Powavg](#) pipeline (only for Resting state scans) for exemplar subject 012345 unpack to the following directory structure:

MEG/Restin/powavg/

012345_MEG_3-Restin_powavg.mat
012345_MEG_4-Restin_powavg.mat
012345_MEG_5-Restin_powavg.mat

figures/

012345_MEG_3-Restin_powavg_multiplot.png
012345_MEG_3-Restin_powavg_singleplot.png
012345_MEG_4-Restin_powavg_singleplot.png
012345_MEG_5-Restin_powavg_multiplot.png
012345_MEG_5-Restin_powavg_singleplot.png

Tmegpreproc results

The results of the [Tmegpreproc](#) pipeline (only for Task scans) for exemplar subject 012345 unpack to the following directory structure:

MEG/Wrkmem/tmegpreproc/

012345_MEG_6-Wrkmem_tmegpreproc_TIM.mat
012345_MEG_6-Wrkmem_tmegpreproc_TRESP.mat
012345_MEG_6-Wrkmem_tmegpreproc_trialinfo.mat
012345_MEG_7-Wrkmem_tmegpreproc_TIM.mat
012345_MEG_7-Wrkmem_tmegpreproc_TRESP.mat
012345_MEG_7-Wrkmem_tmegpreproc_trialinfo.mat

There are similar results for the other task scans, each with the corresponding scan type and number in the directory and in the file names:

MEG/StoryM/icaclass/

MEG/Motort/icaclass/

Eravg results

The results of the [Eravg](#) pipeline (only for Task scans) for exemplar subject 012345 unpack to the following directory structure for the Working Memory task:

MEG/Wrkmem/eravg/

012345_MEG_Wrkmem_eravg_[LM-TIM-0B]_[BT-diff]_[MODE-mag].mat
012345_MEG_Wrkmem_eravg_[LM-TIM-0B]_[BT-diff]_[MODE-planar].mat

012345_MEG_Wrkmem_eravg_[LM-TIM-0B-versus-2B]_[OP-diff]_[BT-diff]_[MODE-mag].mat
 012345_MEG_Wrkmem_eravg_[LM-TIM-0B-versus-2B]_[OP-diff]_[BT-diff]_[MODE-planar].mat
 012345_MEG_Wrkmem_eravg_[LM-TIM-2B]_[BT-diff]_[MODE-mag].mat
 012345_MEG_Wrkmem_eravg_[LM-TIM-2B]_[BT-diff]_[MODE-planar].mat
 etc. for all Wrkmem datagroups and contrasts

figures/

012345_MEG_Wrkmem_eravg_[LM-TIM-0B]_[BT-diff]_[MODE-mag]_plot.png
 012345_MEG_Wrkmem_eravg_[LM-TIM-0B]_[BT-diff]_[MODE-planar]_plot.png
 012345_MEG_Wrkmem_eravg_[LM-TIM-0B-versus-2B]_[OP-diff]_[BT-diff]_[MODE-mag]_plot.png
 012345_MEG_Wrkmem_eravg_[LM-TIM-0B-versus-2B]_[OP-diff]_[BT-diff]_[MODE-planar]_plot.png
 012345_MEG_Wrkmem_eravg_[LM-TIM-2B]_[BT-diff]_[MODE-mag]_plot.png
 012345_MEG_Wrkmem_eravg_[LM-TIM-2B]_[BT-diff]_[MODE-planar]_plot.png
 etc. for all Wrkmem datagroups and contrasts

The results of the Eravg pipeline (with the appropriate event-related condition and contrast detail differences) for the Language Processing (Story-Math) and Motor tasks similarly unpack to the following directories:

MEG/StoryM/eravg/
 figures/

MEG/Motort/eravg/
 figures/

Tfavg results

The results of the [Tfavg](#) pipeline (only for Task scans) for the Working memory task for exemplar subject 012345 unpack to the following directory structure:

MEG/Wrkmem/tfavg/

012345_MEG_Wrkmem_tfavg_[LM-TIM-0B]_[MODE-planar].mat
 012345_MEG_Wrkmem_tfavg_[LM-TIM-2B]_[MODE-planar].mat
 012345_MEG_Wrkmem_tfavg_[LM-TRESP-0B]_[MODE-planar].mat
 012345_MEG_Wrkmem_tfavg_[LM-TRESP-2B]_[MODE-planar].mat
 012345_MEG_Wrkmem_tfavg_[LM-TIM-0B]_[MODE-mag].mat
 012345_MEG_Wrkmem_tfavg_[LM-TIM-0B-versus-2B]_[OP-diff]_[MODE-mag].mat
 012345_MEG_Wrkmem_tfavg_[LM-TIM-0B-versus-2B]_[OP-diff]_[MODE-planar].mat
 etc. for all Wrkmem datagroups and contrasts

figures/

012345_MEG_Wrkmem_tfavg_[LM-TIM-0B]_[MODE-planar]_plot.png
 012345_MEG_Wrkmem_tfavg_[LM-TIM-2B]_[MODE-planar]_plot.png
 012345_MEG_Wrkmem_tfavg_[LM-TRESP-0B]_[MODE-planar]_plot.png
 012345_MEG_Wrkmem_tfavg_[LM-TRESP-2B]_[MODE-planar]_plot.png
 012345_MEG_Wrkmem_tfavg_[LM-TIM-0B]_[MODE-mag]_plot.png
 012345_MEG_Wrkmem_tfavg_[LM-TIM-0B-versus-2B]_[OP-diff]_[MODE-mag]_plot.png
 012345_MEG_Wrkmem_tfavg_[LM-TIM-0B-versus-2B]_[OP-diff]_[MODE-planar]_plot.png
 etc. for all Wrkmem datagroups and contrasts

The results of the Tfavg pipeline (with the appropriate time-frequency response conditions and contrast detail differences) for the Language Processing (Story-Math) and Motor tasks similarly unpack to the following directories:

MEG/StoryM/tfavg/
figures/

MEG/Motort/tfavg/
figures/

Icamne results

The results of the [icamne](#) pipeline (only for Resting state scans) for exemplar subject 012345 unpack to the following directory structure:

MEG/Restin/icamne/figures

012345_MEG_3-Restin_icamne_1.png
012345_MEG_3-Restin_icamne_2.png
etc.
012345_MEG_4-Restin_icamne_1.png
012345_MEG_4-Restin_icamne_2.png
etc.
012345_MEG_5-Restin_icamne_1.png
012345_MEG_5-Restin_icamne_2.png
etc.

Icablpenv results

The results of the [icablpenv](#) pipeline (only for Resting state scans) for exemplar subject 012345 unpack to the following directory structure:

MEG/Restin/icablpenv/

012345_MEG_3-Restin_icablpenv_alpha.power.dtseries.nii
012345_MEG_3-Restin_icablpenv_betahigh.power.dtseries.nii
012345_MEG_3-Restin_icablpenv_betalow.power.dtseries.nii
012345_MEG_3-Restin_icablpenv_delta.power.dtseries.nii
012345_MEG_3-Restin_icablpenv_gammahigh.power.dtseries.nii
012345_MEG_3-Restin_icablpenv_gammalow.power.dtseries.nii
012345_MEG_3-Restin_icablpenv_gammamid.power.dtseries.nii
012345_MEG_3-Restin_icablpenv_theta.power.dtseries.nii
012345_MEG_3-Restin_icablpenv_whole.power.dtseries.nii
012345_MEG_4-Restin_icablpenv_alpha.power.dtseries.nii
etc. for all Restin scans and frequency bands

Icablpenv parcellated results

The parcellated results of the [icablpenv](#) pipeline (only for Resting state scans) (using the [Yeo et al. 2011](#) 17 network parcellation) for exemplar subject 012345 unpack from the [SubjectID]_Restin_parcel_yeo package to the following directory structure:

```
MEG/Restin/icablpenv/
  012345_MEG_3-Restin_icablpenv_alpha.power.Yeo2011.ptseries.nii
  012345_MEG_3-Restin_icablpenv_betahigh.power.Yeo2011.ptseries.nii
  012345_MEG_3-Restin_icablpenv_betalow.power.Yeo2011.ptseries.nii
  012345_MEG_3-Restin_icablpenv_delta.power.Yeo2011.ptseries.nii
  etc. for all Restin scans and frequency bands
```

Icablpcorr results

The results of the [icablpcorr](#) pipeline (only for Resting state scans) for exemplar subject 012345 unpack to the following directory structure:

```
MEG/Restin/icablpcorr/
  012345_MEG_Restin_icablpcorr_alpha.blpcorr.dconn.nii
  012345_MEG_Restin_icablpcorr_betahigh.blpcorr.dconn.nii
  012345_MEG_Restin_icablpcorr_betalow.blpcorr.dconn.nii
  012345_MEG_Restin_icablpcorr_delta.blpcorr.dconn.nii
  012345_MEG_Restin_icablpcorr_gammahigh.blpcorr.dconn.nii
  012345_MEG_Restin_icablpcorr_gammalow.blpcorr.dconn.nii
  012345_MEG_Restin_icablpcorr_gammamid.blpcorr.dconn.nii
  012345_MEG_Restin_icablpcorr_theta.blpcorr.dconn.nii
  012345_MEG_Restin_icablpcorr_whole.blpcorr.dconn.nii

  figures/
    012345_MEG_Restin_icablpcorr_alpha.blpcorr.png
    012345_MEG_Restin_icablpcorr_alpha.blpcorr_parcs.png
    012345_MEG_Restin_icablpcorr_alpha.blpcorr_L-CS.png
    012345_MEG_Restin_icablpcorr_alpha.blpcorr_L-PCC.png
    012345_MEG_Restin_icablpcorr_alpha.blpcorr_L-S2.png
    012345_MEG_Restin_icablpcorr_alpha.blpcorr_R-CS.png
    012345_MEG_Restin_icablpcorr_alpha.blpcorr_R-vCS.png
    012345_MEG_Restin_icablpcorr_betahigh.blpcorr.png
  etc. for all frequency bands and views
```

Icablpcorr parcellated results

The parcellated results of the [icablpcorr](#) pipeline (only for Resting state scans) (using the [Yeo et al. 2011](#) 17 network parcellation) for exemplar subject 012345 unpack from the [SubjectID]_Restin_parcel_yeo package to the following directory structure:

```
MEG/restin/icablpcorr
  012345_MEG_Restin_icablpcorr_alpha.blpcorr.Yeo2011.pconn.nii
```

012345_MEG_Restin_icablpcorr_betahigh.blpcorr.Yeo2011.pconn.nii
etc. for all frequency bands

Icaimagcoh results

The results of the [icaimagcoh](#) pipeline (only for Resting state scans) for exemplar subject 012345 unpack to the following directory structure:

MEG/Restin/icaimagcoh/
012345_MEG_3-Restin_icaimagcoh_alpha.dconn.nii
012345_MEG_3-Restin_icaimagcoh_betahigh.dconn.nii
012345_MEG_3-Restin_icaimagcoh_betalow.dconn.nii
012345_MEG_3-Restin_icaimagcoh_delta.dconn.nii
012345_MEG_3-Restin_icaimagcoh_gammahigh.dconn.nii
012345_MEG_3-Restin_icaimagcoh_gammalow.dconn.nii
012345_MEG_3-Restin_icaimagcoh_gammamid.dconn.nii
012345_MEG_3-Restin_icaimagcoh_theta.dconn.nii
012345_MEG_4-Restin_icaimagcoh_alpha.dconn.nii
etc. for all Restin scans and frequency bands

Icaimagcoh parcellated results

The parcellated results of the [icaimagcoh](#) pipeline (only for Resting state scans) (using the [Yeo et al. 2011](#) 17 network parcellation) for exemplar subject 012345 unpack from the [SubjectID]_Restin_parcel_yeo package to the following directory structure:

MEG/restin/icaimagcoh
012345_MEG_3-Restin_icaimagcoh_alpha.blpcorr.Yeo2011.pconn.nii
012345_MEG_3-Restin_icaimagcoh_betahigh.blpcorr.Yeo2011.pconn.nii
etc. for all Restin scans and frequency bands

Bfblpenv results

The results of the [bfblpenv](#) pipeline (only for Resting state scans) for exemplar subject 012345 unpack from the [SubjectID]_Restin_dtseries package to the following directory structure:

MEG/Restin/bfblpenv/
012345_MEG_3-Restin_bfblpenv_alpha.power.dtseries.nii
012345_MEG_3-Restin_bfblpenv_betahigh.power.dtseries.nii
etc. for all Restin scans and frequency bands

Bfblpenv parcellated results

The parcellated results of the [bfblpenv](#) pipeline (only for Resting state scans) (using the [Yeo et al. 2011](#) 17 network parcellation) for exemplar subject 012345 unpack from the [SubjectID]_Restin_parcel_yeo package to the following directory structure:

MEG/Restin/bfblpennv/

012345_MEG_3-Restoin_bfblpennv_alpha.power.Yeo2011.ptseries.nii
012345_MEG_3-Restoin_bfblpennv_betahigh.power.Yeo2011.ptseries.nii
etc. for all Restin scans and frequency bands

Bfblpcorr results

The results of the [bfblpcorr](#) pipeline (only for Resting state scans) for exemplar subject 012345 unpack from the [SubjectID]_Restin_dconn package to the following directory structure:

MEG/Restin/bfblpcorr/

012345_MEG_Restoin_bfblpcorr_alpha.blpcorr.dconn.nii
012345_MEG_Restoin_bfblpcorr_betahigh.blpcorr.dconn.nii
etc. for all other frequency bands

figures/

012345_MEG_Restoin_bfblpcorr_alpha.blpcorr.png
012345_MEG_Restoin_bfblpcorr_alpha.blpcorr_parcc.png
012345_MEG_Restoin_bfblpcorr_alpha.blpcorr_view_L-CS.png
012345_MEG_Restoin_bfblpcorr_alpha.blpcorr_view_L-PCC.png
012345_MEG_Restoin_bfblpcorr_alpha.blpcorr_view_L-S2.png
012345_MEG_Restoin_bfblpcorr_alpha.blpcorr_view_R-CS.png
012345_MEG_Restoin_bfblpcorr_alpha.blpcorr_view_R-vCS.png
012345_MEG_Restoin_bfblpcorr_betahigh.blpcorr.png
012345_MEG_Restoin_bfblpcorr_betahigh.blpcorr_parcc.png
etc. for all other frequency bands/views

Bfblpcorr parcellated results

The parcellated results of the [bfblpcorr](#) pipeline (only for Resting state scans) (using the [Yeo et al. 2011](#) 17 network parcellation) for exemplar subject 012345 unpack from the [SubjectID]_Restin_parcel_yeo package to the following directory structure:

MEG/restin/bfblpcorr

012345_MEG_Restoin_bfblpcorr_alpha.blpcorr.Yeo2011.pconn.nii
012345_MEG_Restoin_bfblpcorr_betahigh.blpcorr.Yeo2011.pconn.nii
etc. for all other frequency bands

Srcavglcmv results

The results of the [srcavglcmv](#) pipeline (only for Working memory and Motor Task scans) for exemplar subject 012345 unpack to the following directory structure:

MEG/Wrkmem/srcavglcmv/

012345_MEG_Wrkmem_srcavglcmv_[LM-TIM-0B]_[IT-avg].power.dtseries.nii
012345_MEG_Wrkmem_srcavglcmv_[LM-TIM-2B]_[IT-avg].power.dtseries.nii
012345_MEG_Wrkmem_srcavglcmv_[LM-TIM-FIX]_[IT-all].power.dscalar.nii
012345_MEG_Wrkmem_srcavglcmv_[LM-TIM-FIX]_[IT-avg].power.dscalar.nii

etc. for all Wrkmem datagroups and contrasts

figures/

012345_MEG_Wrkmem_srcavglcmv_[LM-TIM-0B]_[IT-avg]_plot.png

012345_MEG_Wrkmem_srcavglcmv_[LM-TIM-2B]_[IT-avg]_plot.png

012345_MEG_Wrkmem_srcavglcmv_[LM-TIM-FIX]_[IT-all]_plot.png

etc. for all Wrkmem datagroups and contrasts

MEG/Motort/srcavglcmv/

012345_MEG_Motort_srcavglcmv_[LM-TEMG-FIX]_[IT-all].power.dscalar.nii

012345_MEG_Motort_srcavglcmv_[LM-TEMG-FIX]_[IT-avg].power.dscalar.nii

012345_MEG_Motort_srcavglcmv_[LM-TEMG-LF]_[IT-avg].power.dtseries.nii

etc. for all Motort datagroups and contrasts

figures/

012345_MEG_Motort_srcavglcmv_[LM-TEMG-FIX]_[IT-all]_plot.png

012345_MEG_Motort_srcavglcmv_[LM-TEMG-FIX]_[IT-avg]_plot.png

012345_MEG_Motort_srcavglcmv_[LM-TEMG-LF]_[IT-avg]_plot.png

etc. for all Motort datagroups and contrasts

The results are saved in CIFTI-2 format. This format was chosen so that results can be directly read by the [Connectome Workbench software](#), in which results from other modalities (structural-MR , fMRI, dMRI) can be also read and overlaid.

Srcavglcmv Parcellated Results

The parcellated results of the [srcavglcmv](#) (only for Working Memory and Motor Task scans) pipeline (using the [Yeo et al. 2011](#) 17 network parcellation) for exemplar subject 012345 unpack from the [SubjectID]_[Task]_parcel_yeo package to the following directory structure:

MEG/Wrkmem/srcavglcmv/

012345_MEG_Wrkmem_srcavglcmv_[LM-TIM-0B]_[IT-avg].power.Yeo2011.ptseries.nii

012345_MEG_Wrkmem_srcavglcmv_[LM-TIM-2B]_[IT-avg].power.Yeo2011.ptseries.nii

etc. for all Wrkmem datagroups and contrasts

MEG/Motort/srcavglcmv/

012345_MEG_Motort_srcavglcmv_[LM-TEMG-FIX]_[IT-all].power.Yeo2011.pscalar.nii

012345_MEG_Motort_srcavglcmv_[LM-TEMG-FIX]_[IT-avg].power.Yeo2011.pscalar.nii

etc. for all Motort datagroups and contrasts

Srcavgdics results

The source reconstructed power results of the [srcavgdics](#) pipeline (only for Working memory and Motor Task scans) for the working memory task for exemplar subject 012345 unpack to the following directory structure:

MEG/Wrkmem/srcavgdics/

012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-alpha].power.dtseries.nii
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-betahigh].power.dtseries.nii
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-betalow].power.dtseries.nii
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-delta].power.dtseries.nii
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-gammahigh].power.dtseries.nii
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-gammalow].power.dtseries.nii
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-gammamid].power.dtseries.nii
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-theta].power.dtseries.nii
 etc. for all other WM task contrasts, data groups and frequency bands

figures/

012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-alpha]_plot.png
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-betahigh]_plot.png
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-betalow]_plot.png
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-delta]_plot.png
 etc. for all .dtseries.nii and .dscalar.nii files in MEG/Wrkmem/srcavgdics/

For the motor task, the srcavgdics pipeline includes both source reconstructed power and coherence with the EMG of the corresponding hand or foot. The results unpack to the following directory structure:

MEG/Motort/srcavgdics/

012345_MEG_Motort_srcavgdics_[LM-TEMG-LF]_[FB-alpha].power.dtseries.nii
 012345_MEG_Motort_srcavgdics_[LM-TEMG-LF]_[FB-betahigh].power.dtseries.nii
 012345_MEG_Motort_srcavgdics_[LM-TEMG-LF]_[FB-betalow].power.dtseries.nii
 012345_MEG_Motort_srcavgdics_[LM-TEMG-LF]_[FB-delta].power.dtseries.nii
 012345_MEG_Motort_srcavgdics_[LM-TEMG-LF]_[FB-gammahigh].power.dtseries.nii
 012345_MEG_Motort_srcavgdics_[LM-TEMG-LF]_[FB-gammalow].power.dtseries.nii
 012345_MEG_Motort_srcavgdics_[LM-TEMG-LF]_[FB-gammamid].power.dtseries.nii
 012345_MEG_Motort_srcavgdics_[LM-TEMG-LF]_[FB-theta].power.dtseries.nii
 etc. for all other Motor task events, data groups and frequency bands

figures/

012345_MEG_Motort_srcavgdics_[LM-TEMG-FIX]_[FB-alpha]_plot.png
 012345_MEG_Motort_srcavgdics_[LM-TEMG-FIX]_[FB-betahigh]_plot.png
 012345_MEG_Motort_srcavgdics_[LM-TEMG-FIX]_[FB-betalow]_plot.png
 012345_MEG_Motort_srcavgdics_[LM-TEMG-FIX]_[FB-delta]_plot.png
 etc. for all .dtseries.nii and .dscalar.nii files in MEG/Motort/srcavgdics/

Srcavgdics Parcellated Results

The parcellated results of the [srcavgdics](#) (only for Working Memory and Motor Task scans) pipeline (using the [Yeo et al. 2011](#)¹⁷ network parcellation) for exemplar subject 012345 unpack from the [SubjectID]_[Task]_parcel_yeo package to the following directory structure:

MEG/Wrkmem/srcavgdics/

012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-alpha].power.Yeo2011.ptseries.nii
 012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-betahigh].power.Yeo2011.ptseries.nii

012345_MEG_Wrkmem_srcavgdics_[LM-TIM-0B]_[FB-betalow].power.Yeo2011.ptseries.nii
etc. for all Wrkmem datagroups, contrasts, and frequency bands

MEG/Motort/srcavgdics/

012345_MEG_Motort_srcavgdics_[LM-TEMG-FIX]_[FB-alpha].power.Yeo2011.pscalar.nii
012345_MEG_Motort_srcavgdics_[LM-TEMG-FIX]_[FB-betahigh].power.Yeo2011.pscalar.nii
012345_MEG_Motort_srcavgdics_[LM-TEMG-FIX]_[FB-betalow].power.Yeo2011.pscalar.nii
etc. for all Motort datagroups, contrasts, and frequency bands

Chapter 4: HCP Processing Pipelines

MR Preprocessing pipelines

The HCP MR data preprocessing pipelines (detailed in [Glasser et al. 2013](#)) are designed to capitalize on the high quality data offered by the HCP. The minimal preprocessing pipelines for structural, functional, and diffusion MRI accomplish many low level tasks, including spatial artifact/distortion removal, surface generation, cross-modal registration, and alignment to standard space. The final standard space makes use of a recently introduced CIFTI file format and the associated grayordinates spatial coordinate system. This allows for combined cortical surface and subcortical volume analyses while reducing the storage and processing requirements for high spatial and temporal resolution data.

Version 3 preprocessing pipelines. All preprocessing of the 1200 Subjects MR data was done using Version 3 (3.1-3.21, depending on the subpipeline) of the preprocessing pipelines. Most of the pipelines are as detailed in [Glasser et al. 2013](#) with updates as noted below (changes from Version 2 pipelines are noted with an asterisk *) and on the [Github HCP pipeline repository](#) where pipeline scripts are available for download. All currently available released data have been processed using the same pipelines to maximize compatibility for analyses going forward.

As a reminder, Q1 data was released in March 2013 using Version 1 of the preprocessing pipelines. In the Q2 release in June 2013, we re-released all Q1 subjects data reprocessed using Version 2 preprocessing pipelines. In the Q3 release in September 2013, we released new subjects processed with the Version 2 preprocessing pipelines. For the 500 Subjects release, we reprocessed all released subjects' data with the Version 3 preprocessing pipelines. For the 900 Subjects release, we added the [MSM-All pipeline](#) that produces data with improved intersubject registration of functional cortical areas.

With the 1200 Subjects release, as with the 900 Subjects release, we are adding new subjects processed with the Version 3 preprocessing pipelines. Therefore, all subjects in the 1200 Subjects release have compatibly processed data that can be directly compared.

For the S1200 Subjects Release, we have added new subpipelines (also v3) that update from EDDY 5.0.9 to 5.0.10 in the diffusion pipeline (v3.19.0), added the bedpostX pipeline, updates to the bias field correction in the functional preprocessing pipeline (PreFreeSurfer pipeline applied only to 7T data, v3.15.1), and updates to the ICA-FIX (v3.16.1), DeDriftAndResample (v3.17.0), Resting State Stats pipelines for 7T data (v3.21.0). See the code and notes on [Github](#) for more details.

We recommend for users to keep track of the processing versions, taking care not to mix data processed with different versions in the same analyses, and update any data acquired from HCP before the 500 Subjects release, as necessary.

Each dataset includes Release Notes (in a Subject_ID/release-notes directory) that state the version number(s) of the relevant HCP pipelines used to process the data. For example, the archive 100307_3T_fmMRI_EMOTION_preproc.zip contains the file release-notes-fmMRI_EMOTION_preprocess.txt:

100307_3T_fmMRI_EMOTION_preproc.zip

Wed Apr 2 11:17:20 CDT 2014
Structural Pipeline v3.1
Functional Pipeline v3.1
Execution 1

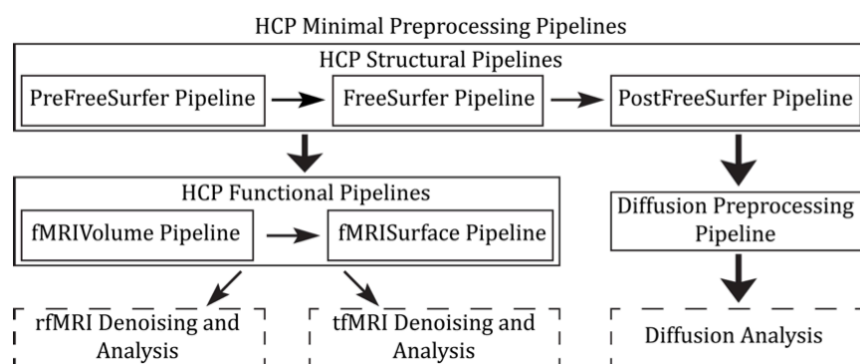
Improvements in processing pipelines that occur over time may warrant retrospective application of updated pipelines to previously released subject data, as we have done with Version 3. The HCP documents any such changes in processing pipelines between versions.

Overview of MR Minimal Processing Pipelines

Structural MR, fMRI, and dMRI datasets were processed by a set of 6 pipelines (see below) aimed at providing high quality volume and surface data. These pipelines use freely available software from the FSL ([Jenkinson M et al., 2012](http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/); <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>) and FreeSurfer ([Dale AM et al., 1999](http://surfer.nmr.mgh.harvard.edu/); <http://surfer.nmr.mgh.harvard.edu/>) and Connectome Workbench ([Marcus et al. 2013](http://www.connectome-workbench.org/)) image analysis suites, and are discussed in detail in ([Glasser et al. 2013](http://www.humanconnectomeproject.org/pipelines/)).

*Pipelines Version 3 were upgraded to use FSL 5.0.6, FreeSurfer 5.3.0-HCP, and Connectome Workbench v1.1.1. These software changes have each made measurable improvements the outputs of the HCP preprocessing pipelines.

The figure to the right shows the overall workflow for preprocessing and data analysis in the HCP. The six minimal preprocessing



Overview of HCP Minimal Processing Pipelines [Glasser et al. 2013](http://www.humanconnectomeproject.org/pipelines/)

pipelines include three structural pipelines (PreFreeSurfer, FreeSurfer, and PostFreeSurfer), two functional pipelines (fMRIVolume and fMRISurface), and a Diffusion preprocessing pipeline.

A summary of the major processing steps for each pipeline is provided below, full descriptions are in [Glasser et al. 2013](#), and **version 3 updates are noted with an asterisk***. The following section, [Directory structure for processed data](#) summarizes the directory structure created by the preprocessing pipelines. [Appendix 3B](#) contains a complete list of all resulting preprocessing file names and subdirectories.

MR Structural Pipelines

PreFreeSurfer pipeline

The main goals of the first structural pipeline, PreFreeSurfer, are to produce an undistorted “native” structural volume space for each subject, align the T1w and T2w images, perform a B1 (bias field) correction, and register the subject’s native structural volume space to MNI space. Thus, there are two volume spaces in HCP data: 1) The subject’s undistorted native volume space (rigidly “aligned” to the axes of MNI space), which is where volumes and areas of structures should be measured and where tractography should be performed, as this space is the best approximation of the subject’s physical brain. 2) Standard MNI space is useful for comparisons across subjects and studies, particularly of subcortical data, which is more accurately aligned by nonlinear volume registration than is cortical data.

Steps:

1. Gradient distortion correction.
2. Coregistration and averaging of T1w and T2w runs.
3. ACPC (i.e. 6 dof) registration for T1w and T2w (FLIRT plus a customized script). The acpc registered distortion corrected T1w image is in the “native” volume space.
4. Initial brain extraction for T1w and T2w (FNIRT-atlas-based-brain-mask).
5. Field map distortion correction and registration of T2w to T1w using a customized FLIRT BBR algorithm.
6. Bias field correction using $\sqrt{T1w \times T2w}$ ([Rilling et al., 2011](#)).
7. Atlas registration (FLIRT linear + FNIRT nonlinear to MNI152).

Note: Difference between the PreFreeSurfer pipeline applied on 3T vs. 7T data:

For 3T data, the bias field correction that is applied to the fMRI scans as part of the minimal processing was estimated from the T1 and T2 anatomical scans as described in 6. above (known currently as the “LEGACY” method in the GenericfMRIVolumeProcessingPipeline.sh

script). For 7T data, we used a better method for the bias field correction of the fMRI scans, in which the coil bias field was estimated from the “SpinEchoFieldMap” scans acquired within the same scan session as the fMRI (and applied using `--biascorrection=SEBASED` in the `GenericfMRIVolumeProcessingPipeline.sh` script). This functionality was introduced starting with v3.15 of the HCP Pipelines.

FreeSurfer pipeline

The FreeSurfer pipeline is based on FreeSurfer 5.3.0-HCP, with a number of enhancements specifically designed to capitalize on HCP data. The main goals of this pipeline are to segment the volume into predefined structures (including the subcortical parcels used in CIFTI), reconstruct white and pial cortical surfaces, and perform FreeSurfer’s standard folding-based surface registration to their surface atlas (`fsaverage`).

Steps:

1. Downsampling of 0.7mm T1w to 1mm using splines.
2. Initial FreeSurfer steps (`autorecon1` except `-skullstrip`).
3. Initialize FreeSurfer with skull registration using PreFreeSurfer brain mask for a more robust registration.
4. FreeSurfer skullstripping.
5. Further early FreeSurfer processing steps (`-autorecon2 -nosmooth2 -noinflate2`).
6. Adjustment of white matter surface placement using hires T1w (full 0.7mm data) with a customized `mri_normalize`, `mrisc_make_surfaces`, and fine tuning of T2w to T1w registration using `bbregister`.
7. Middle FreeSurfer steps (`-smooth2 -inflate2 -sphere -surfreg -jacobian_white -avgcurv -cortparc`).
8. Creation of pial surface using hires T1w (full 0.7mm data) and adjustment of pial surface placement using hires T2w (full 0.7mm data) with a customized `mrisc_make_surfaces` to remove vessels and dura. Then grey matter intensity normalization of T1w, regeneration of pial surface and adjustment with T2w.
9. Final FreeSurfer steps (`-surfvolume -parcstats -cortparc2 -parcstats2 -cortribbon -segstats -aparc2aseg -wmparc -balabels -label-exvivo-ec`).

PostFreeSurfer pipeline

The final structural pipeline, PostFreeSurfer, produces all of the GIFTI surface and auxiliary files necessary for viewing the data in Connectome Workbench, performs individual surface registration using cortical folding surface features and the MSM algorithm (`MSMSulc`),

downsampling registered surfaces for connectivity analysis, creating the final brain mask, and creating myelin maps.

*The switch to MSM_{Sulc} for registration was the most significant change in the version 3.1 structural pipeline compared with previous versions. Instead of using the FreeSurfer spherical surface rotation and warping for intersubject registration, we extract only the surface rotation component from FreeSurfer and use that as the initial input into the novel multimodal surface matching (MSM) algorithm ([Robinson et al. 2014](#)). For the 500 Subjects release, individual cortical folding surface features (FreeSurfer's 'sulc' measure of average convexity) are used with the FreeSurfer spherical surface in MSM_{Sulc} to perform individual surface registration to a group-average surface template created from 196 HCP Q1-Q3 subjects (R196 atlas). MSM_{Sulc} registration is “gentler” in that it introduces less distortion, performing substantially better in distortion and slightly better in cross-subject alignment than FreeSurfer cortical folding registration alone. In addition, MSM_{Sulc} served as a precursor to more substantial improvements available from MSM registration using the information from cortical folding, myelin maps, and rfMRI correlations together (MSM_{All}), described below.

MSM_{All} pipeline

*With the 900 Subjects Release, we added MSM_{All} (MSM-All) registration, which uses information on areal features derived from myelin, Resting State Network (RSN), and rfMRI visuotopic maps in a joint multi-modal registration for improved intersubject registration of functional cortical areas. This pipeline is run after the PostFreeSurfer pipeline to produce similar output files using this additional registration (denoted with “MSM_{All}” in the file names to distinguish from FS [FreeSurfer] and MSM_{Sulc} outputs).

The motivation for adding this registration to the structural preprocessing pipeline comes from the observation that while registration based on cortical folding patterns performs well in early areas such as V1 in the calcarine sulcus and areas 4, 3a, and 3b in the central sulcus, folding patterns are substantially less correlated with association areas implicated in ‘higher’ levels of processing (e.g. areas 44 and 45 in the inferior frontal gyrus) ([Fischl et al., 2008](#); [Van Essen et al., 2005](#); [Van Essen et al., 2012](#)). As demonstrated in [Robinson et al. 2014](#), registration using myelin and RSN maps can improve cross-subject alignment of independent tfMRI datasets, while at the same time decreasing the cross-subject alignment of cortical folding patterns that do not correlate with cortical areal locations. Details on the MSM_{All} method employed by the MSM-All HCP pipeline will be available in an upcoming publication (Glasser et al. 2016, submitted). In the meantime, the MSM-All pipeline is included in the [HCP pipelines released on GitHub](#) and more documentation on the steps of this processing will be forthcoming there.

There are three surface spaces in HCP data: the native surface mesh for each individual (~136k vertices, most accurate for volume to surface mapping), the high resolution Conte69 registered standard mesh (~164k vertices, appropriate for cross-subject analysis of high resolution data

like myelin maps) and the low resolution Conte69 registered standard mesh (~32k vertices, appropriate for cross-subject analysis of low resolution data like fMRI or diffusion). These spaces are also shown in Figure 8 [Glasser et al. 2013](#). The 91,282 standard grayordinates (CIFTI) space is made up of a standard subcortical segmentation in 2mm MNI space (from the Conte69 subjects) and the 32k Conte69 mesh of both hemispheres (Figure 1 [Glasser et al. 2013](#)). *All Version 3 pipelines, including structural pipeline v3.1, output [CIFTI-2.0 format](#) files.

Steps:

1. Creation of Connectome Workbench data files and spec files, creation of FreeSurfer segmentation brain mask.
2. Surface registration with MSM_{Sulc} and downsampling, volume transformation of surfaces nonlinearly into MNI space. Creation of 2mm grayordinates space.
3. Creation of FreeSurfer ribbon files with full 0.7mm resolution data (masks of grey matter and white matter).
4. Myelin mapping and combination of all transforms for one step resampling of T1w and T2w images from original to MNI space ([Glasser and Van Essen, 2011](#)). Normalization of myelin maps ([Glasser et al. 2013](#)).

Following completion of the structural pipelines, the functional or diffusion pipelines may run.

MR Functional Pipelines

fMRIVolume pipeline

The first functional pipeline, fMRIVolume, removes spatial distortions, realigns volumes to compensate for subject motion, registers the fMRI data to the structural, reduces the bias field, normalizes the 4D image to a global mean, and masks the data with the final brain mask. Standard volume-based analyses of the fMRI data could proceed from the output of this pipeline. Care is taken in the fMRIVolume pipeline to minimize the smoothing from interpolation, and no overt volume smoothing is done.

Steps:

1. Gradient distortion correction.
2. FLIRT-based motion correction using the “SBRef” volume as the target (Smith SM *et al.*, submitted).
3. TOPUP-based field map preprocessing using Spin echo field map (for each day of each BOLD run)
4. Distortion correction and EPI to T1w registration of the “SBRef” volume using a customized FLIRT BBR algorithm for distortion correction and bbrregister to fine tune the

registration. A Jacobian file representing the intensity modulations of the distortion correction is stored but not applied to the data.

5. One step spline resampling from the original EPI frames to atlas space including all transforms (motion, EPI distortion, EPI to T1w from FLIRT BBR, fine tuning of EPI to T1w with bbregister, nonlinear T1w to MNI).
6. Intensity normalization to mean of 10000 (like in FEAT) and bias field removal. Brain mask based on FreeSurfer segmentation.

fMRISurface pipeline

The main goal of the second functional pipeline, fMRISurface, is to bring the timeseries from the volume into the CIFTI grayordinates standard space. This is accomplished by mapping the voxels within the cortical grey matter ribbon onto the native cortical surface, transforming them according to the surface registration onto the 32k Conte69 mesh, and mapping the set of subcortical grey matter voxels from each subcortical parcel in each individual to a standard set of voxels in each atlas parcel. The result is a standard set of grayordinates in every subject (i.e. the same number in each subject, and with spatial correspondence) with 2mm average surface vertex and subcortical volume voxel spacing. These data are smoothed with surface (novel algorithm, see [Glasser et al. 2013](#)) and parcel constrained smoothing of 2mm FWHM to regularize the mapping process. The output of these pipelines is a CIFTI dense timeseries that can be used for subsequent resting-state or task fMRI analyses.

Steps:

1. Cortical ribbon-based volume to surface mapping (from MNI space 2mm volume to MNI space native mesh surface).
 1. Voxels between white and pial surface are included in mapping. Voxels that are only partially between white and pial surface are weighted according to their partial volume that is between the two surfaces.
 2. Voxels with a temporal coefficient of variance greater than 0.5 standard deviations of their local neighborhood (sigma=5mm) are excluded from volume to surface mapping. The practical effect of this is to remove any small vessels and brain rim voxels remaining in the ribbon.
 3. Transformation of timeseries from native mesh to fs_LR registered 32k mesh (2mm average vertex spacing) in a single step.
2. Surface-based smoothing (2mm FWHM).
3. Subcortical parcel-constrained smoothing and atlas resampling (2mm FWHM).
4. Creation of dense timeseries (on a standard set of brainordinates).

MR Diffusion Pipeline

The diffusion preprocessing pipeline does the following: normalizes the b0 image intensity across runs; removes EPI distortions, eddy-current-induced distortions, and subject motion; corrects for gradient-nonlinearities; registers the diffusion data with the structural; brings it into 1.25mm structural space; and masks the data with the final brain mask.

Diffusion processing pipeline

1. Basic preprocessing: Intensity normalization across runs, preparation for later modules.
2. 'TOPUP' algorithm for EPI distortion correction.
3. 'EDDY' algorithm for eddy current and motion correction.
4. Gradient nonlinearity correction, calculation of gradient bvalue/bvector deviation.
5. Registration of mean b0 to native volume T1w with FLIRT BBR+bbregister and transformation of diffusion data, gradient deviation, and gradient directions to 1.25mm structural space. Brain mask based on FreeSurfer segmentation.

Note, in the initial Q1 data release (March 2013), the preprocessed dMRI data were in the coordinate system of the individual diffusion scans. All diffusion data released post Q1 (including reprocessed Q1 datasets for the 500 Subjects release) are aligned to the native structural space in order to facilitate various cross-modal comparisons.

With the S1200 release, we updated the diffusion pipeline to support updating FSLs [EDDY](#) from v5.0.9 to v5.0.10. The new version of EDDY performs image slice outlier detection to identify diffusion slices where signal has been lost as a consequence of subject movement during the diffusion encoding. These slices are replaced by non-parametric predictions by the Gaussian Process modeling that is at the heart of EDDY.

HCP pipelines v3.19.0 and later on [Github](#) have this change. As of the S1200 release, all HCP diffusion data (even those subjects that were previously released) were reprocessed with the new v3.19.0 diffusion minimal preprocessing pipeline.

bedpostX pipeline for MR Diffusion Analysis

A pipeline for running FSL's BEDPOSTX on a GPU cluster in the context of the HCP data structure was added in December 2016 for processing all HCP diffusion preprocessing outputs. [BEDPOSTX](#) stands for Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques. The X stands for modelling Crossing Fibres. bedpostx runs Markov Chain Monte Carlo sampling to build up probability distributions on diffusion parameters at each voxel. It creates all the files necessary for running probabilistic tractography.

As of the S1200 release, all minimally preprocessed HCP diffusion data was also run through the bedpostX pipeline and results are available as a separate diffusion analysis package.

Overview of MEG Processing pipelines

HCP MEG datasets were processed by a set of custom pipelines aimed at providing high quality data. The figure below shows the overall workflow for MEG data processing in the HCP.

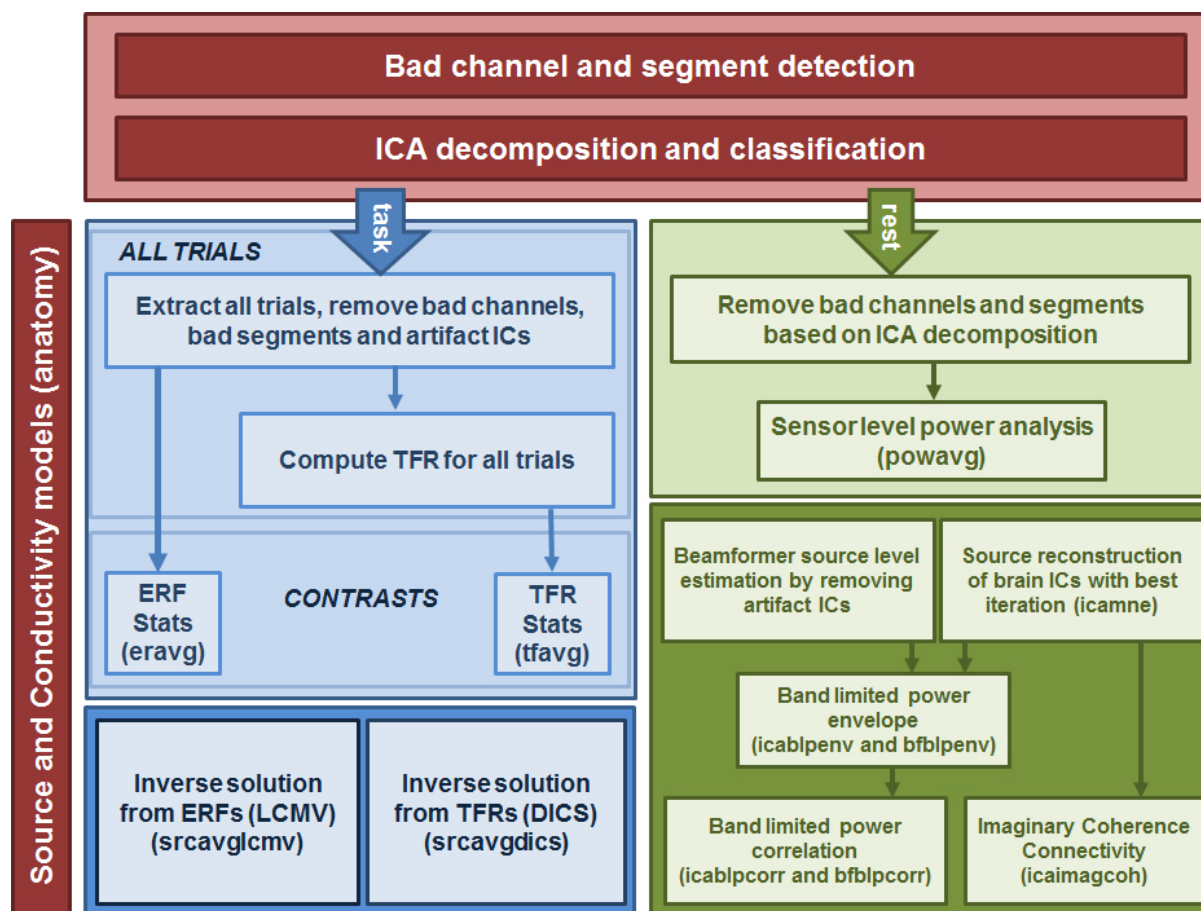
Artifacts, bad channels, and bad segments are identified and removed from rMEG and tMEG data which are processed by an ICA-based pipeline to identify artifacts (environmental and subject's artifacts) and non-artifacts components of the overall signal.

The rMEG preprocessing (channel-level) pipelines produce: i) a set of non-artifact ICs comprising time-courses and sensor maps; ii) cleaned data (raw minus artifacts) at the channel level in 2 second chunks; iii) channel-level power spectrum densities.

The rMEG source-level processing pipelines i) project the sensor maps of the brain ICs produced in channel-level processing into source space, ii) estimate the band-limited power (BLP) envelope of the source-level signals, iii) produce source-level Pearson correlation coefficient-estimated dense and parcellated connectomes on BLP timeseries, and iv) produce frequency-specific connectomes based on the Multivariate Interaction Measure. Two kinds of source reconstruction strategies have been implemented based on two very common inverse algorithms namely the Minimum Norm Estimate (MNE) and Linearly Constrained Minimum Variance (LCMV) Beamformer filters. Most of the connectome level results are obtained using both methods.

The channel-level tMEG preprocessing pipelines provide: i) clean data for the Task experimental paradigms, split in various groups of trials according to different conditions of interest, ii) Event Related Field (ERF) averages at the sensor level for a number of conditions and for comparisons between pairs of conditions for each Task paradigm, and iii) Time-Frequency Representation (TFR) averages at the sensor level for a number of conditions and for comparisons between pairs of conditions. The tMEG source-level processing pipelines produce: i) ERF band-limited power time series averages at the source level for a number of conditions and for comparisons between pairs of conditions for each Task paradigm, and ii) Time-Frequency Representation (TFR) power series averages at the source level for a number of conditions and for comparisons between pairs of conditions.

A summary of the major processing steps for each pipeline is provided below. [Directory structure for Channel- \(preprocessed\) and Source-level processed MEG data](#) summarizes the



Overview of HCP MEG Processing

directory structure created by the processing pipelines. [Appendix 3G](#) contain complete lists of all resulting processed file names and subdirectories.

The pipelines use freely available software from the *FieldTrip* analysis toolbox and custom scripts developed by the HCP that are collectively referred to as “megconnectome”. Both megconnectome software and the FieldTrip toolbox are available for download [on the HCP website](#). The megconnectome software consists of analysis scripts, analysis functions and template data, e.g. the channel layout and a canonical MRI. Each analysis script represents one of the steps of the analysis pipeline as outlined below and in [Larson-Prior et al., 2013](#).

Improvements in MEG processing pipelines that occur over time may warrant retrospective application of updated pipelines to previous data releases. The HCP will document any such

changes in processing pipelines between versions and indicate which versions are used in the release notes provided with each dataset.

MEG Data Processing Output File Formats

The MEG data processing pipelines output files in .txt, .mat, .xml, and CIFTI (*.dtseries.nii, *.dscalar.nii, and *.dconn.nii) file formats. Below are details on each of these formats.

*.txt Files

The files ending with the *.txt extension are ASCII text files with 8-bit Unicode Transformation Format (UTF-8) content. The content of the files is structured to be self-explanatory. An example is given below:

```
badsegment.zscore = [  
419334 426841  
427909 428408  
];  
badsegment.ica = [  
1 113472  
417760 427193  
617686 620126  
702465 704442  
743761 745389  
928217 979475  
];  
badsegment.manual = [  
1 113472  
928217 979475  
];
```

Although the files have the extension *.txt, their content can be directly interpreted by MATLAB as a valid script. Since the file names start with numbers and have special characters in them, such as the '-' (minus) sign, you cannot simply rename the extension from *.txt to *.m, as MATLAB script file names cannot start with a number and cannot have a '-' in their name. The megconnectome/analysis_functions/hcp_read_ascii.m function can be used to read the content into MATLAB workspace, just like the standard MATLAB "load" command would be used for *.mat files.

*.mat Files

The files ending with the *.mat extension are binary MATLAB files. The file format of these files is specified at http://www.mathworks.com/help/pdf_doc/matlab/matfile_format.pdf. Over the

years, MATLAB has introduced a number of file formats. The following overview is derived from the MATLAB documentation:

- v4: Can only contain two-dimensional double, character, and sparse arrays.
- v6: Supports all version 4 features plus N-dimensional arrays, cell and structure arrays, and variable names greater than 19 characters.
- v7: Supports all version 6 features plus data compression and Unicode character encoding.
- v7.3: Supports all version 7.0 features plus support for data items greater than or equal to 2GB on 64-bit systems

From mat-file format version 7.3 onward, the files are based on the Hierarchical Data Format (HDF) standard. Although this makes them in principle more accessible in other platforms, in practice the v7 files can more easily be accessed, as it is a simpler file format for which the specification is available. Furthermore, the v7 format is in general slightly smaller than the v7.3 format, hence we have chosen to store data in the v7 file format.

In MATLAB you can read these files using the "load" command.

In Python you can read these files using

```
>>> import scipy.io  
>>> data = scipy.io.loadmat('file.mat')
```

***.xml Files**

The files ending with the *.xml extension contain provenance information in Extensible Markup Language (XML) format. These files always accompany another file, e.g. the file example.mat is accompanied by the provenance file example.mat.xml. The content of the XML file is encoded with 8-bit Unicode Transformation Format (UTF-8) and can be read in any text editor. The format should be self-explanatory, an example is given below:

```
<?xml version="1.0" encoding="utf-8"?>  
<megconnectome xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
xsi:noNamespaceSchemaLocation="megconnectome.xsd">  
  <version>  
    <matlab>  
      <Name>MATLAB</Name>  
      <Version>8.0</Version>  
      <Release>(R2012b)</Release>  
      <Date>20-Jul-2012</Date>  
    </matlab>  
  </version>  
</megconnectome>
```



```
<Name>megconnectome</Name>
<Version>2.2</Version>
<Release>www.humanconnectome.org</Release>
<Date>10-Nov-2014</Date>
</megconnectome>
<fieldtrip>
  <Name>FieldTrip</Name>
  <Version>r9924</Version>
  <Release>fieldtriptoolbox.org</Release>
  <Date>27-Oct-2013</Date>
</fieldtrip>
</version>
<compiled>true</compiled>
<username>schoffelen</username>
<hostname>node054</hostname>
<architecture>glnxa64</architecture>
<buildtimestamp>10-Nov-2014 10:06:08</buildtimestamp>
<pwd>/home/schoffelen</pwd>
<matlabstack>
In hcp_write_provenance at 39
In hcp_write_matlab at 62
In megconnectome at 129
</matlabstack>
  <filename>100307_MEG_anatomy_sourcemodel_3d6mm.mat </filename>
  <md5sum>d4d168a9ad33ee2fbb3becc9a3acdaf7 </md5sum>
</megconnectome>
```

CIFTI Files (*.dtseries.nii, *.dscalar.nii, and *.dconn.nii)

The files ending in *d[type].nii are CIFTI-2 files that contain timeseries, map, and connectivity data outputs from the HCP source-level processing pipelines. These CIFTI files contain values that are to be mapped onto a Gifi surface geometry for the cortical sheet for visualization. The MEG results-based CIFTI files produced by the HCP MEG pipelines require the 4k vertices per hemisphere Gifi files that are present on the individual subjects' anatomy packages, e.g. 012344.L.midthickness.4k_fs_LR.gii and 012344.R.midthickness.4k_fs_LR.gii. The cortical sheet coordinate system defined for these files is the subject-specific ACPC-aligned headcoordinate space. Currently, Connectome Workbench is one of the few software packages that are fully CIFTI-compatible, therefore performing many operations on this data and visualization assumes the user to have Connectome Workbench v1.0 installed. C++ and MATLAB import and export functions are available from <https://github.com/Washington-University>. FieldTrip is already CIFTI-compatible and other software packages including FSL

and FreeSurfer are working on becoming CIFTI-compatible. More information on the details of the CIFTI-2 format is available at: <http://www.nitrc.org/projects/cifti/>. The CIFTI standard extensions are listed in the specification, see Appendix A in http://www.nitrc.org/forum/attachment.php?attachid=342&group_id=454&forum_id=1955 for the list.

MEG Anatomy processing pipeline

The MEG anatomy pipeline generates data, constructed from anatomical information, that are needed for MEG source reconstruction:

1. coregistration information that defines transformations between the different relevant coordinate systems: (1) voxel-based, (2) MEG-sensor based coordinate system, and (3) an ACPC-based RAS coordinate system, where the origin is defined in the anterior commissure, and the X, Y and Z axes run in right, anterior and superior direction, respectively.
2. generation of a single shell volume conduction model.
3. generation of a set of source models (cortical sheet and 3-D grids) defined on a regular 3D grid in normalized space.

The single shell volume conduction model is defined in the MEG-system based head coordinate system and was obtained by segmentation of the anatomical MRI, using SPM8.

One set of source models are defined on a regular 3D grid in normalized MNI-space. For each participant the anatomical image was non-linearly volumetrically warped to a template image (using SPM8), and the inverse of this warp was applied to the positions on the grid. As a consequence, different source reconstructions that are defined on this grid (e.g. from different sessions or subjects) can be directly averaged, because each position in the grid represents the same location in normalized space. There is also a source model that is defined on the individual subject's cortical sheet, with ~4k vertices per hemisphere. Across subjects, the dipole positions on these sheets are registered in surface-space, allowing for direct averaging across subjects.

How is the MEG coregistered to the anatomical MRI and volume conduction model?

Coregistration is performed using an unmasked 1mm resolution T1-weighted anatomical image (which, due to anonymity requirements, is not part of the release). The geometric information that is needed for source reconstruction (i.e. MEG sensor positions and information derived from the anatomical MRI) is expressed in the head coordinate system, which is defined as being relative to the participant's head and based on a set of landmarks on the head, and is thus

subject-specific. This coordinate system is expressed in meters, with the principal (X, Y, Z) axes going through external anatomical landmarks on the head (i.e. fiducials). The details are:

- the origin is exactly between the left and right pre-auricular points (LPA and RPA)
- the X-axis goes towards the nasion (NAS)
- the Y-axis goes approximately towards LPA, orthogonal to X and in the plane spanned by the fiducials
- the Z-axis goes approximately towards the vertex, orthogonal to X and Y

The magnetometer locations in the raw and processed channel level data are represented in this coordinate system. The "anatomy" pipeline performs the coregistration of the MEG with the MRI data and results in a volume conduction model and various source models (cortical sheet and 3-D grids), which are all expressed in this head coordinate system. The "anatomy" pipeline also produces the 4x4 homogenous transformation matrices to convert between individual head coordinates, MNI coordinates and voxel coordinates. These transformation matrices are represented in the 012345_MEG_anatomy_transform.txt files.

What is the interpretation of the different homogeneous transformation matrices?

The transformation matrices represented in the 012345_MEG_anatomy_transform.txt files represent the transformation between different coordinate systems. Each matrix is named [from]2[to] where it denotes the transformation from coordinate system [from] into coordinate system [to]. The matrices that have 'vox' or 'vox07mm' in their name transform from or to anatomical MRI voxels. The matrices that contain 'vox' are of limited use to external users, because the release does not contain the original anatomical images that were used to create these transformation matrices. The matrices with 'vox07mm' refer to transformations that relate the voxels in the T1w_acpc_dc_restore.nii.gz image (part of the structural preprocessing pipeline output) to the other coordinate systems used.. The transformation matrices that are of most relevance are the bti2spm and spm2bti, which convert between the MEG-system based coordinate system (bti) and a coordinate system based on the ACPC-based RAS (spm).

How are the MEG source models defined?

The 3-dimensional grids that can be used for (beamformer) source reconstruction are based on 3-dimensional grids with a regular spacing of the dipoles, defined in normalized MNI space. The spacing between the dipoles (in normalized space) is represented in the filename (i.e. SUBJECTID_MEG_anatomy_sourcemodel_3d4mm means that the dipoles are 4mm apart). For each participant, the template grid has been warped to individual head space in order to express the dipoles' coordinates in the head coordinate system that is used for source reconstruction (see above). This warp is a non-linear one, and is based on the inverse of the nonlinear volumetric spatial normalization from individual head space to normalized MNI space. Consequently, the dipoles' coordinates are not placed on a regular 3-dimensional grid in individual head space. Yet, the source models allow for easy averaging across subjects, without

the need for an intermediate interpolation step, because individual grid points coincide in normalized space.

The cortical sheet-based source models allow for distributed source reconstruction, e.g. weighted minimum norm estimates. These models have been constructed from the individual 32k-node per hemisphere surface-registered cortical sheets which are output files of the [MR structural preprocessing pipeline](#), and have ~4k nodes per hemisphere.

MEG Preprocessing Pipelines

A set of (utility) preprocessing pipelines are run on all MEG data prior to further resting-state or task-specific analysis. Overall, these pipelines perform various sanity checks and quality assurance tasks.

“datacheck” pipeline

In the “datacheck” pipeline, the header and events of each MEG dataset are checked for basic features (such as sampling rate, number of channels and whether EOG and ECG are present). The dataset check pipeline results in a text file and figures that provide qualitative information on MEG data quality (see exemplar files in **Directory structure for Channel-level Processing Pipelines- [Datacheck](#)**). It does not perform a detailed analysis of artifacts, bad channels or data segments. Measures of data quality are examined (output in six figures per run):

1. SQUID jumps: (values $\leq 1e12$), to identify bad sensors or acute noise increases during a run.
2. Trigger / Response events: (values 0-400), to check for appropriate numbers and values of trigger events and response events for a given run type.
3. Power Line Noise: (values $< 1.5e-27 \text{ T}^2/\text{Hz}$), to check for sensors with unusually great power, (e.g. A246 is usually noisy indicating a bad sensor).
4. MEG Powerspectra – to confirm three features: rough linearity in the range of sampling frequency, $1/f$ coarse overall pattern, and expected 60Hz noise. Deviant channels (unusual spikes or unusual power) are noted.
5. MEGREF Powerspectra - to confirm rough linearity in the range of sampling frequency and $1/f$ coarse overall pattern. Deviant channels (unusual spikes or unusual power) are noted.
6. Neighbor Correlation – to check for smooth transitions between neighbors. Correlations are expected to be greater than 0.6 (except for empty room).

"baddata" pipeline

MEG processing requires that artifact signals are removed from data. The artifacts displaying excessive signal amplitude are generated by non-working channels and bursts of noise and are referred to as “bad channels” and “bad segments”, respectively. The purpose of this pipeline is to detect these bad channels and bad segments.

The pipeline input consists of the raw MEG data and a list of manually defined bad channels and bad segments. Different strategies for detection of bad channels and segments are applied, and their outputs can be combined or used separately for further processing of the data. At the end of the computation, output QC figures are produced. This QC step is mainly used to evaluate the quality of the scan, in terms of number of bad channels and number and duration of bad segments. Additionally, QC figures allow one to identify whether further non-working channels or bad segments not detected by the pipeline are still depleting the data, so that the pipeline should be run again. In this case, if any channel or segment is manually selected, the baddata pipeline is run again using these selected channels/segments as input.

Epochs at the beginning and at the end of each resting state fixation session are automatically selected as bad, since in these epochs the subject's noise (eye blinks and saccades) is considerably larger than during the central part of the run (see below).

For task runs, data between the first and last triggers limit the portion of data to be preprocessed by the baddata pipeline. The output of the baddata pipeline consists of two ASCII files, called `subject_session_baddata_badchannels.txt` and `subject_session_baddata_badsegments.txt` (see below for samples).

- ***What does it mean if a channel is marked as "bad"?***

A channel is marked as “bad” when the signal recorded by that channel is abnormal and cannot be used for further processing. The presence of these channels would have a negative impact on both sensor level analysis and source reconstruction. There might be a set of reasons invalidating the “bad” channel output: i) the SQUID is not working; ii) the SQUID is extremely noisy; iii) the channel read out electronics is not working; iv) the channel read out electronics is not locked (that is, the flux locked loop scheme is temporarily not working and should be reset) during that recording and the output is meaningless.

- ***Why do subsequent scans of the same subject have different bad channels?***

If the SQUID is damaged or noisy, or if the readout electronics is not working, the related channel will always be classified as “bad”. Conversely, if the channel loop was not temporarily locked, the channel may not be flagged as “bad” in the subsequent recording. Indeed, the channel array is reset before every run, and this operation might recover the “former bad” channel. This implies that the set of “bad” channels will include

both permanently and temporary non-working channels, and the temporary ones might be different for each run.

- ***What does it mean if a segment is marked as "bad"?***

During the recordings, short time epochs of high variance might occur. During these epochs, sources lying outside the brain (noise sources) generate a magnetic field which is several times larger than the brain magnetic field. Those fields might be generated by the subject him/herself, as for example eye squints, or might be related to external sources, such as car traffic inducing mechanical vibrations of the shielded room or power supply bursts.

- ***What is the consequence that in some Resting-state sessions the data acquisition was started prior to starting the Resting-state E-Prime script?***

During the Resting-state scans of some of the subjects, the data acquisition was started prior to starting the E-Prime script. As a consequence the initial segment of MEG data may be noisy, which could reflect the fact that the participant was not yet in the resting-state. In these instances it is advisable to discard for analysis any data that has been specified manually as bad segment in the accompanying ASCII files (as in e.g. 012345_MEG_3-Restin_baddata_manual_badsegments.txt). The manual bad segments have been specified in order to realize a 300-second long scan, where any data in excess of this has manually been specified to be a bad segment. 2/3 of this data in excess of 300 was specified at the beginning of the scan, the remaining 1/3 at the end.

The pipeline can be represented by the following logical steps (see figure below):

1. Bad channels are selected through the estimation of the correlation between each channel and its neighbors. This approach is based on the assumption that the signal of nearby sensors is very similar due to the spread in the propagation of the brain magnetic field ([Winter et al., 2007](#)). A bad channel is then identified by searching for outliers in the neighbor correlation distribution by means of a statistical threshold (3σ below the mean neighbor correlation from all the channels). Once these channels are identified they are removed from subsequent analysis by this pipeline.
2. Each MEG channel is scanned for the identification of noisy segments based on the Z-score of each data point relative to the statistical characteristics of the entire data time-series of a channel. First the Z-score of each data point is computed and the data points that have a Z-score higher than a threshold (20) is flagged as bad. Then each channel is scanned for clipping segments, during which the signal is flat. This is performed using FieldTrip's built in function `ft_artifact_clip.m`. Finally each channel is scanned for segments with muscle related artifacts.

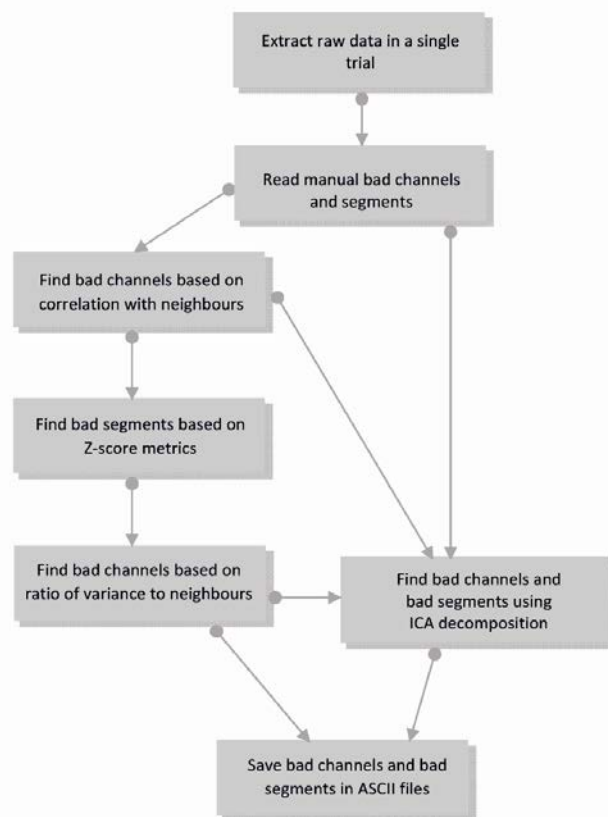
This is performed using FieldTrip's built in function `ft_artifact_muscle.m`. Once these segments are identified they are removed from data used in step 3 and are stored in the output.

3. Bad channel identification is performed by calculating the ratio between the standard deviations of each channel and its neighbors. This approach is based on the assumption that the signal in nearby sensors should have a comparable variance ([Winter et al., 2007](#)). A bad channel is then identified if its standard deviation is more the 50% higher than that of its neighbors. Once these channels are identified they are removed from subsequent analysis by this pipeline.

4. An iterative approach based on the Independent Component Analysis (ICA) is used to decompose the sensor signals into independent components (ICs) in order to find components associated with non-working channels or epochs affected by artifacts such as large spikes and bumps. This approach uses the raw data and the manual list of bad segments (obtained in step 2) and the list of bad channels (obtained in steps 1 and 3) as input.

First, one run of ICA is applied to the data and bad channels are selected based on the sensor map topographies associated with the ICs produced by the approach. Indeed, sensor maps associated with non-working channels show a very large value on the problematic channel and values close to zero for the other channels. Automatic selection is then achieved by a statistical threshold (10σ) applied on the distribution of the sensor map values.

Bad segment selection is performed on each IC in two main steps. First, the IC time course is segmented in epochs with 1 s duration and the variance of the time course is estimated for each epoch. Epochs with high variance are selected through a statistical threshold (12σ). Second, to identify the bad segment with a finer time resolution, the power time course of the IC is estimated in a window of 10 seconds around the epochs with high variance. A statistical threshold is applied to the distribution of the power time course. This ICA approach is iterative, as it is run every time a bad channel or a bad segment is identified, with upgraded input, until no further bad channels and bad segments are found. The outputs of the baddata pipeline can be



Steps of the baddata pipeline.

merged together (as it is done in the task analysis and channel level rest analysis) or can be used separately (as it is done in the ICA based analysis of resting state data, using only the bad segments produced by the ICA based approach).

- ***How are bad channels represented?***

Bad channels are detected for every individual scan and represented in an ASCII file with extension *.txt. The content of the text file can directly be parsed by MATLAB or can be imported using the megconnectome `hcp_read_ascii` function. An example is given here:

```
badchannel.neigh_corr = {
    'A2', 'A90'
};
badchannel.neigh_stdratio = {
    'A195', 'A20', 'A246', 'A41', 'A73'
};
badchannel.all = {
    'A195', 'A2', 'A20', 'A246', 'A41', 'A73', 'A90'
};
```

- ***How are bad segments represented?***

Bad segments are detected for every individual scan and represented in an ASCII file with extension *.txt. The content of the text file can directly be parsed by MATLAB or can be imported using the megconnectome `hcp_read_ascii` function. Segments are indicated by the begin and end sample in the raw data file, which starts with sample 1. An example is given here:

```
badsegment.zscore = [
    291544 297604
    589700 591042
    591748 595895
];

badsegment.ica = [
    1 89094
    384791 395319
];
```

```
398478 405186
561693 570009
573970 576941
699447 743995
];
```

```
badsegment.manual = [
1 89094
699447 743995
];
```

```
badsegment.all = [
1 89094
291544 297604
384791 395319
398478 405186
561693 570009
573970 576941
589700 591042
591748 595895
699447 743995

];
.
```

The directory structure, including the results from the baddata pipeline, is reported in [Appendix 3G](#).

“icaclass” pipeline

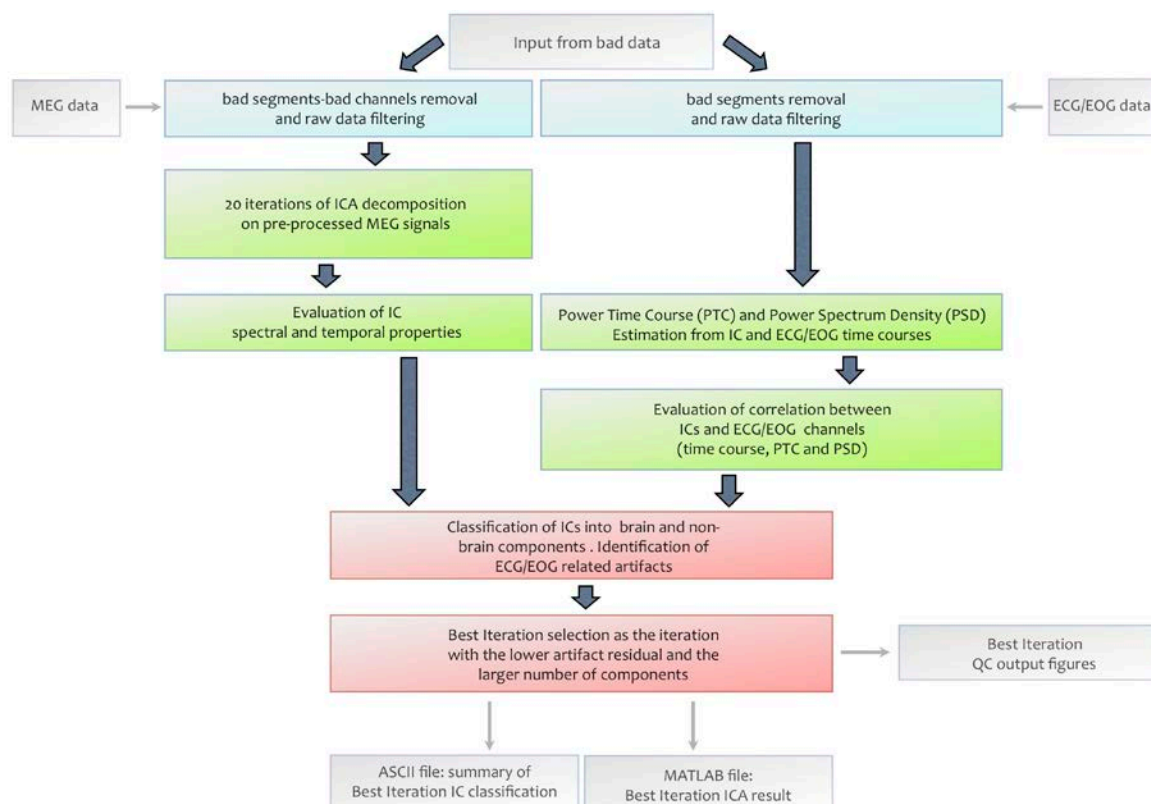
This pipeline performs decomposition of MEG raw data into brain and non-brain components. This is achieved by means of an Independent Component Analysis followed by an automatic classification of the obtained Independent Components (ICs).

Differently from the baddata pipeline, which is not classifying ICs and is only detecting bad channels and bad segments, here the IC decomposition is repeated several times starting from different initial conditions, and the decomposition with the lowest artifact residual and the largest number of brain components is retained as the best decomposition. This best decomposition is selected through an automatic classification of the ICs, which is based on the evaluation of 6

parameters quantifying temporal and spectral properties of the IC signals (3 parameters) and contribution of the eye or heart magnetic signals (3 parameters). The IC time courses and the related sensor maps related to the best decomposition are stored as the output. The pipeline produces also a summary txt file and output figures useful for the quality check step. Each figure shows summary plots of a single IC as well as the values of the classification parameters and the corresponding flag (brain or artifact).

The pipeline can be represented by the following logical steps (see figure below):

1. Data are filtered through a band-pass (1.3-150 Hz) and a notch (59-61/119-121 Kz) Butterworth filters. Bad channels and bad segments are removed from data.
2. ECG and EOG data are preprocessed with the same procedures as in 1.
3. Twenty iterations ICA decomposition (FastICA with deflation approach) of the preprocessed data are run from different initializations.
4. Power Spectrum Density (PSD) and Power Time Course (PTC) are estimated from the IC time courses and the EOG/ECG time courses. Three correlation coefficients between: i) the IC



Steps of the icaclass pipeline.

and the reference time courses, ii) the IC and the reference PTC, and iii) the IC and the reference PSD are estimated.

5. Additional 3 parameters are estimated from the kurtosis of the IC time courses, the 1/f trend and the flatness in the PSD of each IC. These parameters allow classification of environmental and instrumental artifacts.

6. For each iteration of the ICA decomposition, classification into brain and non-brain components is carried out as follows: an IC is classified as a brain component is identified if none of the 6 parameters exceed its corresponding threshold. Moreover, if at least one of the 3 parameters based on the correlation with the EOG/ECG channels exceeds its threshold, the IC is flagged as ECG/EOG related artifact. The ECG/EOG components will be removed from data in the task pipelines. In the resting state pipeline only the brain components are used for further analysis.

7. The best iteration is selected as the one with the largest value of the product between the number of brain ICs and a number quantifying the quality of artifact rejection, defined as 1-artifact residual. The artifact residual is estimated as the maximum correlation between the time courses and PTC of all the IC and the EOG/ECG channels.

8. Results are saved as MATLAB and ASCII files. A summary on the icaclass output, including the number of total components, which ones are brain IC and which are not, and which ones are flagged as ECG/EOG artifacts, is in the file subject_session_icaclass.txt. A sample of the summary included in this file is given below:

```
vs.total_ic_number = 39;
vs.brain_ic_vs_number = 37;
vs.brain_ic_vs = [
  2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
 30 32 33 34 35 36 37 38 39
];
vs.ecg_eog_ic = [
  1 31
];
```

The directory structure including the results from the icaclass pipeline is reported in [Appendix 3G](#).

“icaclass_qc” pipeline

The purpose of this pipeline is to verify the automatic classification of independent components identified in the icaclass pipeline (above) by visual inspection. In the cases in which a component is misclassified, this step allows one to manually modify its classification. Misclassification can occur for a number of reasons. As an example, reference electrical

channels might not be correctly placed on the subject or might not work. EOG/ECG reference channels are specifically used to identify artifacts that will be discarded in further analysis. If the ECG/EOG signals are inappropriate, the ICs capturing the heart and eye-related muscular activity may be classified like brain activity or as non-brain (but not ECG/EOG) components.

Another reason for misclassification might be that the EOG channel is capturing brain activity due to positioning (one of the electrodes is on the forehead) and thus some brain components are flagged as non-brain due to correlation with the EOG channel.

Finally, automatic classification of ICs is based on a set of thresholds, which were estimated from a ROC analysis on more than three hundred components classified by visual inspection. These components were obtained from Phase 1 subjects. Nevertheless, for statistical reasons a component can be misclassified because it slightly exceeds one of the thresholds. As an example, this might occur when the magnetic noise due to car traffic is much larger than in the ROC set, thus the 1/f threshold would not be adequate in this case.

If a component is modified, according to the new classification, the label "Brain Corrected", "Artifact corrected" or "Physiological Artifact Corrected" is shown in the output figure. In the file `subject_session_icaclass_vs.txt`, the field `vs.flag = 1`. The indexes of the corrected ICs are displayed in the fields: `vs.good`, `vs.bad`, and `vs.physio` for re-classified brain, non-brain and ECG/EOG components, respectively. Result files and figures include `_vs` in their name (see in [Appendix 3G](#)). If the automatic classification is able to properly classify all the components, then the above fields are empty.

Manual improvements/corrections to Anatomy and Preprocessing pipeline results

HCP pipelines for MEG data analysis are mostly automatic. Manual interventions are required in some case for different reasons. For instance, in the automatic artefact identification pipeline some independent component may be incorrectly classified, which can be corrected by visual inspection of the initial ICA quality check output figures. In the following, we summarize the manual interventions and the rationale behind them.

Anatomy pipeline manual interventions

The anatomy pipeline starts with an interactive step. The initial coregistration of the anatomical MRI with the MEG data is done using head localizer coils attached to the scalp. The position of these localizer coils relative to some anatomical landmarks (nose and ears) is measured prior to MEG acquisition using a Polhemus electromagnetic digitizer, with which the complete head shape is digitized. The interactive steps in the anatomy pipeline consist of identifying the same anatomical landmarks (nasion and pre-auricular points, as well as the anterior and posterior commissures) in the anatomical MRI as a starting point for the coregistration. Subsequently, the coregistration is refined by matching the digitized head shape to the scalp surface from the

anatomical MRI. This latter refinement is also done interactively (i.e. by manual rotations and translations).

“baddata” pipeline manual interventions

The main output of the "baddata" pipeline is the selection of bad segments and bad channels to be discarded from the data in the following analysis steps. In general, the "baddata" pipeline is able to automatically select these bad segments and channels and no further interventions are required. However, in some cases not all artifactual segments/channels are detected and manual intervention is used to optimize the selection. Baddata manual intervention may be performed after the visual inspection of baddata and icaclass_vs figures. For instance, since the icaclass pipeline performs a many-iteration ICA decomposition, looking at the "icaclass_vs" output figures may show further channels/segments to be excluded. Also looking at the "baddata" output figures, one can find segments or channels not selected by the algorithm.

The "baddata" manual intervention is performed by editing the corresponding ASCII files: `<expID>_<scanID>_baddata_manual_badsegments.txt` and `<expID>_<scanID>_baddata_manual_badchannels.txt` where the variables `manual_badchannels` and `manual_badsegments` are defined. After manual segments or channels are added, the "baddata" and "icaclass" pipelines must be re-run.

“icaclass” pipeline manual interventions

"Icaclass" output consists of a set of independent components classified as "Brain" or "Artifacts". The automatic classification uses a set of 6 parameters, which makes it, in general, a very efficient algorithm. However, it may happen that an independent component is misclassified for some reason. For instance, if the EOG/ECG reference electrodes are not working properly, the automatic classification may fail. The manual intervention allows the user to overrule the incorrect classification of the obtained components.

rMEG Channel-level Processing Pipelines

The rMEG Channel-level analysis pipelines 1) use the outputs of the “baddata” and “icaclass” pipelines (above) to clean and segment the rMEG data, and 2) generate the averaged resting state power spectrum for each sensor (channel).

"rmegpreproc" pipeline

The resting state preprocessing pipeline results in a cleaned representation of the sensor-level MEG data. The raw data are segmented into 2-second pieces. Bad channels and segments are

removed. ICA components that have been classified as artifacts are projected out of the data. The cleaned data are represented as FieldTrip structure in a MATLAB file.

"powavg" pipeline

This pipeline results in the averaged power spectrum of the resting state MEG data at the sensor level. The continuous data are segmented, Hanning-tapered, Fourier-transformed and the estimated power is averaged over all segments. The power spectrum is represented as FieldTrip structure in a MATLAB file.

What are the filter settings used in the analysis of the resting-state MEG?

In the resting state pipeline, data filtering is performed by means of a Butterworth filter both in band-pass and band-stop case. Raw data are first band-pass filtered by means of a two-step procedure performing consequently high-pass and low-pass filtering (cutoff frequency 1.3 Hz and 150 Hz). Notch filters are then applied to the band-pass filtered data in order to remove the power line noise (cutoff frequencies 59-61 Hz and 119-121 Hz).

rMEG Source-level Processing Pipelines

The rMEG Source-level processing pipelines 1) use the outputs of the "icaclass" pipeline (above) to project the sensor maps of the brain ICs from the channel space into the source space, 2) generate frequency specific connectomes based on the "Multivariate Interaction Measure (MIM)", 3) estimate the band limited power envelope of the reconstructed source level signals, and 4) produce source-level Pearson correlation coefficient-estimated dense and parcellated connectomes on band limited power time series. Two kinds of source reconstruction strategies have been implemented based on two very common inverse algorithms namely the Minimum Norm Estimate (MNE) ([icamne pipeline](#)) and Beamformer ([bfbldenv pipeline](#)) filters. Most of the connectome level results are obtained using both methods.

"icamne" pipeline

The "icamne" (ICA minimum norm source reconstruction) pipeline applies weighted minimum norm estimation to project the sensor maps of the brain ICs (Independent Components) obtained by [icaclass](#) from the channel space into the source space. The source space onto which the IC topographies are projected consists of the individual surface-registered cortical sheet comprising 8004-vertices. The results of this pipeline are eventually multiplied by the corresponding IC timecourses and linearly combined to obtain brain activity at each vertex. Specifically, the columns of the mixing matrix corresponding to the brain ICs are individually

projected into the source space using a Tikhonov-regularized minimum-norm estimator, according to:

$$\hat{s}_i = W^{-1}L^T(LW^{-1}L^T + \lambda_i I)^{\dagger} a_i$$

where a_i is the i -th IC topography, L is the leadfield matrix for the MEG forward problem, W is a diagonal weighting matrix the elements of which are defined as $W_j = ||L_j||$ (with L_j being the three field components for unit dipoles in three directions at the j -th voxel and the symbol $||$ the Frobenius norm), λ_i is the regularization parameter for each IC, I is the identity matrix and the symbols $^{-1}$, T and † indicate matrix inverse, transpose and pseudoinverse, respectively.

Before MNE projection, the IC sensor maps are scaled to norm 100, and the scaling factor is then applied back to the IC source maps. This scaling factor allows one to define a consistent noise level for all the mixing matrices, independently from the subject or the specific run. The forward solution is based on a spherical harmonic expansion (Lead Field Method) with location-wise SVD, and depth weighting normalization is applied using the Frobenius norm of the $[Lx_i, Ly_i, Lz_i]$ vector for each dipole location i . Prewhitening is applied to both the sensor maps and the leadfield, using a diagonal covariance matrix with elements equal to the noise level, which is set to 8 for our data scaled to $[-100, 100]$. The regularization λ_i is determined for each brain component according to (Lin et. al, 2004), from the covariance matrix and the signal to noise ratio (SNR) of each IC.

The result consists of a fieldtrip structure containing the three orientations of the current at the vertex sites for each independent component. This structure is stored in a Matlab file. Additionally, quality check figures showing the IC weights projected into the source space are produced by this pipeline. However, both the Matlab file and the quality check figures are not part of the release.

"icablpenv" pipeline

The "icablpenv" (ICA source level power envelope) pipeline estimates the band limited power envelope of the reconstructed source level signals, defined on the 8004-vertex surface-registered cortical sheet.

To this aim, both the output from the [icaclass](#) and [icamne](#) pipelines are required. First, the time courses obtained by the icaclass pipeline are band-pass filtered in the physiological oscillation bands, according to the table in [icaimagcoh pipeline](#) below.

High pass and low pass zero-phase forward and reverse Butterworth filters are subsequently applied. The filter order is optimized for each frequency band in order to obtain the highest slope compatible with the cutoff frequencies. Then, for each vertex of the cortical sheet and for each frequency band, the current source distribution over time is obtained for the 3 directions x, y, z

by linearly combining the filtered time courses of the brain components, weighted with their source space maps (obtained by the [icamne pipeline](#)). Subsequently, the power envelope is estimated by a moving average integrating the squared amplitude of the vertex signal. The integration window duration is 400 milliseconds and it slides with a time step of 20 milliseconds.

For each frequency band, the band limited power time-course is then saved in separate CIFTI files named *.power.dtseries.nii that can be displayed on the subject's cortex using Connectome Workbench. These files contain the dense cortical representation and their content can be open by Matlab through the ft_read_CIFTI function available in Fieldtrip. The pipeline also produces a .mat file containing the band limited power envelope structure. However, this file is not part of the release package.

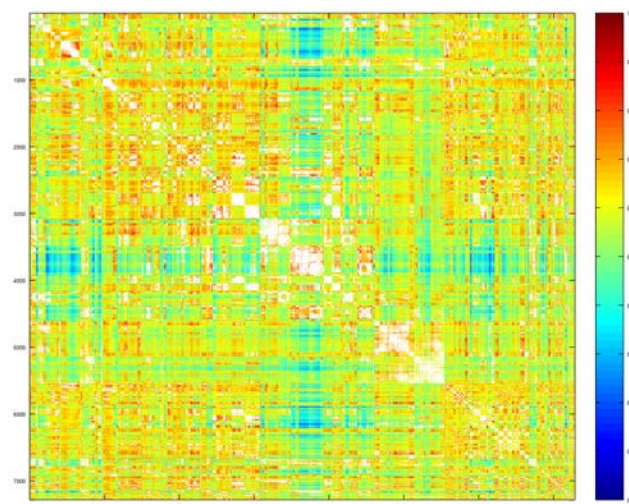
"icablpcorr" pipeline

The "icablpcorr" (Band Limited Power dense connectome) pipeline produces a dense connectome at the source-level through the Pearson correlation coefficient estimated on slowly varying band limited power time series, obtained with the icablpenv pipeline in the delta, theta, alpha, beta, low gamma, high gamma and wide bands (see the description of the [icablpenv pipeline](#) for the list of frequencies delimiting the oscillation bands).

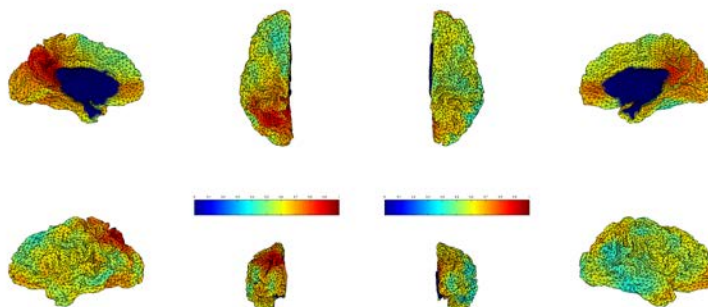
Dense connectomes are obtained by using all the three sessions of a single subject. Specifically, for every vertex pair on the surface-registered cortical sheet comprising 8004 vertices, Pearson correlation is estimated across epochs of 25 seconds of non-overlapping data. The correlation coefficients estimated over these epochs in each run are subsequently averaged across epochs and runs. The output of the pipeline at each frequency band is saved in CIFTI format, as a *.dconn.nii file, and can be displayed on the cortex using Connectome Workbench.

The pipeline also produces a Matlab file in which the connectome matrix is stored. Additionally, 2 bitmap figures are also saved for quality checking. The Matlab matrix file and the 2 figure files are not included in data for this release, however, examples of these files are shown here.

The first figure shows the dense connectome with the vertices ordered according to the RSN parcellation from the common consensus for one subject (right).



The second figure shows the correlation map obtained by selecting from the connectome matrix the row corresponding to the vertex associated to the posterior cingulate cortex (PCC) seed-region. An example for one subject is shown below:



"icaimagcoh" pipeline

The "icaimagcoh" (ICA source level imaginary coherence) pipeline generates frequency specific connectomes based on the "Multivariate Interaction Measure (MIM)" (Ewald et al., 2012; Marzetti et al., 2013) for source level data from non-artifactual independent components (as estimated from the [icamne pipeline](#)).

The "Multivariate Interaction Measure" (MIM) maximizes the imaginary part of coherence between MEG signal at a given reference voxel with respect to the signal at any other target voxel. More specifically, the estimated MEG signal at each brain voxel is a vector quantity that can be represented, in a given reference system, through its three components and the MIM is designed to maximize the imaginary part of coherence between vector quantities.

More specifically, the Source Level Imaginary Coherence pipeline consists of several conceptual steps:

1. The cross spectral density matrix (CSD) from independent component time series is estimated. The estimation is performed by segmenting the data into time epochs of 1.0s duration with 50% overlap and applying Fast Fourier analysis after signal linear de-trending and Hanning windowing. The CSD estimation is performed in the selected frequency band according to the definitions at right.

| Frequency band | Frequency range (Hz) |
|----------------|----------------------|
| delta | 1-4 |
| theta | 4-8 |
| alpha | 8-15 |
| beta low | 15-26 |
| beta high | 26-35 |
| gamma low | 35-50 |
| gamma mid | 50-76 |
| gamma high | 76-120 |

2. The CSD is projected to source level by appropriately multiplying it by the independent component topographies obtained by the [icamne](#) pipeline, thus resulting in the source space CSD being a matrix of size $3N \times 3N$ where N is the number of cortical nodes;
3. MIM is estimated between cortical node pairs according to equation 14 in (Ewald et al., 2012) and a connectome of size $N \times N$ for each frequency bin is thus derived;

Connectivity results are stored as CIFTI output.

"bfbldenv" pipeline

This pipeline estimates the source level band limited power from resting state data. In contrast to [icabldenv](#), which uses the output of [icamne](#), a different approach is adopted in the "bfbldenv" pipeline to project channel-level data into the source space. First, this pipeline uses the result of the [icaclass](#) pipeline to remove the components identified as artifacts from the raw data, which are preprocessed as at the input of [icaclass](#). A Linearly Constrained Minimum Variance (LCMV) beamforming method is then applied on the clean channel-level data to reconstruct source activity. A realistic, single shell head model obtained from the individual MRI is used in the reconstruction. When constructing the forward model, the Leadfield vectors have been normalized by their quadrance (squared euclidean norm) in order to remove the inherent bias of beamformer solutions towards the center of the brain. The beamformer solution is derived from the covariance matrix obtained from the cleaned time courses filtered in the physiological frequency band under consideration.

The output from the [icaclass](#) and [baddata](#) pipelines are required as input of this pipeline. After the artifacts are removed, the cleaned sensor level data are band-pass filtered in the physiological oscillation bands, according to the frequency band table above.

High pass and low pass zero-phase forward and reverse Butterworth filters are subsequently applied. The filter order is optimized for each frequency band in order to obtain the highest slope compatible with the cutoff frequencies.

Following the estimation of the filtered activity over the cortical sheet, the BLP calculation proceeds in exactly the same way as the [icabldenv](#) case. The power envelope is estimated by a moving average, integrating the squared amplitude of the vertex signal. The integration window duration is 400 milliseconds and it slides with time steps of 20 milliseconds. For each frequency band, the band limited power time-course is then saved in separate CIFTI2 files named *.power.dtseries.nii that can be displayed on the subject's cortex using [Connectome Workbench](#). These files contain the dense cortical representation and their content can be opened by Matlab through the `ft_read_cifti` function available in Fieldtrip. The pipeline also produces a .mat file containing the band limited power envelope structure. This file is not part of the release package.

"bfbldcorr" pipeline

This pipeline produces dense connectomes at the source-level through the Pearson correlation coefficient estimated on slowly varying band limited power time series using exactly the same procedures as the [icabldcorr](#) pipeline, but using [bfbldenv](#) results, rather than [icabldenv](#) ones.

"icabldpdyn" pipeline

This pipeline is intended to produce parcellated connectome level results by means of a sliding window approach calculating the Pearson-correlation connectivity measure on Band Limited Power data (from the [icabldenv](#) pipeline). Both the time window in which the correlation is evaluated and the step between two consecutive evaluations can be modified by users. Default values are respectively 10 and 0.5 seconds. The pipeline output file name contains the length of the window and the time step for which the correlation is evaluated. This pipeline is part of the [megconnectome v3.0 software release](#), but outputs of this pipeline are not included in the MEG data released as part of the 1200 Subjects release.

The pipeline needs a parcellation file as input and produces results according to it. In order to work properly, [Connectome Workbench](#) must be installed and `wb_command` must be made executable in matlab.

tMEG Channel-level Processing Pipelines

The tMEG Channel-level analysis pipelines `tmegpreproc`, `eravg`, and `tfavg` perform the following analysis steps:

- 1) divide the data from each tMEG experiment into groups of trials that represent various events of interest
- 2) use the outputs of the previously run preprocessing pipeline steps to clean the tMEG data (now grouped as events of interest)
- 3) compute the average (across trials/runs) event related fields and calculate averaged condition contrasts, at the sensor (channel) level
- 4) compute the average Time-Frequency response at the sensor level for conditions and contrasts of conditions in the Task-MEG scans

How are different conditions and contrasts (between conditions) of Task scans analyzed?

Contrasts are used in the tMEG pipelines so that all different conditions and comparisons of conditions at different stages of analysis may be processed with the same pipelines.

A contrast is a MATLAB structure representing the settings for performing a specific type for a specific condition or between 2 different conditions of a single experiment.

The settings for the analysis of a single condition or comparison between 2 conditions can vary in terms of the time periods, the baseline or comparison operation and other settings. So in order to add the flexibility for varying such parameters for a single condition or for the comparison of different conditions, more than one contrast is used for each single condition or for the comparison of different conditions.

The table below defines and describes the fields of the Contrast structure for the eravg and tfavg pipelines.

| Field Name | Print Mnemonic prefix | Description | Example | Possible Values | Default value |
|------------------|--|--|---------------------------------------|---|---------------|
| .pipeline | | name of pipeline where contrast is to be run | 'eravg' | 'eravg','tfavg' | [] |
| .lockmode | LM- | describes data group from which trials for contrast have been selected | {'TIM'} | different datagroups within each task, currently: For Motort: 'TEMG','TFLA'. For Workmem: 'TIM', 'TRESP'. For StoryM: 'TEV','TRESP','BSENT', 'BU' | {[]} |
| .mnemtrl | (The trial set mnemonics are appended after the lockmode mnemonic) | a mnemonic describing the specific sets of trials (conditions) in the contrast | {'0B', '2B'} (i.e. 0-Back, 2-Back) | trial set mnemonics for each of the different conditions used in the contrasts A table presenting all the currently considered contrasts is presented further below in this page. Check there for all the available contrasts. | {[]} |



| | | | | | |
|---------------------|--------------------|---|----------------------------|---|------|
| .operation | OP- | describes how metrics from the 2 different conditions will be compared – only meaningful where 2 conditions are compared within a contrast | 'diff','rel', or 'relch' | 'diff' (Difference), 'rel' (relative:ratio), 'relch' (relative change:ratio-1) | [] |
| .timeperiods | Not part of naming | describes time points or periods for which analysis will be performed If a one column vector (one column, many rows), analysis will be performed for each of these time points. If two columns exist, this is interpreted as time windows (each row a time window). | [-0.5 0 0 0.5 0.5 1] | no specific constraint This is only constrained by the data time information | {[]} |
| .timedef | TD- | used when time periods are defined in windows (see above) It defines how the individual time points within each window will be processed during the analysis. For example, when computing the coherence between brain areas in time window [0 0.5], coherence can be computed by estimating on spectral value for the entire window for each trial. Or the spectral coefficients can be computed for all time points within the window and they can all be used in the computation of coherence. | 'concat' | 'concat' (compute metric by all time points in windows concatenated for all trials) 'avg' (take the avg of all times within the window before computing the metric) 'sing' ('Pick a single value at the center of the window for each trial to perform analysis') | [] |

| | | | | | |
|----------------------|--------------------|---|--------------------------------|--|--------|
| .baseline | Not part of naming | contains the time window to be used as baseline in the contrast -- mostly used in single condition contrasts | { [0 0.5] } | no specific constraint This is only constrained by the data time information | { [] } |
| .baselinetype | BT- | defines by what operation the baseline will be applied | 'diff' | 'diff' (Difference) 'rel' (relative:ratio) 'relch' (relative change:ratio-1) | [] |
| .selection | Not part of naming | contains indices of the trials for the specific contrast conditions from the pool of all trials for the specific data group | { '[1:20]' , ... [21:40]' } | trial indices for each contrast are selected by the contrast function of each experiment | { [] } |
| .description | Not part of naming | contains a very short description of the contrast in a form that can be used for printing in a report -- automatically produced by the contrast function of each experiment | 'memberTypes : Left_Hand\n' | | [] |
| .mnemprint | | Print Mnemonic constructed from all the above fields and used for the naming of the produced files by the pipeline | 'tfavg_[LM-TFLA-LH]_[BT-diff]' | | [] |

All the available conditions and contrasts between conditions for each Task paradigm are extracted using a dedicated MATLAB function named as:

contrast_\$SCANMNEMONIC.m

where \$SCANMNEMONIC can be "Wrkmem", 'Motort' or 'StoryM'.

These functions extract the list of all contrasts for all pipelines.

"tmegpreproc" pipeline

The hcp_tmegpreproc.m pipeline script is the first level of processing of the task specific analysis. There are 3 different Task paradigms in the HCP MEG protocol, very different from each other in content and structure. In each of these task experiments there are different events whose perception by and effect on the operation of the human brain are subjects of research.

For example in the motor task, one type of event is the onset of the flashing symbol instructing the subject to move his/her fingers or toes. The processing and perception of this visual instruction is a fast process well time locked to the onset of the visual stimulus. However, the onset of the movement is not as well time locked to the corresponding instruction. Most of this time jittering is introduced by anticipation of the instruction timing, by the relatively short interstimulus interval (ISI) translated into the motor system not being entirely settled before the next instruction and by the brain networks that make sense of a perceived instruction and translate it into a motor command. A more appropriate event for studying the onset of the movement is the onset of the EMG signal from the activated muscle. The above example simply demonstrates that within the same dataset one can cut trials relative to different events in order to study different aspects of brain's operation. The same holds true also for the working memory and story/math (language processing) task.

In order to accommodate this diversity in the analysis of task MEG data, the first step is to divide the data from each experiment into groups of trials according to the different events of interest. Each of these groups is then cleaned from artifacts according to the results from the preceding analysis of `hcp_baddata.m` and `hcp_icaclass.m`. Each group of trials is then saved with a characteristic mnemonic and used by all subsequent analysis.

The following table outlines the existing division into groups of trials locked on different events in each experiment.

| Experiment | Data group mnemonic | Defining event | Trial duration | What can be studied |
|-----------------------|---------------------|--|---|---|
| Motor Task | TFLA | onset of flashing cross that instructs subject to perform movement by hand or foot | -1.2 to 1.2 sec relative to onset of flashing cross | <ul style="list-style-type: none"> brain networks of processing of the presented instruction activation flow from visual to motor areas brain networks of anticipation of the flashing cross onset |
| Motor Task | TEMG | onset of the EMG signal from hand or foot recorded muscles | -1.2 to 1.2 sec relative to onset of EMG signal | <ul style="list-style-type: none"> brain networks employed during movement execution functional mechanism and hierarchy between central and peripheral nervous system |
| Working Memory | TIM | onset of an image that the subject has to match or not with the target image | -1.5 to 2.5 sec relative to the onset of the image | <ul style="list-style-type: none"> networks employed during image processing (faces vs tools) networks of memory recall for different memory loads |
| Working Memory | TRESP | onset of button press by the subject | -1.5 to 1.5 sec relative to button press | <ul style="list-style-type: none"> mechanism of memory recall and decision making just prior to the response mechanism of memory storage after response when the memory load is dynamic (2-back) as |



| Experiment | Data group mnemonic | Defining event | Trial duration | What can be studied |
|-------------------|---------------------|---|--|--|
| | | | | <p>compared to static</p> <ul style="list-style-type: none"> motor network |
| Story/Math | TEV | onset of any task event during the question and option period in stories and math problems | -1.5 to 1.5 secs relative to story math event | <ul style="list-style-type: none"> networks employed during math calculations compare networks during first parts of a calculation and later parts with heavier load networks during the onset of a story sentence relative to that of a math problem sentence networks during the presentation of the correct answer as compared to the wrong one activated areas when a number is presented as compared to when an operand is presented |
| Story/Math | TRESP | onset of button press by the subject | -1.5 to 1.5 sec relative to button press | <ul style="list-style-type: none"> motor network when a finger has already been picked for response compare immediate responses with later responses which can mean uncertainty in the answer |
| Story/Math | BSENT | <p>trials contain entire sentences</p> <p>For stories, this is a sentence during narration without the option sentence at the end of the story.</p> <p>For math, this is the sentence of the math problem excluding the option sentence at the end.</p> | 1 sec before onset of sentence and ends 1 sec after its offset | <ul style="list-style-type: none"> compare how brain networks change at the later parts of the story as compared to earlier parts of a story compare how brain networks change at the later math problems in a block as compared to earlier math problems in a block |
| Story/Math | BUN | <p>trials containing entire Blocks of stimulus Units</p> <p>As stimulus unit is defined an entire story or an entire math problem including the option part.</p> | 1 sec before onset of unit and ends 1 sec after its offset | <ul style="list-style-type: none"> compare the overall networks deployed during the processing of stories vs the processing of math problems <p>The logic behind selecting entire units as trials comes from the fact that the total duration of stories is balanced with the total duration of math problems. However the number of math problems is significantly higher than the number</p> |

| Experiment | Data group mnemonic | Defining event | Trial duration | What can be studied |
|------------|---------------------|----------------|----------------|--|
| | | | | of stories and the inverse holds for their duration. So actual comparison between stories and math can be performed at the block level and not at event level. |

The definition of the parameters for all the groups of a specific scan are described in a configuration file named after the convention:

alltrialdefparameters_\$\$SCANMNEMONIC.m, where \$SCANMNEMONIC can be, for example, 'Wrkmem', 'StoryM' or 'Motort'. These files are located in the megconnectome/trial_functions/ folder.

These parameters are fed through ft_definetrial.m into the corresponding trial definition functions located in megconnectome/trial_functions/ and named in a corresponding manner as:

trialfun_\$\$SCANMNEMONIC.m

This function returns the starting and ending samples as well as the offset of each trial in a data group for the specific scan.

This information is then used to extract the group trials data by ft_preprocessing.m

The trial definition function also extracts additional information about each trial, which is stored in the data structure in the trialinfo field.

All the above data groups whose mnemonic starts with T* have fixed trial length, which is relatively short. For this reason the trials that coincide even partially with bad data segments are completely removed. The data groups whose mnemonic starts with B* have variable trial length and the represent longer blocks of data. For these data sets the number of trials is small. They represent entire blocks of stimuli, i.e. an entire Story or an entire Math problem. As the number of trials is small, trials that coincide with bad data segments are not entirely removed. In order to preserve the trial definition structure inside the data, these segments are replaced by NaNs (Not a Number).

The tmegpreproc pipeline extracts and saves the trial definition and additional information for each group. For each tMEG task, descriptions of trial subsets by data group produced by the tmegpreproc pipeline are available in [Appendix 7. tMEG: Extracting trial subsets from data produced by tmegpreproc pipeline](#).

The data for each group based on the above information are extracted and cleaned for bad segments, bad channels and eye and heart related Independent Components.

The data are resampled to one fourth of the original sampling frequency in order to reduce the size of the dataset. The new sampling frequency is 508.63 Hz.

The clean data for each data group are saved in a separate file to be used by the later steps of the pipeline.

The output file `$EXPERIMENTID_$SCANID_tmegpreproc_$GROUPMNEMONIC.mat` contains the clean data split in trials for the specific data group in FieldTrip format.

The output file `$EXPERIMENTID_$SCANID_tmegpreproc_trialinfo.mat` contains a structure 'trlInfo' which contains the trial definition and additional information for each of the data groups for a specific scan. This trial information describes the clean set of trials within each data group.

The ASCII output file `$EXPERIMENTID_$SCANID_tmegpreproc_rawtrialinfoQC.txt` contains in MATLAB readable format general trial information about the raw data file. For example, it contains the number of trials in each of the main conditions, the number of trials in which there was a response, etc. For the EMG defined trials in the motor task, it contains how many trials were identified by the EMG. This information can be used as Quality Control in order to identify that the triggers have been read correctly and that the correct number of trials has been identified.

"eravg" pipeline

The purpose of this pipeline is the computation of the average of the event related fields and for calculation of condition differences (contrasts) at the sensor (channel) level.

This function inputs the data created by the `hcp_tmegpreproc` pipeline. This data contains the entire set of trials for a given data group, i.e. for Working Memory the data group TIM contains all the trials locked on the onset of a stimulus image. Within this data structure there is also information, in the form of a matrix, about the sequence and type of each trial.

The `eravg` pipeline uses the `contrast` function for the given experiment, e.g., Working Memory (see above, [How are different conditions and contrasts between conditions of Task scans analyzed?](#)), to select the trials corresponding to the conditions that comprise a contrast. The `contrast` function returns a default set of pre-defined experimental comparisons and information with respect to the pipeline(s) in which it is to be computed. The same comparison between conditions (e.g. 2-Back vs 0-Back) can be made with different specifications (e.g. stimulus-versus response-locked) at various levels of analysis (e.g. channel or source). For the above-mentioned example, the average of the event related fields at the sensor level can be computed for every original data sample in the trial, however for time frequency analysis a lower time resolution suffices, e.g. 25 ms steps. Additionally, the frequencies for TFR are defined, e.g. 1 to 100 Hz in steps of 1 Hz. For connectivity analysis, this high time and frequency resolution is again more resolution than is necessary for analysis, therefore the analysis is performed in a

few time windows and in a few frequency bands. These are examples of the wide range of algorithmic specifications for the same comparison between conditions at the different analysis levels. To accommodate for this in the data analysis pipelines, a single contrast defines not only the selection of trials, but also the settings for the desired analyses in each of the pipelines.

The 'eravg' pipeline starts by loading the entire list of contrasts for a given experiment. Subsequently it selects the ones for the 'eravg' pipeline and feeds them into the core analysis function, named 'hcp_eravg_contrasts.m'.

The first task that this core function performs is to rearrange the list of contrasts so that those from the same data group are combined. This helps to avoid loading the same data files more than once, as this requires significant time. As the contrasts are clustered according to the data group, the data containing all the trials for one data group are only loaded once.

Once the data for all trials is loaded, the trials corresponding to the specific conditions of a contrast are selected and the event related average is computed and saved to disk.

At this time, the only operation implemented to compare different conditions in a contrast is to compute the difference in the mean. Furthermore, a contrast to the baseline is implemented for single conditions and for two-condition-contrasts, in which case the baseline correction is applied prior to the computation of the average of each individual condition. The only type of baseline operation implemented at this time is subtraction, i.e. absolute difference.

Prior to baseline correction, the ERF is band pass filtered between 0.5 and 60 Hz.

The 'eravg' pipeline computes the event related average for both the axial magnetic field and for the magnitude of the estimated planar gradient. The planar gradient magnitude provides a more interpretable localization of the underlying sources, on the sensor level.

Additionally the 'eravg' pipeline has been designed to combine both scans from a given paradigm, e.g. "10-Motort" and "11-Motort", computing the average over all selected trials in both scans. The results are saved using a SCANID that omits the individual run number, e.g. "Motort". If one of the scans is not available, the pipeline will operate on a single scan.

For the eravg pipelines, the following conditions and contrasts between conditions are computed:

Motor Task: The following contrasts are computed for both **data groups TFLA** (Trials locked on the onset of the Flashing Cross that instructs the subject to perform a movement) and **TEMG** (Trials locked on the onset of the EMG signal).

| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|-------|-------|-----------------------|-----------------------|-----------------|--|
| LH | | Left Hand | | -1.2 to 1.2 sec | \$EXPID_Motort_eravg_[LM-TFLA-LH]_[BT- |

| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|-----------|-------|-----------------------|-----------------------|--|--|
| | | | | with 508.63 Hz resolution baseline Period= [-0.3 0]; baseline Type = 'diff' | diff_[MODE-mag] (magnetic field) \$EXPID_Motort_eravg_[LM-TFLA-LH]_[BT- diff]_[MODE-planar](planar gradient) (for data group TEMG , LM-TFLA in the naming is replaced by LM-TEMG) |
| RH | | Right Hand | | same as above | \$EXPID_Motort_eravg_[LM-TFLA-RH]_[BT- diff]_[MODE-mag] (magnetic field) \$EXPID_Motort_eravg_[LM-TFLA-RH]_[BT- diff]_[MODE-planar](planar gradient) (for data group TEMG , LM-TFLA in the naming is replaced by LM-TEMG) |
| LF | | Left Foot | | same as above | \$EXPID_Motort_eravg_[LM-TFLA-LF]_[BT- diff]_[MODE-mag] (magnetic field) \$EXPID_Motort_eravg_[LM-TFLA-LF]_[BT- diff]_[MODE-planar](planar gradient) (for data group TEMG , LM-TFLA in the naming is replaced by LM-TEMG) |
| RF | | Right Foot | | same as above | \$EXPID_Motort_eravg_[LM-TFLA-RF]_[BT- diff]_[MODE-mag] (magnetic field) \$EXPID_Motort_eravg_[LM-TFLA-RF]_[BT- diff]_[MODE-planar](planar gradient) (for data group TEMG , LM-TFLA in the naming is replaced by LM-TEMG) |

The same contrasts are performed for the data group TEMG. For this group, LM-TFLA is replaced by LM-TEMG in the naming of the resultant files.

Working Memory: data group **TIM** (Trials locked on the onset of the image)

| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|-----------|-------|----------------------------|-----------------------|--|---|
| 0B | | 0-Back Memory Trials | | [-1.5 to 2.5] sec with 508.63 Hz resolution baseline Period=[- 0.5 0]; baseline Type = 'diff' | \$EXPID_Wrkmem_eravg_[LM-TIM-0B]_[BT- diff]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_eravg_[LM-TIM-0B]_[BT- diff]_[MODE-planar](planar gradient) |
| 2B | | 2-Back Memory | | same as above | \$EXPID_Wrkmem_eravg_[LM-TIM-2B]_[BT- diff]_[MODE-mag] (magnetic field) |

| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|-------------|-------------|----------------------|----------------------|---|--|
| | | Trials | | | \$EXPID_Wrkmem_eravg_[LM-TIM-2B]_[BT-diff]_[MODE-planar] (planar gradient) |
| face | | face image trials | | same as above | \$EXPID_Wrkmem_eravg_[LM-TIM-face]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_eravg_[LM-TIM-face]_[BT-diff]_[MODE-planar] (planar gradient) |
| tool | | tool image trials | | same as above | \$EXPID_Wrkmem_eravg_[LM-TIM-tool]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_eravg_[LM-TIM-tool]_[BT-diff]_[MODE-planar] (planar gradient) |
| OB | 2B | 0-Back Memory Trials | 2-Back Memory Trials | [-1.5 to 2.5] sec with 508.63 Hz resolution baseline Period=[- 0.5 0]; baseline Type = 'diff' operation:'diff' | \$EXPID_Wrkmem_eravg_[LM-TIM-OB-versus-2B]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_eravg_[LM-TIM-OB-versus-2B]_[OP-diff]_[BT-diff]_[MODE-mag] (planar gradient) |
| face | tool | face image trials | tool image trials | same as above | \$EXPID_Wrkmem_eravg_[LM-TIM-face-versus-tool]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_eravg_[LM-TIM-face-versus-tool]_[OP-diff]_[BT-diff]_[MODE-planar] (planar gradient) |

Working Memory: data group **TRESP** (Trials locked on the onset of the button press by the subject). Similar to the TIM data group above, with the only difference in the baseline interval used. In each of the eravg contrasts of the TRESP data group the baseline is set to interval [- 0.75 to -0.25] seconds. For this group in the naming of the result files the part LM-TIM is replaced by LM-TRESP.

Story/Math: data group **TEV** (Onset of any task event during the question and option period in stories and math problems).

| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|-------------------|-------|------------------------------------|--------------------|---|--|
| mathnumque | | Onset of numbers in math questions | | [-1.5 to 4] sec with 508.63 Hz resolution baseline Period=[- 0.5 0]; | \$EXPID_StoryM_eravg_[LM-TEV-mathnumque]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-mathnumque]_[BT-diff]_[MODE-planar] (planar gradient) |



| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|-----------------------|------------------------|--|--|--|--|
| | | | | baseline Type = 'diff' | |
| storsentnon | | Onset of sentence in stories (not option part) | | same as above | \$EXPID_StoryM_eravg_[LM-TEV-storsentnon]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-storsentnon]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathsentnon | | Onset of sentence in math question (not option part) | | same as above | \$EXPID_MEG_StoryM_eravg_[LM-TEV-mathsentnon]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_MEG_StoryM_eravg_[LM-TEV-mathsentnon]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathnumopt | | Onset of numbers in math options | | same as above | \$EXPID_StoryM_eravg_[LM-TEV-mathnumopt]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-mathnumopt]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathoper | | Onset of operand (+ -) in math questions | | same as above | \$EXPID_StoryM_eravg_[LM-TEV-mathoper]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-mathoper]_[BT-diff]_[MODE-planar] (planar gradient) |
| storsentnon | mathsentnon | Onset of sentence in stories (not option part) | Onset of sentence in math question (not option part) | [-1.5 to 4] sec with 508.63 Hz resolution baseline Period = [-0.5 0]; baseline Type = 'diff' operation : 'diff' | \$EXPID_StoryM_eravg_[LM-TEV-storsentnon-versus-mathsentnon]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-storsentnon-versus-mathsentnon]_[OP-diff]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathnumquelate | mathnumqueearly | Onset of the | Onset of | same as | \$EXPID_StoryM_eravg_[LM-TEV- |



| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|----------------------|----------------------|---|---|---------------|--|
| | | latter half of numbers in math questions | the earlier half of numbers in math questions | above | mathnumquelate-versus-mathnumqueearly_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-mathnumquelate-versus-mathnumqueearly]_[OP-diff]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathnumoptcor | mathnumoptwro | Onset of correct option numbers in math options | Onset of wrong option numbers in math options | same as above | \$EXPID_StoryM_eravg_[LM-TEV-mathnumoptcor-versus-mathnumoptwro]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-mathnumoptcor-versus-mathnumoptwro]_[OP-diff]_[BT-diff]_[MODE-planar] (planar gradient) |
| storoptcor | storoptwro | Onset of correct option in story options | Onset of wrong option in story options | same as above | \$EXPID_StoryM_eravg_[LM-TEV-storoptcor-versus-storoptwro]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-storoptcor-versus-storoptwro]_[OP-diff]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathnumque | mathoper | Onset of numbers in math questions | Onset of operand (+ -) in math questions | same as above | \$EXPID_StoryM_eravg_[LM-TEV-mathnumque-versus-mathoper]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-mathnumque-versus-mathoper]_[OP-diff]_[BT-diff]_[MODE-planar] (planar gradient) |

Story/Math: data group **TRESP** (Onset of Button press in response).

| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|-------|-------|--------------------|--------------------|---------|------------------------|
|-------|-------|--------------------|--------------------|---------|------------------------|

| | | | |
|-----|---------------------------|--|--|
| all | all response trials | [-1.5 to 1.5] sec with 508.63 Hz resolution baseline Period= [-0.75 -0.25]; baseline Type = 'diff' | \$EXPID_StoryM_eravg_[LM-TRESP-all]_[BT- diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TRESP-all]_[BT- diff]_[MODE-planar] (planar gradient) |
|-----|---------------------------|--|--|

"tfavg" pipeline

The purpose of this pipeline is the computation of the average Time-Frequency response at sensor level for a set of conditions and contrasts of conditions in the Task-MEG scans.

This function inputs the data created by the `hcp_tmegpreproc` pipeline. These data contain the entire set of trials for a given data group, i.e. for Working Memory the data group "TIM" contains all the trials locked on the onset of a stimulus image. Within this data structure there is also information, in the form of a matrix, about the sequence and type of each trial.

The `tfavg` uses the contrast function for the given experiment, e.g., Working Memory, to select the trials corresponding to the conditions that comprise a contrast. The contrast functions returns a default set of pre-defined experimental comparisons and information with respect to the pipeline(s) in which it is to be computed. The same comparison between conditions (e.g. 2-Back vs 0-Back) can be made with different specifications (e.g. stimulus versus response locked) at various levels of analysis (e.g. channel or source). For the above mentioned example, the average of the Time-Frequency response at the sensor level can be computed for every original data sample in the trial, however a lower time resolution suffices, e.g. 25 ms steps. Additionally the frequencies for TFR are defined in the range 1 to 100 Hz in steps of 1 Hz.

The 'tfavg' pipeline starts by loading the entire list of contrasts for a given experiment. Subsequently it selects the ones for the 'tfavg' pipeline and feeds them into the core analysis function, named 'hcp_tfavg_contrasts.m'.

The first task that this core function performs is to rearrange the list of contrasts so that those from the same data group are clustered so that the data containing all the trials for one data group is only loaded once. This helps to avoid loading the same data files multiple times, as this requires significant time. When the data for all trials are loaded, the trials corresponding to the specific conditions of a contrast are selected and the event related average is computed and stored to disk.

At this time there are 2 operations implemented for comparison between different conditions: difference and ratio. Furthermore, a contrast to the baseline is implemented for single conditions and for two-condition contrasts, in which case the baseline correction is applied prior to the computation of the average of each individual condition. The only type of baseline operation implemented at this time is ratio.

The 'tfavg' pipeline computes the average Time-Frequency for both the axial magnetic field and for the magnitude of the estimated planar gradient. The planar gradient magnitude provides on the sensor level a more interpretable localization of the underlying sources.

Additionally the 'tfavg' pipeline has been designed to combine both scans from a given paradigm, e.g. "10-Motort" and "11-Motort", computing the average over all selected trials in both scans. The results are saved using a SCANID that omits the individual run number, e.g. "Motort". In case one of the scans is not available, the pipeline will operate on a single scan.

Time-frequency analysis is performed using the following settings:

| TFR analysis settings | | |
|-----------------------|-------------|---|
| Frequencies | f | 1 to 100 Hz in steps of 1 Hz |
| Time Step | tstep | 0.025 sec (time resolution) |
| Time Window | 5./f | For each different frequency a different time window length is used for the computation of TFR at a time point. The length of the time window is set to 4 periods of the given frequency. For low frequencies this time window length becomes long so a threshold has been set to 500 msec as the longest window. |
| TFR method | 'mtmconvol' | multitaper time-frequency transformation based on multiplication in the frequency domain |
| TFR taper type | 'hanning' | Hanning window taper |

For the tfavg pipelines the following conditions and contrasts between conditions are computed:

Motor Task: The following contrasts are computed for both data groups **TFLA** (Trials locked on the onset of the Flashing Cross that instructs the subject to perform movement) and **TEMG** (Trials locked on the onset of the EMG signal).

| Cond1 | Cond2 | Cond1 Description | Cond2 Description | Details | Naming of Output Files |
|-----------|-------|-------------------|-------------------|---|--|
| LH | | Left Hand | | Time points =[-1.2:0.025:1.2] sec; 100 frequencies=[1:1:100]; baseline Period=[-0.3 0]; baseline type:'diff' | \$EXPID_Motort_tfavg_[LM-TLFA-LH]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_Motort_tfavg_[LM-TLFA-LH]_[BT-diff]_[MODE-planar] (planar gradient) |
| RH | | Right Hand | | same as above | \$EXPID_Motort_tfavg_[LM-TLFA-RH]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_Motort_tfavg_[LM-TLFA-RH]_[BT-diff]_[MODE-planar] (planar gradient) |

| | | | |
|-----------|------------|--|--|
| LF | Left Foot | same as above | \$EXPID_Motort_tfavg_[LM-TLFA-LF]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_Motort_tfavg_[LM-TLFA-LF]_[BT-diff]_[MODE-planar] (planar gradient) |
| RF | Right Foot | same as above | \$EXPID_Motort_tfavg_[LM-TLFA-RF]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_Motort_tfavg_[LM-TLFA-RF]_[BT-diff]_[MODE-planar] (planar gradient) |
| LH | Left Hand | Time points = [-1.2:0.025:1.2] sec; 100 frequencies=[1:1:100]; Connectivity Metric: emgcoh (Coherence with EMG signal) | \$EXPID_Motort_tfavg_[LM-TLFA-LH]_[CM-emgcoh]_[MODE-mag] (magnetic field) \$EXPID_Motort_tfavg_[LM-TLFA-LH]_[CM-emgcoh]_[MODE-planar] (planar gradient) |
| RH | Right Hand | same as above | \$EXPID_Motort_tfavg_[LM-TLFA-RH]_[CM-emgcoh]_[MODE-mag] (magnetic field) \$EXPID_Motort_tfavg_[LM-TLFA-RH]_[CM-emgcoh]_[MODE-planar] (planar gradient) |
| LF | Left Foot | same as above | \$EXPID_Motort_tfavg_[LM-TLFA-LF]_[CM-emgcoh]_[MODE-mag] (magnetic field) \$EXPID_Motort_tfavg_[LM-TLFA-LF]_[CM-emgcoh]_[MODE-planar] (planar gradient) |
| RF | Right Foot | same as above | \$EXPID_Motort_tfavg_[LM-TLFA-RF]_[CM-emgcoh]_[MODE-mag] (magnetic field) \$EXPID_Motort_tfavg_[LM-TLFA-RF]_[CM-emgcoh]_[MODE-planar] (planar gradient) |

The same contrasts are performed for the data group **TEMG**. For this group in the naming of the result files the part LM-TFLA is replaced by LM-TEMG.

Working Memory: data group: **TIM** (Trials locked on the onset of the image)

| Cond1 | Cond2 | Cond1 Description | Cond1 Description | Details | Naming of Output Files |
|-------------|-------------|----------------------|----------------------|---|--|
| 0B | | 0-Back Memory Trials | | 100 frequencies=[1:1:100]; Time points = [-1:0.025:2] sec baseline Period=[-0.5 0]; baseline type:'diff' | \$EXPID_Wrkmem_tfavg_[LM-TIM-0B]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_tfavg_[LM-TIM-0B]_[MODE-planar] (planar gradient) |
| 2B | | 2-Back Memory Trials | | same as above | \$EXPID_Wrkmem_tfavg_[LM-TIM-2B]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_tfavg_[LM-TIM-2B]_[MODE-planar] (planar gradient) |
| face | | face image trials | | same as above | \$EXPID_Wrkmem_tfavg_[LM-TIM-face]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_tfavg_[LM-TIM-face]_[MODE-planar] (planar gradient) |
| tool | | tool image trials | | same as above | \$EXPID_Wrkmem_tfavg_[LM-TIM-tool]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_tfavg_[LM-TIM-tool]_[MODE-planar] (planar gradient) |
| 0B | 2B | 0-Back Memory Trials | 2-Back Memory Trials | 100 frequencies=[1:1:100]; Time points = [-1:0.025:2] sec operation:'diff' | \$EXPID_Wrkmem_tfavg_[LM-TIM-0B-versus-2B]_[OP-diff]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_tfavg_[LM-TIM-0B-versus-2B]_[OP-diff]_[MODE-planar] (planar gradient) |
| face | tool | face image trials | tool image trials | same as above | \$EXPID_Wrkmem_tfavg_[LM-TIM-face-versus-tool]_[OP-diff]_[MODE-mag] (magnetic field) \$EXPID_Wrkmem_tfavg_[LM-TIM-face-versus-tool]_[OP-diff]_[MODE-planar] (planar gradient) |

Working Memory: data group **TRESP** (Trials locked on the onset of the button press by the subject)

Similar to the TIM data group above, with only difference in the time interval used. In each of the tfavg contrasts of the TRESP data group there are 121 timepoints in the interval [-1.25 to 1.75] seconds with time step of 25 mseconds. For this group in the naming of the result files the part LM-TIM is replaced by LM-TRESP.



Story/Math: data group: **TEV** (Onset of any task event during the question and option period in stories and math problems.)

| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|--------------------|-------|--|-----------------------|--|--|
| mathnumque | | Onset of numbers in math questions | | 100 frequencies= [1:1:100]; Time points = [-1:0.025:3.5] sec baseline Period=[-0.5 0]; baseline Type = 'diff' | \$EXPID_StoryM_eravg_[LM-TEV-mathnumque]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_eravg_[LM-TEV-mathnumque]_[BT-diff]_[MODE-planar] (planar gradient) |
| storsentnon | | Onset of sentence in stories (not option part) | | same as above | \$EXPID_StoryM_tfavg_[LM-TEV-storsentnon]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_tfavg_[LM-TEV-storsentnon]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathsentnon | | Onset of sentence in math question (not option part) | | same as above | \$EXPID_MEG_StoryM_tfavg_[LM-TEV-mathsentnon]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_MEG_StoryM_tfavg_[LM-TEV-mathsentnon]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathnumopt | | Onset of numbers in math options | | same as above | \$EXPID_StoryM_tfavg_[LM-TEV-mathnumopt]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_tfavg_[LM-TEV-mathnumopt]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathoper | | Onset of operand (+ -) in math questions | | same as above | \$EXPID_StoryM_tfavg_[LM-TEV-mathoper]_[BT-diff]_[MODE-mag] (magnetic field) |



| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|-----------------------|------------------------|---|--|--|--|
| | | | | | \$EXPID_StoryM_tfavg_[LM-TEV-mathoper]_[BT-diff]_[MODE-planar] (planar gradient) |
| storsentnon | mathsentnon | Onset of sentence in stories (not option part) | Onset of sentence in math question (not option part) | 100 frequencies=[1:1:100]; Time points = [-1:0.025:3.5] sec operation : 'diff' | \$EXPID_StoryM_tfavg_[LM-TEV-storsentnon-versus-mathsentnon]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_tfavg_[LM-TEV-storsentnon-versus-mathsentnon]_[OP-diff]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathnumquelate | mathnumqueearly | Onset of the latter half of numbers in math questions | Onset of the earlier half of numbers in math questions | same as above | \$EXPID_StoryM_tfavg_[LM-TEV-mathnumquelate-versus-mathnumqueearly]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_tfavg_[LM-TEV-mathnumquelate-versus-mathnumqueearly]_[OP-diff]_[BT-diff]_[MODE-planar] (planar gradient) |
| mathnumoptcor | mathnumoptwro | Onset of correct option numbers in math options | Onset of wrong option numbers in math options | same as above | \$EXPID_StoryM_tfavg_[LM-TEV-mathnumoptcor-versus-mathnumoptwro]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_tfavg_[LM-TEV-mathnumoptcor-versus-mathnumoptwro]_[OP-diff]_[BT-diff]_[MODE-planar] (planar gradient) |
| storoptycor | storoptywro | Onset of correct option in story options | Onset of wrong option in story options | same as above | \$EXPID_StoryM_tfavg_[LM-TEV-storoptycor-versus-storoptywro]_[OP-diff]_[BT-diff]_[MODE-mag] (magnetic field) \$EXPID_StoryM_tfavg_[LM-TEV-storoptycor-versus- |

| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|-------------------|-----------------|------------------------------------|--|---------------|--|
| | | | | | storoptywo][OP-diff][BT-diff][MODE-planar] (planar gradient) |
| mathnumque | mathoper | Onset of numbers in math questions | Onset of operand (+ -) in math questions | same as above | \$EXPID_StoryM_tfavg_[LM-TEV-mathnumque-versus-mathoper][OP-diff][BT-diff][MODE-mag] (magnetic field) \$EXPID_StoryM_tfavg_[LM-TEV-mathnumque-versus-mathoper][OP-diff][BT-diff][MODE-planar] (planar gradient) |

Story/Math: data group TRESP (Onset of Button press in response.)

| Cond1 | Cond2 | Cond 1 Description | Cond 2 Description | Details | Naming of Output Files |
|------------|-------|-----------------------|-----------------------|--|--|
| all | | all response trials | | 100 frequencies=[1:1:100]; Time points = [-1.25:0.025:1.25] sec baseline Period= [-0.75 -0.25]; baseline Type = 'diff' | \$EXPID_StoryM_tfavg_[LM-TRESP-all][BT-diff][MODE-mag] (magnetic field) \$EXPID_StoryM_tfavg_[LM-TRESP-all][BT-diff][MODE-planar] (planar gradient) |

How was the E-Prime trigger sequence reproduced for the scans in which no parallel port triggers were recorded?

In the Working Memory and Motor task scans, the signal in the TRIGGER channel consists of 2 superimposed bit patterns. One from the stimulus presentation computer that is running the E-Prime protocol, the other is from a photodiode placed in a corner on the stimulus presentation screen.

The E-Prime triggers encode all the events that are of interest within a scan. The photodiode triggers contain a subset of these events, corresponding mainly to the onset of images on the screen. For the Working Memory scans, the photodiode is on when a stimulus image is presented and is off otherwise. For the Motor Task scans, the photodiode is on when the Cue Screen (e.g. Left Hand) or the pacing flashing cross screen for each movement are presented, and the photodiode is off otherwise.

The photodiode triggers provide more accurate timing than the E-Prime triggers, because they measure the instance of the actual stimulus onset on the screen and do not represent the time

at which the display command was sent by the software. However, the value of the photodiode trigger can take only two values: 255 for ON and 0 for OFF. In contrast, the E-Prime triggers values are diverse and encode multiple aspects of the experimental sequence, such as stimulus content and task demands.

As a consequence of the two superimposed triggers, the event code sequence is extracted based on the E-Prime triggers, but its timing is extracted from the corresponding photodiode triggers.

Due to hardware problems, for some subjects during the Working Memory and Motor tasks, only the photodiode triggers were recorded in the TRIGGER channel in the raw MEG data file. This happened for the following subjects:

158035
189349
191841
250427
352132
352738
500222

For these subjects, the the E-Prime log files that are created during stimulus presentation are used to extract the E-Prime trigger information and reconstruct the E-Prime triggers as follows: The E-Prime log files describe the sequence of events with time stamps referenced to the beginning of the stimulation script run. These time stamps are represented with 1 Khz resolution. Based on these timestamps, the parallel port trigger sequence can be reproduced.

As the start of the E-Prime stimulation script does not correspond precisely to the start of the MEG acquisition, the log file timing needs to be aligned to the MEG data file TRIGGER channel by means of the photodiode triggers.

First the sequence of events from the E-Prime log file was extracted. The trigger events for which the corresponding photodiode events exist were identified. Then the E-Prime timestamps and the photodiode trigger timestamps for these events were referenced to the very first of these events, so that the first event time stamp was 0 for both. Subsequently the slope between the E-Prime and 4D timestamps for these events was computed. Based on this slope the E-Prime event sequence timing was transformed to the 4D sampling resolution of 2KHz. Then based on their common events the e-prime and photodiode sequences were aligned in time and an accurate mapping from E-Prime timestamps to 4D timestamps was derived.

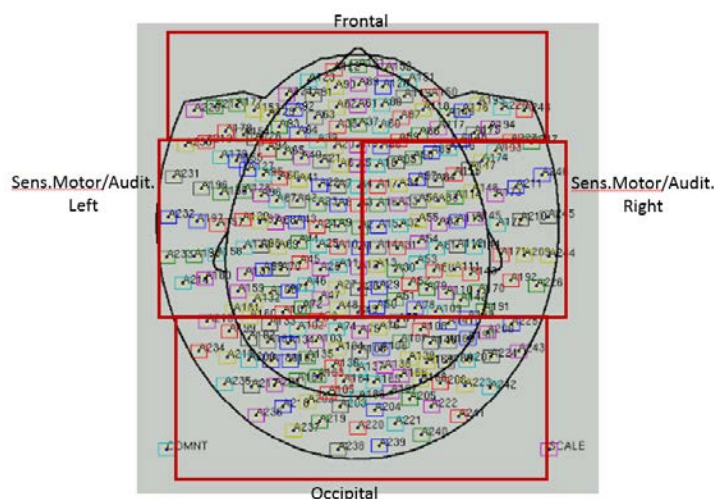
The reconstructed E-Prime trigger sequence (expressed in MEG timestamps) was superimposed on the existing photodiode trigger sequence and then written in the TRIGGER channel of a new version of the raw data file. This version is the one available for download.

Alpha Beta peak frequency processing

Alpha and beta frequency bands represent two of the strongest oscillations in the human brain. Alpha oscillations are believed to have an inhibitory role and to originate from the thalamus. Beta oscillations are mostly dominant in the sensorimotor cortex and are associated with a "status quo" signaling mechanism. Alpha and beta peak frequencies were calculated for each subject in the 2 pipelines below.

avgspecextract pipeline

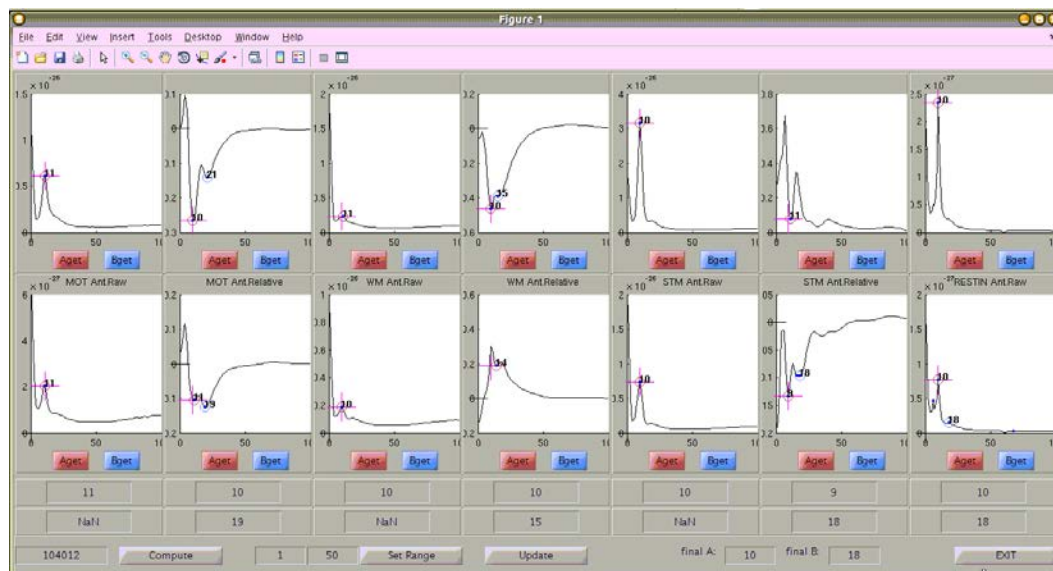
In this pipeline, the sensor array is divided into 4 quadrants (2 anterior and 2 posterior, see figure). Spectra are averaged for each group of sensors. For each Task, results of the [tfavg pipeline](#) (time-frequency spectra) are averaged within each sensor group and across time. The same averaging is applied for the results of the [powavg pipeline](#) for Resting state scans, which contain only spectral information (non-time resolved).



avgspecmanualalphabeta pipeline

This pipeline uses the Graphical User Interface (GUI) function `avgspecmanualalphabeta` where the spectra across all tasks can be assessed for a given subject. tMEG and rMEG average spectra from the [avgspecextract pipeline](#) are loaded into the GUI where spectral peaks in the alpha and beta band are automatically estimated.

In the GUI, the first 2 columns of plots correspond to MOTOR task, the next 2 correspond to the WORKING MEMORY task, the next 2 to the STORY MATH task and the last column to the RESTING STATE. The upper rows show data from the posterior quadrant sensors and the bottom rows show data from the anterior sensors. After automatically determining an alpha and a beta peak for each of the 7 conditions represented in the seven GUI columns, the overall alpha and beta peaks are determined by taking the median of all the alpha peaks and all the beta peaks, respectively. However, using the GUI, the user may also manually set and save the peaks, giving the user the ability to get an overview of the spectra in all different tasks and resting state so that the choice of alpha and beta is more informed.



The automatically estimated median values for peak alpha and peak beta across all tMEG and rMEG scans are reported as individual subject data in ConnectomeDB under the category “MEG Subjects”.

tMEG Source-level Processing Pipelines

The tMEG Source-level processing pipelines “srcavgdics”, “srcavgldmv”, and “tmegconnebasic” 1) compute the power series of the average event related fields in source space for the main conditions of each task, 2) estimate time courses of frequency band limited power in source space for the main conditions of each task, and 3) estimate time-resolved dense connectomes in a number of frequency bands at the level of the neuronal sources for the task MEG scans.

“srcavgdics” pipeline

The purpose of the “srcavgdics” (Source level average narrowband induced activity) pipeline is to estimate time courses of frequency band limited power at the level of the neuronal sources for the task MEG scans. These activity time courses are time-locked to task-specific events-of-interest. The pipeline uses the data created by the [tmegpreproc pipeline](#). This is sensor-level data, which contains the entire set of trials for a given data group, e.g. for Working Memory the data group TIM contains all the trials locked to the onset of a stimulus image.

For each main condition in the Motor Task and Working Memory tasks (e.g. left hand/right hand/left foot/right foot in the Motor Task) and each frequency band (see below) a separate file is generated.

The results are saved in CIFTI-2 format. This format was chosen so that results can be directly read by the [Connectome Workbench software](#), in which results from other modalities (structural-MR, fMRI, dMRI) can be also read and overlaid.

The power series are derived by projecting the sensor level Cross spectral density in source space through an inverse solution based on a Dynamic Imaging of Coherent Sources (DICS) beamformer.

The power series for the main trial conditions are derived by projecting the sensor level time-resolved cross-spectral density matrices in source space through an inverse solution based on a Dynamic Imaging of Coherent Sources (DICS) beamformer. This inverse solution is derived based on a single shell forward model. In this forward model, the Leadfield vectors have been normalized by their quadrance (squared euclidean norm) in order to remove the inherent bias towards the center of the brain by beamformer solutions.

For all conditions, the beamformer solution is derived from the average cross-spectral density matrix at the sensor level. A different inverse solution is computed for each Frequency Band (see [FAQ for using MEG Data](#)). In this release, the main conditions for which source level power series have been computed are:

Motor Task: Left Hand, Right Hand, Left Foot, Right Foot, Fixation

Working Memory Task: 0-Back, 2-Back, Faces, Tools, Fixation

For the Fixation blocks, there is no series of power values but rather a single value for each voxel. This is because there are no events in the Fixation blocks. Therefore, one power value for each voxel is computed for this case. This power is only computed in order to serve as a contrast for the rest of the conditions.

The output source level data has not been referenced with respect to any baseline at source level. Any baseline defined is only used for plotting Quality Control figures.

For the Fixation blocks, there is no series of power values but rather a single power value for each voxel. This is because there are no events in the Fixation blocks. Therefore, one power value for each voxel is computed for this case. This power is experimentally only computed in order to serve as a contrast for the rest of the conditions.

The 'srvavgdics' pipeline has been designed to combine both scans from a given paradigm, e.g. "10-Motort" and "11-Motort", computing the average over all selected trials in both scans. The results are saved using a SCANID that omits the individual run number, e.g. "Motort". In case one of the scans is not available, the pipeline will operate on a single scan.

The data is originally represented in Tesla units, which results in values with high negative exponents. This is because the multiplication by the inverse filters further reduces the values

and because the squaring of the projected values for power computation doubles the value of the negative exponent. The resulting values are in the order of 10^{-41} . For such range of results, the color scaling in Connectome Workbench appears problematic, therefore, to facilitate direct visualization of results in this platform, the power is scaled through multiplication by 10^{42} prior to being saved.

Time-frequency analysis is performed using the following settings:

TFR analysis settings

| | | |
|----------------|-------------|---|
| Time Window | 3./f | For each different frequency a different time window length was used. The length of the time window is set to 3 periods of the given frequency. For low frequencies this time window length becomes long so a threshold has been set to 500 msec as the longest window. |
| TFR method | 'mtmconvol' | Single taper time-frequency transformation based on multiplication in the frequency domain. |
| TFR taper type | 'hanning' | Hanning window taper |

Main Steps

Load the entire list of contrasts for a given experiment and selects the ones that are set for the 'eravg' pipeline. Call core function `hcp_srcavglcmv_contrasts.m`.

- 2a. Arrange the list of contrasts so that those from the same data group are stacked together.
- 2b. The clean data trials are loaded for the current data group.
- 2c. For each contrast in this data group select the trials for the corresponding conditions.
- 2e. Compute fourier coefficients for each trial
- 2f. Compute Cross spectral density per time instance in trial
- 2g. Compute average Cross Spectral density for all trial time instances
- 2h. Compute inverse solution with DICS
- 2i. Project Cross spectral Density for each time instance and save power value
- 2j. Save data file with bandlimited power and image file with figure.
- 2k. Proceed to same process for next frequency band
- 2h Proceed to the next contrast from the same data group.
- 2i. If all contrasts from data group have been processed, then load data from the next data group and perform steps 2a to 2i for this new data group.

"srcavglcmv" pipeline

The purpose of the "srcavglcmv" (Source level average broadband evoked activity) pipeline is the computation of the power of the average event related fields in source space. It produces a time series of power for each voxel in source space.

This function inputs the data created by the [tmegpreproc pipeline](#). This data contains the entire set of trials for a given data group, i.e. for Working Memory the data group TIM, it contains all the trials locked on the onset of a stimulus image. Within this data structure, there is also information, in the form of a matrix, about the sequence and type of each trial.

The srcavglcmv pipeline produces event related power series for the main conditions of each task (i.e. LH, RH, LF, RF) and saves them in a Matlab format.

The results are saved in CIFTI-2 format. This format was chosen so that results can be directly read by the [Connectome Workbench software](#), in which results from other modalities (structural-MR, fMRI, dMRI) can be also read and overlaid.

The power series for the main trial conditions are derived by projecting the sensor level Event Related Field in source space through an inverse solution based on a Linearly Constrained Minimum Variance (LCMV) beamformer. This inverse solution is derived based on a single shell forward model. In this forward model, the Leadfield vectors have been normalized by their quadrance (squared euclidean norm) in order to remove the inherent bias towards the center of the brain by beamformer solutions.

For the majority of the main conditions, the beamformer solution is derived from the covariance matrix used for the inverse solution derived from the average Event Related Fields. In this release, the main conditions for which source level power series have been computed are:

Motor Task: Left Hand, Right Hand, Left Foot, Right Foot, Fixation

Working Memory Task: 0-Back, 2-Back, Faces, Tools, Fixation

For the Fixation blocks, there is no series of power values but rather a single value for each voxel. This is because there are no events in the Fixation blocks. Therefore, one power value for each voxel is computed for this case. This power is experimentally only computed in order to serve as a contrast for the rest of the conditions.

For the Fixation trials, 2 different cases are computed. In one case, the fixation blocks are split in pseudo trials of the same length and averaged and the inverse solution is computed on this average pseudo-Event Related Field, in order to have the same signal to noise ratio with the other conditions. Of course, this average does not produce any meaningful Event Related Field, but it computes a type of bias from areas that are persistently active during the fixation period. In addition to that, for the Fixation trials the power is also computed from the covariance matrix

of all data points (and not of the pseudo ERF) in order to capture the power of strong induced activity.

The output source level data has not been referenced with respect to any baseline at source level. Any baseline defined is only used for plotting Quality Control figures.

The 'sravavglcmv' pipeline has been designed to combine both scans from a given paradigm, e.g. "10-Motort" and "11-Motort", computing the average over all selected trials in both scans. The results are saved using a SCANID that omits the individual run number, e.g. "Motort". In case one of the scans is not available, the pipeline will operate on a single scan.

The data is originally represented in Tesla units which results in values with high negative exponents. This is because the multiplication by the inverse filters further reduces the values and because the squaring of the projected values for power computation doubles the value of the negative exponent. The resulting values are in the order of 10^{-41} . For such range of results, the color scaling in Connectome Workbench appears problematic, therefore, to facilitate direct visualization of results in this platform, the power is scaled through multiplication by 10^{42} prior to being saved.

Main Steps

The pipeline performs the following steps:

- Read list of input contrasts
- Sort contrasts so that they are grouped in the list according to the data group they apply so that contrasts from the same data group are processed in sequence and data is not loaded every time,
- LOOP over contrasts : For each Contrast DO:
 - + Check if current contrast is from same data group as previous processed contrast
 - IF SAME: Don't load anything just use the already loaded data
 - IF NOT SAME: Load data data for current data group.
 - If this is the first processed contrast then use gradiometer definition from loaded data, load source space model and compute leadfield
 - + Get subset of trials for given contrast from both session files and merge
 - + Get the type of inverse solution: filter computed from covariance matrix of the average over trials OR of all data points. Indicated in variable "invFiltFlag"
 - + Apply appropriate filtering, depending on if the source localization will be based on the ERF or on the entire data
 - + Get timing and baseline settings for current contrast.

- + Select data portion that spans from the earliest to the latest timepoints or timeperiod boundaries (and baseline boundaries, if provided).
- + Compute covariance Matrix
 - If invFiltFlag =0 (avg) then compute covariance from averaged trials
 - If invFiltFlag =1 (all) then compute covariance from NON-averaged trials
- + Compute Inverse Solution
- + Project data at appropriate time/periods through the inverse Spatial Filters
- + If baseline defined, compute power in baseline period (Used only for plotting)
- + Plot results
- + Scale power results by $10^{(42)}$
- + Save results
- + Go to next contrast

"tmegconnebasic" pipeline

The purpose of this pipeline is to estimate time-resolved dense connectomes in a number of frequency bands at the level of the neuronal sources for the task MEG scans. These time-resolved connectomes are time-locked to task-specific events-of-interest. The pipeline uses the data created by the [tmegpreproc](#) pipeline. This is sensor-level data, which contains the entire set of trials for a given data group, e.g. for Working Memory, the data group TIM contains all the trials locked to the onset of a stimulus image.

For each main condition in the Motor Task and Working Memory tasks (e.g. left hand/right hand/left foot/right foot in the Motor Task) and each frequency band (see below), a separate file is generated.

The results are saved in CIFTI-2 format. This format was chosen so that results can be directly read by [Connectome Workbench](#) software. Workbench reads and enables viewing of results from other modalities (structural-MR, fMRI, dMRI) as overlays on brain surfaces that may be easily compared.

The dense connectomes are derived by projecting the sensor level cross spectral density in source space through an inverse solution based on a Dynamic Imaging of Coherent Sources (DICS) beamformer. This inverse solution is derived based on a single shell forward model. In this forward model, the Leadfield vectors have been normalized by their quadrance (squared euclidean norm) in order to remove the inherent bias towards the center of the brain by beamformer solutions.

For all conditions, the beamformer solution is derived from the average cross-spectral density matrix at the sensor level. A different inverse solution is computed for each Frequency Band. Currently, the main conditions for which source level power series have been computed are:

Motor Task

Conditions: Left Hand, Right Hand, Left Foot, Right Foot, Fixation

For the Motor task, the Time windows of time-resolved connectomes (fixation is not time-resolved, so it has only one connectome per frequency) are:

```
[ -0.3    0
      0    0.3
      0.3  0.6
      0.6  0.9] sec
```

These windows were used in both the TFLA (trials locked on Flashing pacing stimulus) and TEMG (trials locked on the onset of EMG) data groups.

Working Memory Task

Conditions: 0-Back, 2-Back , Faces, Tools, Fixation

For the Working Memory task the Time windows of time-resolved connectomes (fixation is not time-resolved, so it has only one connectome per frequency) are:

for data group TIM (trials locked on image onset):

```
[ 0 0.5
    0.5 1
    1 1.5
    1.5 2] sec
```

for datagroup TRESP (trials locked on response onset):

```
[ -0.25 0.25
      0.25 0.75
      0.75 1.25
      1.25 1.75] sec
```

For the Fixation blocks, the dense connectomes are not time-resolved but rather a single connectome per frequency band. This is because there are no events in the Fixation blocks. This connectome is experimentally only computed in order to serve as a contrast for the rest of the conditions.

The output source level data has not been referenced with respect to any baseline at source level.

The 'tmegconnebasic' pipeline has been designed to concatenate across scans from a given paradigm, e.g. "10-Motort" and "11-Motort", computing the average over all selected trials in both scans. The results are saved using a SCANID that omits the individual run number, e.g. "Motort". In the case that one of the scans is not available, the pipeline will operate on a single scan.

This pipeline currently creates connectomes based on three different connectivity metrics, namely Imaginary Coherence, Coherence and Band Limited Power Correlation. The Coherence and Band Limited Power Correlation connectomes suffer from field spread and should be used for comparisons between time windows within a condition or between conditions. The Imaginary Coherence connectomes do NOT suffer from field spread and can be directly interpreted.

The DICS based inverse solution is performed including data from *all time windows* in a single condition and frequency band. This is done so that differences in connectivity between time windows are not biased due to different spatial filters being used.

Frequency bands used at right.

| Frequency band | Frequency range (Hz) |
|----------------|----------------------|
| delta | 1-4 |
| theta | 4-8 |
| alpha | 8-15 |
| beta low | 15-26 |
| beta high | 26-35 |
| gamma low | 35-50 |
| gamma mid | 50-76 |
| gamma high | 76-120 |

Time-frequency analysis has been performed using the following settings:

| TFR analysis settings | | |
|-----------------------|-------------|---|
| Time Window | 3./f | For each different frequency a different time window length was used. The length of the time window is set to 3 periods of the given frequency. For low frequencies this time window length becomes long so a threshold has been set to 500 msec as the longest window. |
| TFR method | 'mtmconvol' | Single taper time-frequency transformation based on |

multiplication in the frequency domain.

| | | |
|----------------|-----------|----------------------|
| TFR taper type | 'hanning' | Hanning window taper |
|----------------|-----------|----------------------|

Main Steps

Load the entire list of contrasts for a given experiment and selects the ones that are set for the 'tmeqconnebasic' pipeline. Call core function `hcp_tmeqconnebasic_contrasts.m`.

- 2a. Arrange the list of contrasts so that those from the same data group are stacked together.
- 2b. The clean data trials are loaded for the current data group.
- 2c. For each contrast in this data group, select the trials for the corresponding conditions.
- 2d. Compute inverse solution with DICS based on cross spectral density across all time windows in given contrast.
- 2e. Compute fourier coefficients for given Frequency Band and given time-window in each trial.
- 2f. Compute Cross spectral density for given time-window.
- 2g. Project Cross spectral Density for given time window (for Power Correlation only the fourier coefficients are projected).
- 2h. Compute connectome for given Connectivity metric, given frequency band and given time-window.
- 2i. Proceed to the next time-window and repeat from 2e.
- 2j. Save data file with bandlimited time-resolved dense connectome.
- 2k. Proceed to same process for next frequency band from 2d.
- 2l. Proceed to the next contrast from the same data group from 2c.
- 2m. If all contrasts from data group have been processed, then load data from the next data group and perform steps 2b to 2l for this new data group.

Chapter 5: Behavioral and Individual Subject measures details

There are important individual differences in patterns of connectivity and there is evidence that this variability is associated with alterations in important cognitive and behavioral variables that affect real world function. Thus, the HCP is collecting behavioral measures of a range of motor, sensory, cognitive and emotional processes that will delineate a core set of functions relevant to understanding the relationship between brain connectivity and human behavior.

The HCP collects many behavioral measures developed for the [NIH Toolbox](#) Assessment of Neurological and Behavioral function and several additional measures to assess domains not covered by the NIH Toolbox. Our goal in choosing assessments was to identify and utilize a reliable and well-validated battery of measures that assess a wide range of human functions and behaviors in a reasonable amount of time (3-4 hours total, to satisfy subject burden considerations). For more information on the rationale behind the development of the behavioral and individual differences batteries used in HCP see [Barch et al. 2013](#).

Many of the behavioral and demographic measures are included in the data that is released under the [Open Access Data Use Terms](#). Due to their sensitive and possibly identifying nature, several of the measures are being released only to those who have been approved for [Restricted Data](#). Restricted Data categories are marked with a * below.

Details on the NIH Toolbox behavioral measures included in the HCP behavioral battery can be found at <http://www.nihtoolbox.org> and are individually linked in the table below. Details and references for each behavioral measure are described below by domain.

HCP battery of behavioral and individual difference measures

| Domain | Subdomain (Measure Name) |
|-----------|---|
| Alertness | Cognitive Status (Mini Mental Status Exam) |
| | Sleep (Pittsburgh Sleep Quality Index) |
| Cognition | Episodic Memory (Picture Sequence Memory) |
| | Executive Function/Cognitive Flexibility (Dimensional Change Card Sort) |
| | Executive Function/Inhibition (Flanker Inhibitory Control and Attention Task) |
| | Fluid Intelligence (Penn Progressive Matrices) |
| | Language/Reading Decoding (Oral Reading Recognition) |
| | Language/Vocabulary Comprehension (Picture Vocabulary) |
| | Processing Speed (Pattern Comparison Processing Speed) |
| | Self-regulation/Impulsivity (Delay Discounting) |



| Domain | Subdomain (Measure Name) |
|--------------------------------------|--|
| | Spatial Orientation (Variable Short Penn Line Orientation Test) |
| | Sustained Attention (Short Penn Continuous Performance Test) |
| | Verbal Episodic Memory (Penn Word Memory Test) |
| | Working Memory (List Sorting) |
| Emotion | Emotion Recognition (Penn Emotion Recognition Test) |
| | Negative Affect (Sadness, Fear, Anger) (self-report) |
| | Psychological Well-being (Positive Affect, Life Satisfaction, Meaning and Purpose) (self-report) |
| | Social Relationships (Social Support, Companionship, Social Distress, Positive Social Development) (self-report) |
| | Stress and Self-Efficacy (Perceived Stress, Self-Efficacy) (self-report) |
| Motor | Endurance (2 minute walk test) |
| | Locomotion (4-meter walk test) |
| | Dexterity (9-hole Pegboard) |
| | Strength (Grip Strength Dynamometry) |
| Personality | Five Factor Model (NEO-FFI) |
| Health and Family History | Height, Weight, BMI* |
| | Hematocrit Levels* |
| | Blood Pressure* |
| | Thyroid Function (Thyroid Stimulating Hormone Levels)* |
| | Glucose levels (Hemoglobin A1c)* |
| | Endocrine disorders and age of onset (self-report)* |
| | Menstrual Cycle Information* |
| | Parental History of Psychiatric or Neurologic illnesses* |
| MEG Subjects | Alpha and Beta Peak Values (MEG Subjects only) |
| Psychiatric and Life Function | Life Function (Achenbach Adult Self-Report, Syndrome Scales and DSM-Oriented Scale)* |
| | Psychiatric Clinical Symptoms (Semi-Structured Assessment for the Genetics of Alcoholism-SSAGA)* |
| | Substance Use (breathalyzer, self-report of current alcohol and tobacco use)* |
| | Urine Drug Screen* |
| Sensory | Audition (Words in Noise) |
| | Olfaction (Odor Identification Test) |

| Domain | Subdomain (Measure Name) |
|------------------------|--|
| | Pain (Pain Intensity and Interference Surveys) (self-report) |
| | Taste (Regional Taste Intensity Test) |
| | Color Vision (Farnsworth Test)* |
| | Visual Acuity (Electronic Visual Acuity System)* |
| | Contrast Sensitivity (Mars Contrast Sensitivity) |
| Substance Abuse | Breathalyzer and Drug Tests* |
| | Alcohol and Tobacco 7-day Retrospectives* |
| | Alcohol, Tobacco, and Marijuana Use and Dependence (SSAGA self-report)* |
| | Illicit Drug Use (SSAGA self-report)* |

* Access to this restricted data is granted to approved users only.

The [HCP Data Dictionary wiki](#) provides definitions and descriptions of the variable short names for each of the measures.

Results of open access behavioral measures for all tested Q1-Q6 subjects are available as a comma separated variable (.csv) file on the ConnectomeDB splash page (accessible after log in at <https://db.humanconnectome.org/>). If a user has Restricted access, links to a .csv download of all Restricted data for tested Q1-Q6 subjects will appear on the ConnectomeDB splash page and Dashboard. The behavioral measures are also available per subject and can be filtered upon in the ConnectomeDB HCP Dashboard (accessible by clicking the Explore buttons from the ConnectomeDB splash page, see [How to explore HCP data in ConnectomeDB](#)).

FreeSurfer measures

In addition to the behavioral measures, individual difference measures for physical brain features (morphometry statistics) that are automatically generated by the [FreeSurfer processing pipeline](#) are included as a searchable domain in ConnectomeDB. References to the Neurolex definitions for these statistics are included in the ConnectomeDB data dictionary.

Descriptions for HCP Behavioral and Demographic Measures

Subject Information

Demographics

Each HCP participant is given a participant identification number for tracking and asked a number of demographic questions including gender, age, twin status (including self-reported zygosity), race, ethnicity, educational level, household income, and relationship status. Gender and five year-age range are available as Open Access data. Due to the possible identification of

participants when several data types are known and used in combination, the other demographic data are available as restricted data.

Zygosity, Genotyping, Family, and Parent IDs

All of the following family structure measures are available as restricted data.

HCP recruited twins and their non-twin siblings with assistance from the Missouri Family Registry, which tracks birth record data for twins born in Missouri. Each HCP participant was asked in the screening call to self report or confirm information from the registry database on their twin status and related family members. In previous releases, these self reports provided the data in the measures Zygosity (NotTwin, NotMZ [equivalent to DZ], or MZ), Mother_ID, and Father_ID.

With the S1200 release, we now have genotyping data derived from blood or saliva samples for 1142 of the released subjects (at least for one of the 2 gene chips used, custom Illumina Mega Chip + ImmunoArray + PsychArray and the Neuro Consortium chip). The subject genotyping data allowed us to verify the self reports of shared parents among siblings and zygosity for twins (note: in order for a sibling or twin zygosity relationship to be verified, both siblings/twins must have genotyping data available). For subjects with genotyping data, we used the pattern of SNP markers selected at 2-3Mbp intervals along the autosomal chromosomes to graph the sibling relationships in a Graphical Representation of Relationships (GRR) between subjects (full or half sibling, MZ or DZ for twins) according to the degree to which their SNP markers matched (percentage alleles shared). We updated Mother_ID and/or Father_ID if the mother, father, or both parents were suspected to differ among siblings.

It should be noted that the GRR analysis does involve some interpretation because SNP heritability falls in ranges among full siblings (includes DZ twins) and half siblings. There were also a few cases where the genetic samples appeared to be switched or cross contaminated with other samples, for which we did our best to correct.

We were conservative in making the call that subjects shared parents (were full siblings), meaning that if a sibling relationship was an outlier in the percentage alleles shared in what was expected with two shared parents, the Mother_ID and/or Father_ID were updated to new "999xx" numbers.

In summary, for some subjects, the Mother_ID or Father_ID has changed from what they were in previous releases based on the genotyping data. Mother_IDs and Father_IDs that have been changed based on genotyping have a 999xx ID number to facilitate filtering for affected subjects.

We have updated the restricted family structure measures as follows:

HasGT (true/false) indicates if genotyping data from at least for one of the 2 gene chips used: (1) Illumina Mega Chip plus content from the Illumina ImmunoArray and PsychArray; (2) Illumina Neuro Consortium Array. 7 subjects only have data for chip (1), all others with HasGT=TRUE have data for both chips (1) and (2). Family structure (i.e., Mother_ID and Father_ID) is genetically verified for pairs of siblings (including twins and half-siblings) for whom HasGT=TRUE.

ZygositySR (NotTwin, NotMZ, and MZ) is the original self-reported zygosity. Note: A handful of subjects who do not have a value for this measure (blank) self reported as twins, but did not self report their twin zygosity.

ZygosityGT (MZ, DZ) is the twin zygosity verified by genotyping. Both twins must have genotyping data for a subject (and their twin) to have a ZygosityGT value (if genotyping is not available for either of a twin pair, no values are given for ZygosityGT). Non-twin subjects also do not have a value for ZygosityGT. Note that some subjects self-reported as dizygotic twins (ZygositySR='NotMZ') but genotyping established that they were monozygotic twins (thence ZygosityGT='MZ' for those subjects). ZygosityGT should be given precedence over ZygositySR.

Mother_ID and **Father_ID** are now verified by genotyping between siblings for whom HasGT=TRUE, and, if not, are based on self report. Prior to the S1200 release, Mother_ID and Father_ID were based strictly on self-report. For some subjects, Mother_ID or Father_ID have changed from what they were in previous releases based on the genotyping data. Mother_IDs and Father_IDs that have been changed in the S1200 release based on genotyping have a 999xx ID number to facilitate filtering for affected subjects.

Family_ID is a listing of all Mother_IDs and Father_IDs (linked by underscores) associated with subjects who share at least one parent. Family_ID in ConnectomeDB was created using the [CreateKinshipMatrix.R script](#) available on GitHub, courtesy of Thomas Nichols.

Note: Family_ID in ConnectomeDB is ExtFamID in the output of the CreateKinshipMatrix.R script.

Family_ID allows one to find families/sibships easily, even those with one or more half siblings (families with multiple mothers or fathers, indicating the presence of half siblings, will have long Family_IDs). Family_ID does not indicate that siblings grew up in the same household. HCP subjects were not asked about their rearing environment and we do not have a measure to indicate a shared household/upbringing environment.

Because Family_ID is based on Mother_IDs and Father_IDs, which are based on genotyping or self report (if genotyping is not available), Family_ID should not be interpreted to be fully genetically verified unless all siblings have genotyping data (HasGT=TRUE).

Users who are interested in incorporating family structure into their analyses should be aware that a substantial number (36 twin pairs) of twins who self-reported to be dizygotic (DZ) twins (ZygositySR measure) were found by the genotyping (ZygosityGT measure) to be monozygotic (MZ) twins. Other twin studies have found that it is common for some MZ twins to believe they are DZ twins based on what hospital staff told parents at child birth without genetic verification. An overview of the families included in the final S1200 HCP cohort is detailed in [S1200 HCP Subject Family Structure](#).

Handedness

Handedness is assessed using the Edinburgh Handedness questionnaire (Oldfield 1971). Due to the possible identification of participants when several data types are known and used in combination, handedness is available as restricted data.

Study Completion and MR Sessions

Study Completion

For each subject, we record the data completeness for each study imaging session and measurement to enable the ability to search for complete datasets for analysis. “Complete” = TRUE if a session of that MR modality exists. Percent Complete measures for each MR modality allow one to filter for subjects with 100% of the data to be acquired in that session per the HCP protocol.

Session Information

Detailed information on each of the magnetic resonance scans collected on each participant can be viewed and searched and ConnectomeDB.

Alertness measures

Cognitive Status and Sleep

Participants are administered the Mini Mental Status Exam ([Folstein et al., 1975](#)) as a broad measure of cognitive status (participants are excluded if they score below a 27) ([Crum et al., 1993](#)) and the Pittsburgh Sleep Questionnaire (PSQI, [Buysse et al., 1989](#)) as a measure of sleep quality. Summary scores, PSQI Component scores and individual question responses are reported.

Cognition measures

Episodic Memory

Non-verbal episodic memory is assessed using the [NIH Toolbox Picture Sequence Memory Test](#).

Executive Function

Executive function is measured by two tests in the NIH Toolbox battery. Cognitive flexibility is assessed using the [NIH Toolbox Dimensional Change Card Sort Test \(DCCS\)](#) and attention and inhibitory control are assessed using the [NIH Toolbox Flanker Inhibitory Control and Attention Test](#).

Fluid Intelligence

Although the Toolbox contains measures of crystallized IQ (e.g., vocabulary acquisition), an aspect of IQ strongly influenced by educational opportunities, and measures of executive function (which are both theoretically and empirically related to fluid intelligence), it does not contain a specific measure of fluid intelligence. This construct is strongly linked to specific functional outcomes and to variations in neuronal structure and function in humans ([Duncan et al. 2000](#); [Duncan 2003](#); Duncan 2005). The most commonly used measure of fluid intelligence is Raven's Progressive Matrices ([Prabhakaran et al. 1997](#); [Christoff et al. 2001](#); [Gray et al. 2003](#); Conway et al. 2005; Gray et al. 2005; [Wendelken et al. 2008](#)). We use Form A of an abbreviated version of the Raven's developed by Gur and colleagues ([Bilker et al. 2012](#)). Participants are presented with patterns made up of 2x2, 3x3 or 1x5 arrangements of squares, with one of the squares missing. The participant must pick one of five response choices that best fits the missing square on the pattern. The task has 24 items and 3 bonus items, arranged in order of increasing difficulty. However, the task discontinues if the participant makes 5 incorrect responses in a row.

Language

Language skills are assessed using the [NIH Toolbox Oral Reading Recognition Test](#) and the [NIH Toolbox Picture Vocabulary Test](#).

Processing Speed

Speed of participants mental processing is measured using the [NIH Toolbox Pattern Comparison Processing Speed Test](#).

Self-Regulation/Impulsivity (Delay Discounting)

Delay discounting describes the undervaluing of rewards that are delayed in time. It is illustrated by the fact that humans (and other animals) will often choose a smaller immediate reward over

an objectively larger, but delayed reward. We use a version of the discounting task that identifies ‘indifference points’ at which a person is equally likely to choose a smaller reward (e.g., \$100) sooner versus a larger reward later (e.g., \$200 in 3 years). Based on the work of Green and Myerson (Estle *et al.* 2006; [Green et al. 2007](#)), we use an adjusting-amount approach, in which delays are fixed and reward amounts are adjusted on a trial-by-trial basis based on participants’ choices, to rapidly hone in on indifference points. This approach has been repeatedly validated to provide reliable estimates of delay discounting (Estle *et al.* 2006). As a summary measure, we use an area-under-the-curve discounting measure (AUC) that provides a valid and reliable index of how steeply an individual discounts delayed rewards ([Myerson et al. 2001](#)).

Detailed Description of Delay Discounting Measure

In this task, participants are presented with two choices on each trial – a smaller amount “today” or a larger amount at a later point in time. Participants make choices at each of 6 delays (1 month, 6 months, 1 year, 3 years, 5 years and 10 years) and for two delayed amounts (\$200 and \$40,000). For each combination of delay and amount of delayed reward (e.g., \$200 in 1 month or \$40,000 in 6 months), participants make 5 choices, and the value that would have been used for the immediate amount in a 6th choice is taken as the indifference point for that condition. The participants make all five choices for a particular combination of delay and amount before moving on to the next combination of delay and amount. The order is as follows:

Delayed amount of \$200 dollars

- Today versus 6 months
- Today versus 3 years
- Today versus 1 month
- Today versus 5 years
- Today versus 10 years
- Today versus 1 year

Delayed amount of \$40,000 dollars

- Today versus 6 months
- Today versus 3 years
- Today versus 1 month
- Today versus 5 years
- Today versus 10 years
- Today versus 1 year

The first choice at each delay is between the delayed amount (\$200 or \$40,000) and an immediate amount equal to $\frac{1}{2}$ the delayed amount (e.g., \$100 today or \$200 in 1 month, \$20,000 today or \$40,000 in one month). The size of the adjustment after the first choice is

always $\frac{1}{2}$ the amount of the immediate value on the first choice (e.g., a change of \$50 if the first immediate amount is \$100). If the subject chooses the immediate amount, then the immediate amount is reduced on the next choice (e.g., \$50 today versus \$200 in 1 month). If the subject chooses the delayed amount, then the immediate amount is increased (e.g., \$150 today versus \$200 in 1 month). The amount of change on each subsequent choice is $\frac{1}{2}$ the amount of the prior change (e.g., \$25 on the 3rd trial), regardless of whether the subject chooses the immediate or the delayed amount. This procedure rapidly hones in on the amount of immediate gain that is close to the subjective value of the delayed gain.

This design means that for all the choices with \$200 dollars as the delayed amount, the first choice will always be between \$100 today, and \$200 in the specified time period. The second choice will always increment or decrement the immediate value by \$50. The third choice will always increment or decrement the immediate value by \$25. The fourth choice will always increment or decrement the immediate value by \$12.50. The fifth choice will always increment or decrement the immediate value by \$6.25. The “sixth” choice value, which is never presented to the subject, but is entered in the database, is always an increment or decrement of \$3.125 from the immediate value on the 5th choice. Similarly, for all the choices with \$40,000 dollars as the delayed amount, the first choice will always be between \$20,000 today, and \$40,000 in XX time period. The second choice will always increment or decrement the immediate value by \$10,000. The third choice will always increment or decrement the immediate value by \$5,000. The fourth choice will always increment or decrement the immediate value by \$2,500. The fifth choice will always increment or decrement the immediate value by \$1,250. The “sixth” choice value, which is never presented to the subject, but is entered in the database, will always be an increment or decrement of \$625 from the immediate value on the 5th choice.

Thus, for the \$200 amount, we will have 6 Subject Values:

- $SV_{1mo.2}$
- $SV_{6mo.2}$
- $SV_{1yr.2}$
- $SV_{3yr.2}$
- $SV_{5yr.2}$
- $SV_{10yr.2}$

Thus, for the \$40,000 amount, we will have 6 Subject Values:

- $SV_{1mo.40}$
- $SV_{6mo.40}$
- $SV_{1yr.40}$
- $SV_{3yr.40}$
- $SV_{5yr.40}$
- $SV_{10yr.40}$

We compute an Area under the curve measure for each of the two amounts as described below.

$$\begin{aligned} \text{Area under the Curve for \$200} = & ((1+SV_{1mo.2})/(120*200)) + ((SV_{1mo.2}+SV_{6mo.2})/(48*200)) + \\ & ((SV_{6mo.2}+SV_{1yr.2})/(40*200)) + ((SV_{1yr.2}+SV_{3yr.2})/(10*200)) + ((SV_{3yr.2}+SV_{5yr.2})/(10*200)) + \\ & ((SV_{5yr.2}+SV_{10yr.2})/(4*200)) \end{aligned}$$

$$\begin{aligned} \text{Area under the Curve for \$40,000} = & ((1+SV_{1mo.4})/(120*40,000)) + \\ & ((SV_{1mo.4}+SV_{6mo.4})/(48*40,000)) + ((SV_{6mo.4}+SV_{1yr.4})/(40*40,000)) + \\ & ((SV_{1yr.4}+SV_{3yr.4})/(10*40,000)) + ((SV_{3yr.4}+SV_{5yr.4})/(10*40,000)) + \\ & ((SV_{5yr.4}+SV_{10yr.4})/(4*40,000)) \end{aligned}$$

Spatial Orientation

The NIH Toolbox does not contain any measures of visual-spatial processing. Thus, we are measuring spatial orientation processing using the Variable Short Penn Line Orientation Test ([Gur et al. 2001a](#); [Gur et al. 2010](#)). Participants are shown two lines with different orientations. They have to rotate one of the lines (a moveable blue one) so that is parallel to the other line (a fixed red line). The rotation of the blue line is accomplished by clicking buttons on the keyboard that rotate the lines either clockwise or counterclockwise. Across trials, the lines vary in their relative location on the screen, though the distance between the centers of the two lines is always the same. The length of the red line is always the same, but the length of the blue line can be either short or long. There are a total of 24 trials.

Sustained Attention

We measure continuous sustained attention using the Short Penn Continuous Performance Test (CPT, Number/Letter Version) ([Gur et al. 2001a](#); [Gur et al. 2001b](#); [Gur et al. 2010](#)). Participants see vertical and horizontal red lines flash on the computer screen. In one block, they must press the spacebar when the lines form a number and in the other block they press the spacebar when the lines form a letter. The lines are displayed for 300 ms followed by a 700 ms ITI. Each block contains 90 stimuli and lasts for 1.5 minutes.

Verbal Episodic Memory

The NIH Toolbox contains a measure of non-verbal episodic memory. Thus, we are assessing verbal episodic memory using Form A of the Penn Word Memory Test ([Gur et al. 2001a](#); [Gur et al. 2010](#)). Participants are shown 20 words and asked to remember them for a subsequent memory test. They are then shown 40 words (the 20 previously presented words and 20 new words matched on memory related characteristics). They decide whether they have seen the word previously by choosing among “definitely yes,” “probably yes,” “probably no,” and “definitely no.”

Working Memory

We measure working memory using the [NIH Toolbox List Sorting Working Memory Test](#).

Emotion measures

Emotion Recognition

The NIH Toolbox contains only self-report measures of emotional function. Thus, in order to obtain a behavioral measure of emotion processing, we are using the Penn Emotion Recognition Test ([Gur et al. 2001a](#); [Gur et al. 2010](#)). Participants are presented with 40 faces, one at a time. They are asked to choose what emotion the face is showing from five choices: Happy, Sad, Angry, Scared and No Feeling. Half of the faces are males and half are females. There are 8 faces each that have a happy, sad, angry, scared or no feeling expression.

Negative Affect

Self-reported unpleasant feelings or emotions including anger, fear and sadness are accessed by the [NIH Toolbox Negative Affect](#) surveys.

Psychological Well-being

Measures of general life satisfaction, meaning and purpose, and positive affect are self-reported through the [NIH Toolbox Psychological Well-Being](#) surveys.

Social Relationships

The structure, extent, and quality of each participant's social relationships including self-reported social support, companionship, social distress, and positive social development are accessed by [NIH Toolbox Social Relationships](#) survey measures.

Stress and Self-Efficacy

Self-reported perception of each participant's psychological stress and ability to cope with difficult situations are accessed by the [NIH Toolbox Stress and Self-Efficacy](#) surveys.

Motor measures

Endurance

Sub-maximal cardiovascular endurance is measured using the [NIH Toolbox 2-Minute Walk Endurance Test](#).

Locomotion

Gait speed as an indicator of the health of many physiological systems is measured by the [NIH Toolbox 4- Meter Walk Gait Speed Test](#).

Dexterity

Manual dexterity is measured using the [NIH Toolbox 9-Hole Pegboard Dexterity Test](#).

Strength

Full force grip strength is measured by the [NIH Toolbox Grip Strength Test](#).

Personality measures

Five Factor Model

There is consensus that a five factor model captures the major facets of human personality across cultures ([Heine and Buchtel 2009](#)): a) neuroticism; b) extroversion/introversion; c) agreeableness; d) openness; and e) conscientiousness ([Goldberg 1993](#); McCrae and Costa 2008). We are administering the 60 item version of the Costa and McCrae Neuroticism/Extroversion/Openness Five Factor Inventory (NEO-FFI), which has shown excellent reliability and validity (McCrae and Costa 2004). This measure was available as part of the Penn Computerized Cognitive Battery ([Gur et al. 2001a](#); [Gur et al. 2010](#)).

Health and Family History

All Health and Family History measures are Restricted data. The following are part of the current release:

Height, Weight, and Body Mass Index (BMI)

Self-reported height and weight, and BMI calculated from height/weight are recorded for each participant.

Hematocrit and Glucose Levels

We take a blood sample of every participant for various tests. We measure hematocrit levels (2 samples) to assess the volume percentage of red blood cells in the blood.

We also measure Hemoglobin A1c as a measure of glucose levels over time.

Blood Pressure

Each participant's blood pressure is taken and recorded at their visit.

Thyroid Function

Thyroid Stimulating Hormone level in the blood is measured as an endocrine measure.

Endocrine disorders

Participants are asked to self-report if they have ever been diagnosed with an endocrine disorder and, if so, what age they were at the onset of the disorder.

Menstrual Cycle and Hormonal Status

Female participants are asked questions about their menstrual cycle during the intake interview at their first in person session.

Parental Psychiatric or Neurological Disorders

Participants are given a brief assessment of parental history of psychiatric and neurological disorders (yes/no for schizophrenia or psychosis, depression, bipolar, anxiety that needed treatment, drug or alcohol problems, Alzheimer's Disease or dementia, Parkinson's disease, or Tourette's syndrome).

MEG Subjects

Alpha and Beta Peak Values

See [Determining Alpha and Beta Peak Frequencies for Individual Subjects](#) above. These electrophysiological measures are collected for MEG subjects only and are open access.

Psychiatric and Life Function

All Psychiatric, Substance Abuse, and Life Function data is being released to Restricted Access approved users only. Psychiatric clinical symptoms and parental psychiatric illness data will be released in a future data release.

Life Function

The NIH toolbox contains self-report measures of a number of important domains of experience, including positive and negative affect, stress, anxiety, depression and social support. To obtain additional self-report information on an even broader variety of domains, we also administer the Achenbach Adult Self-Report (ASR) for Ages 18-59 (Achenbach 2009). Specifically, we

administer the 123 items from Section VIII. These can be used to generate the ASR Syndrome Scales and the ASR Diagnostic and Statistical Manual (DSM)-Oriented Scales.

Psychiatric History

As part of the screening and assessment process, all participants are given a comprehensive assessment of psychiatric and substance use history over the phone, using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) ([Bucholz *et al.*, 1994](#)). The SSAGA is a well-validated diagnostic instrument used in numerous previous large-scale studies ([Bucholz *et al.*, 1994](#); [Hesselbrock *et al.*, 1999](#)). It assesses a range of diagnostic categories (substance, mood, anxiety, eating disorders and adult ADHD), as well as antisocial personality disorder, using both DSM-IV criteria and either RDC or ICD-criteria, and provides information about both current and lifetime experiences. This instrument also contains the Fagerstrom Test for Nicotine Dependence (Heatherton *et al.*, 1991; Kozlowski *et al.*, 1994).

Substance Use

Drug Screen

Participants are also given a breathalyzer and a urine drug screen (cocaine, THC, opiates, amphetamine, methamphetamine, oxycontin) on each day of testing. These drug screens were not used as an exclusion criterion, but rather for characterization.

Alcohol and Tobacco 7-Day Retrospective

In addition, on the last day of testing, participants fill out a seven day retrospective report of alcohol and tobacco use.

Alcohol, Tobacco, and Marijuana Use and Dependence

As part of the initial phone interview participants respond to a survey of alcohol, tobacco, and marijuana use and dependence as part of the SSAGA.

Illicit Drug Use

As part of the initial phone interview participants respond to a survey of non-marijuana illicit drug (cocaine, hallucinogens, opiates, sedatives, stimulants, other) use as part of the SSAGA.

Sensory measures

Audition

Hearing, or audition, is assessed using the [NIH Toolbox Words-in-Noise Test](#).

Color Vision and Contrast Sensitivity

The NIH Toolbox does not measure color vision or contrast sensitivity. We are assessing color vision using the Farnsworth Test, a valid and reliable measure that provides more quantitative information than the commonly used Ishihara Test ([Cole et al. 2007](#)). In this task, participants order 15 colored blobs as a function of what they think are the closest matching colors. Based on the results, participants are classified as having Normal color vision, Protan (reduced sensitivity to red light), Deutan (reduced sensitivity to green light) or Tritan (reduced sensitivity to blue light) color vision problems. Because color vision deficits are relatively rare and can be identifying, results from this assessment are Restricted Access.

We are assessing contrast sensitivity using the Mars Contrast Sensitivity Test ([Arditi et al. 2005](#)), a brief, valid and reliable measure that improves upon the traditional Pelli-Robson measure ([Dougherty et al. 2005](#); [Haymes et al. 2006](#); [Thayaparan et al. 2007](#)).

Olfaction

Smell, or olfaction, is assessed using the [NIH Toolbox Odor Identification Test](#).

Pain

Participants' self-reported experience of pain is measured by the [NIH Toolbox Pain Interference and Pain Intensity Surveys](#).

Taste

Taste perception is measured by the [NIH Toolbox Regional Taste Intensity Test](#).

Visual Acuity

To assess visual acuity, we are using the Electronic Visual Acuity (EVA) system provided by the Jaeb Center for Health Research and running the Electronic Early Treatment of Diabetic Retinopathy (E-ETDR) protocol ([Beck et al., 2003](#); [Moke et al., 2001](#)). Subjects are scored by the EVA Numerator and Denominator, which are approximations of the Snellen ratio of visual acuity. The EVA Numerator is the virtual distance in feet (always 20) from the EVA system display that the subject's visual acuity was tested. Our test is conducted with the EVA Model 10-WIN system at 3 meters (~10 feet from the subject), but is calibrated to reflect values equivalent to visual acuity testing at 20 feet. Letters are displayed and the subject (with corrected vision, if applicable) is asked to read the letters. The subject's ability to read progressively smaller letters is compared with the ability of a person with normal (20/20 Snellen ratio) vision to read letters at 20 feet. The EVA Denominator is the distance in feet that a person with normal (20/20 Snellen ratio) vision can read letters as well as the subject (with corrected vision, if applicable) can read letters at a virtual distance of 20 feet (value of the EVA numerator). An EVA Denominator value

less than 20 means the subject has better than normal vision, a value greater than 20 means the subject has poorer than normal vision. Because visual acuity can be identifying, results from this assessment are Restricted Access.

References

- Achenbach TM (2009). The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications. Burlington, VT: University of Vermont Research Center for Children, Youth and Families.
- Andersson JLR, Skare S, Ashburner J (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20(2): 870-888.
- Arditi A (2005). Improving the design of the letter contrast sensitivity test. *Invest Ophthalmol Vis Sci* 46(6): 2225-2229.
- Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, Glasser MF, Curtiss S, Dixit S, Feldt C, Nolan D, Bryant E, Hartley T, Footer O, Bjork JM, Poldrack R, Smith SM, Johansen-Berg H, Snyder AZ, Van Essen DC (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage* 80: 169-189.
- Beck RW, Moke PS, Turpin AH, Ferris FL, SanGiovanni JP, Johnson CA, Birch EE, Chandler DL, Cox TA, Blair RC, Kraker RT (2003). A computerized method of visual acuity testing: adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol* 135: 194-205.
- Beckmann CF, Smith SM (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging*. 23(2): 137-152.
- Bilker WB, Hansen JA, Brensinger CM, Richard J, Gur RE, Gur RC (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment* 19(3): 354-369.
- Binder JR, Gross WL, Allendorfer JB, Bonilha L, Chapin J, Edwards JC, Grabowski TJ, Langfitt JT, Loring DW, Lowe MJ, Koenig K, Morgan PS, Ojemann JG, Rorden C, Szaflarski JP, Tivarus ME, Weaver KE (2011). Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study. *NeuroImage* 54(2): 1465-1475.
- Bizzi A, Blasi V, Falini A, Ferroli P, Cadioli M, Danesi U, Aquino D, Marras C, Caldiroli D, Broggi G (2008). Presurgical functional MR imaging of language and motor functions: validation with intraoperative electrocortical mapping. *Radiology* 248(2): 579-589.
- Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger Jr. JI, Reich T, Schmidt I, Schuckit MA (1994). A new, semi-structured psychiatric interview for use in

- [genetic linkage studies: a report on the reliability of the SSAGA](#). *J Stud Alcohol* 55: 149-158.
- Buckner RL, Krienen FM, Castellanos A, Diaz JC, Yeo BT (2011). [The organization of the human cerebellum estimated by intrinsic functional connectivity](#). *J Neurophysiol* 106(5): 2322-2345.
- Buyse DJ, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ (1989). [The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research](#). *Psychiatry Res* 28: 193-213.
- Caceres A, Hall DL, Zelaya FO, Williams SC, Mehta MA (2009). [Measuring fMRI reliability with the intra-class correlation coefficient](#). *Neuroimage* 45(3): 758-768.
- Castelli F, Frith C, Happé F, Frith U (2002). [Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes](#). *Brain* 125(Pt 8): 1839-1849.
- Castelli F, Happé F, Frith U, Frith C (2000). [Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns](#). *NeuroImage* 12(3): 314-325.
- Chen X, Nichols TE, *et al.* (2014). APACE: Accelerated Permutation inference for the ACE model. *OHBM 2014*, Poster 3410.
- Christoff K, Prabhakaran V, Dorfman J, Zhao Z, Kroger JK, Holyoak KJ, Gabrieli JD (2001). [Rostrolateral prefrontal cortex involvement in relational integration during reasoning](#). *NeuroImage* 14(5): 1136-1149.
- Cole BL (2007). [Assessment of inherited colour vision defects in clinical practice](#). *Clin Exp Optom* 90(3): 157-175.
- Conway AR, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW (2005). Working memory span tasks: A methodological review and user's guide. *Psychon Bull Rev* 12(5): 769-786.
- Crum RM, Anthony JC, Bassett SS, Folstein MF (1993). [Population-based norms for the Mini-Mental State Examination by age and educational level](#). *JAMA* 269: 2386-2391.
- Dale AM, Fischl B, Sereno MI (1999). [Cortical surface-based analysis. I. Segmentation and surface reconstruction](#). *NeuroImage* 9(2): 179-194.
- Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000). [Tracking the hemodynamic responses to reward and punishment in the striatum](#). *J Neurophysiol* 84: 3072-3077.

- Dougherty BE, Flom RE, Bullimore MA (2005). [An evaluation of the Mars Letter Contrast Sensitivity Test](#). *Optom Vis Sci* 82(11): 970-975.
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001). [A cortical area selective for visual processing of the human body](#). *Science* 293(5539): 2470-2473.
- Drobyshevsky A, Baumann SB, Schneider W (2006). [A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function](#). *NeuroImage* 31(2): 732-744.
- Duncan J (2003). [Intelligence tests predict brain response to demanding task events](#). *Nat Neurosci* 6(3): 207-208.
- Duncan J (2005). Frontal lobe function and general intelligence: why it matters. *Cortex* 41(2): 215-217.
- Duncan J, Seitz RJ, Kolodny J, Bor D, Herzog H, Ahmed A, Newell FN, Emslie H (2000). [A neural basis for general intelligence](#). *Science* 289(5478): 457-460.
- Estle SJ, Green L, Myerson J, Holt DD (2006). Differential effects of amount on temporal and probability discounting of gains and losses. *Mem Cognit* 34(4): 914-928.
- Ewald A, Marzetti L, Zappasodi F, Meinecke FC, Nolte G (2012). [Estimating true brain connectivity from EEG/MEG data invariant to linear and static transformations in sensor space](#). *Neuroimage*, 60(1):476-88.
- Filippini N, MacIntosh BJ, Hough MG, Goodwin GM, Frisoni GB, Smith SM, Matthews PM, Beckmann CF, Mackay CE (2009). [Distinct patterns of brain activity in young carriers of the APOE-e4 allele](#). *Proc Natl Acad Sci USA*, 106:7209-7214.
- Folstein MF, Folstein SE, McHugh PR (1975). [Mini-mental state: a practical method for grading the cognitive state of patients for the clinician](#). *J Psychiatr Res* 12,189-198.
- Forbes EE, Hariri AR, Martin SL, Silk JS, Moyles DL, Fisher PM, Brown SM, Ryan ND, Birmaher B, Axelson DA, Dahl RE (2009). [Altered striatal activation predicting real-world positive affect in adolescent major depressive disorder](#). *Am J Psychiatry* 166(1): 64-73.
- Fox CJ, Iaria G, Barton JJ (2009). [Defining the face processing network: optimization of the functional localizer in fMRI](#). *Hum Brain Mapp* 30(5): 1637-1651.
- Glasser MF, Van Essen DC (2011). [Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI](#). *J Neurosci* 31:11597-11616.
- Glasser MF, Sotiropoulos SN, Wilson JA, Coalson T, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Van Essen DC, Jenkinson M (2013). [The minimal preprocessing pipelines for the Human Connectome Project](#). *NeuroImage* 80:105-124.

- Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC (2016). [A multi-modal parcellation of human cerebral cortex](#). *Nature*. Aug 11;536(7615):171-8.
- Glasser MF, Smith SM, Marcus DS, Andersson JL, Auerbach EJ, Behrens TE, Coalson TS, Harms MP, Jenkinson M, Moeller S, Robinson EC, Sotiropoulos SN, Xu J, Yacoub E, Ugurbil K, Van Essen DC (2016). [The Human Connectome Project's neuroimaging approach](#) *Nature Neurosci* 19, 1175–1187.
- Glover GH (1999). [Deconvolution of impulse response in event-related BOLD fMRI](#). *NeuroImage* 9(4):416-429.
- Goldberg LR (1993). [The structure of phenotypic personality traits](#). *Am Psychol* 48: 26-34.
- Gray JR, Chabris CF, Braver TS (2003). [Neural mechanisms of general fluid intelligence](#). *Nat Neurosci* 6(3): 316-322.
- Gray JR, Burgess GC, Schaefer A, Yarkoni T, Larsen RJ, Braver TS (2005). Affective personality differences in neural processing efficiency confirmed using fMRI. *Cogn Affect Behav Neurosci* 5(2): 182-190.
- Green L, Myerson J, Shah AK, Estle SJ, Holt DD (2007). [Do adjusting-amount and adjusting-delay procedures produce equivalent estimates of subjective value in pigeons?](#) *J Exp Anal Behav* 87(3): 337-347.
- Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, Moeller S, Xu J, Yacoub E, Baselli G, Ugurbil K, Miller KL, Smith SM (2014). [ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging](#). *NeuroImage* Epub ahead of print 2014 Mar 21. doi: 10.1016/j.neuroimage.2014.03.034.
- Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE (2001a). [Computerized neurocognitive scanning: I. Methodology and validation in healthy people](#). *Neuropsychopharmacology* 25(5): 766-776.
- Gur RC, Ragland JD, Moberg PJ, Bilker WB, Kohler C, Siegel SJ, Gur RE (2001b). [Computerized neurocognitive scanning: II. The profile of schizophrenia](#). *Neuropsychopharmacology* 25(5): 777-788.
- Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, Brensinger C, Gur RE (2010). [A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation](#). *J Neurosci Methods* 187(2): 254-262.

- Hariri AR, Brown SM, Williamson DE, Flory JD, de Wit H, Manuck SB (2006). [Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity.](#) *J Neurosci* 26: 13213-13217.
- Hyvärinen A (1999). [Fast and Robust Fixed-Point Algorithms for Independent Component Analysis.](#) *IEEE Trans Neural Networks* 10(3):626-634.
- Haymes SA, Roberts KF, Cruess AF, Nicoleta MT, LeBlanc RP, Ramsey MS, Chauhan BC, Artes PH (2006). [The letter contrast sensitivity test: clinical evaluation of a new design.](#) *Invest Ophthalmol Vis Sci* 47(6): 2739-2745.
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO (1991). The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *Br J Addict* 86: 1119-1127.
- Heine SJ, Buchtel EE (2009). [Personality: the universal and the culturally specific.](#) *Annu Rev Psychol* 60: 369-394.
- Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V (1999). [A validity study of the SSAGA— a comparison with the SCAN.](#) *Addiction* 94: 1361-1370.
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012). [FSL.](#) *NeuroImage* 62: 782-90.
- Jovicich J, Czanner S, Greve D, Haley E, Kowalewski A, Gollub R, Kennedy D, Schmitt F, Brown G, MacFall J, Fischl B, Dale A (2006). [Reliability in Multi-Site Structural MRI Studies: Effects of Gradient Non-linearity Correction on Phantom and Human Data.](#) *NeuroImage* 30(2): 436-43.
- Kozlowski LT, Porter CQ, Orleans CT, Pope MA, Heatherton T (1994). Predicting smoking cessation with self-reported measures of nicotine dependence: FTQ, FTND, and HSI. *Drug Alcohol Depend* 34: 211-216.
- Kung CC, Peissig JJ, Tarr MJ (2007). [Is region-of-interest overlap comparison a reliable measure of category specificity?](#) *J Cogn Neurosci* 19(12): 2019-2034.
- Larson-Prior LJ, Oostenveld R, Della Penna S, Michalareas G, Prior F, Babajani-Feremi, Schoffelen J-M, Marzetti L, de Pasquale F, Di Pompeo F, Stout J, Woolrich M, Luo Q, Bucholz R, Fries P, Pizzella V, Romani GL, Corbetta M, Snyder AZ (2013). [Adding dynamics to the Human Connectome Project with MEG.](#) *NeuroImage* 80: 190-201.
- Lin FH, Witzel T, Hämäläinen MS, Dale AM, Belliveau JW, Stufflebeam SM (2004). [Spectral spatiotemporal imaging of cortical oscillations and interactions in the human brain.](#) *Neuroimage*, 23(2): 582-595.

- Mantini D, Della Penna S, Marzetti L, de Pasquale F, Pizzella V, Corbetta M, Romani GL (2011). [A signal-processing pipeline for magnetoencephalography resting-state networks](#). *Brain Cogn* 1: 49–59.
- Manuck SB, Brown SM, Forbes EE, Hariri AR (2007). [Temporal stability of individual differences in amygdala reactivity](#). *Am J Psychiatry* 164(10): 1613-1614.
- Marcus DS, Harwell J, Olsen T, Hodge M, Glasser MF, Prior F, Jenkinson M, Laumann T, Curtiss SW, Van Essen DC (2011). [Informatics and Data Mining Tools and Strategies for the Human Connectome Project](#). *Front Neuroinform* 5: 4.
- Marcus DS, Harms MP, Snyder AZ, Jenkinson M, Wilson JA, Glasser MF, Barch DM, Archie KA, Burgess GC, Ramaratnam M, Hodge M, Horton W, Herrick R, Olsen T, McKay M, House M, Hileman M, Reid E, Harwell J, Coalson T, Schindler J, Elam JS, Curtiss SW, Van Essen DC (2013). [Human Connectome Project informatics: Quality control, database services, and data visualization](#). *NeuroImage* 80: 144-168.
- Marzetti L, Della Penna S, Snyder AZ, Pizzella V, Nolte G, de Pasquale F, Romani GL, Corbetta M (2013). [Frequency specific interactions of MEG resting state activity within and across brain networks as revealed by the multivariate interaction measure](#). *Neuroimage* 79:172-83.
- May JC, Delgado MR, Dahl RE, Stenger VA, Ryan ND, Fiez JA, Carter C (2004). [Event-related functional magnetic resonance imaging of reward-related brain circuitry in children and adolescents](#). *Biol psychiatry* 55(4): 359-366.
- McCrae RR, Costa PT (2004). A contemplated revision of the NEO Five Factor Inventory. *Pers Individ Diff* 36(3): 587-596.
- McCrae RR, Costa PT (2008). The five factor theory of personality. *Handbook of Personality: Theory and Research*. O. P. John, R. W. Robins and L. A. Pervin. New York, Guilford: 159-181.
- Milchenko M, Marcus D (2013). [Obscuring surface anatomy in volumetric imaging data](#). *Neuroinformatics* 11: 65-75.
- Moke PS, Turpin AH, Beck RW, Holmes JM, Repka MX, Birch EE, Hertle RW, Kraker RT, Miller JM, Johnson CA (2001). [Computerized method of visual acuity testing: adaptation of the amblyopia treatment study visual acuity testing protocol](#). *Am J Ophthalmol* 132: 903-909.
- Morioka T, Yamamoto T, Mizushima A, Tombimatsu S, Shigeto H, Hasuo K, Nishio S, Fujii K, Fukui M (1995). Comparison of magnetoencephalography, functional MRI, and motor evoked potentials in the localization of the sensory-motor cortex. *Neurol Res* 17(5): 361-367.

- Myerson J, Green L, Warusawitharana M (2001). [Area under the curve as a measure of discounting](#). *J Exp Anal Behav* 76(2): 235-243.
- O'Reilly JX, Beckmann CF, Tomassini V, Ramnani N, Johansen-Berg H (2010). [Distinct and overlapping functional zones in the cerebellum defined by resting state functional connectivity](#). *Cerebral Cortex* 20 (4): 953-965.
- Oldfield RC (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1): 97-113.
- Oostenveld R, Fries P, Maris E, and Schoffelen J-M (2011). [FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data](#). *Computational Intelligence and Neuroscience*, vol. 2011, Article ID 156869, 9 pages, 2011. doi:10.1155/2011/156869
- Peelen MV, Downing PE (2005). [Within-subject reproducibility of category-specific visual activation with functional MRI](#). *Hum Brain Mapp* 25(4): 402-408.
- Power JD (2016). [A simple but useful way to assess fMRI scan qualities](#). *Neuroimage* Aug 7. pii: S1053-8119(16)30387-1. doi: 10.1016/j.neuroimage.2016.08.009. [Epub ahead of print]
- Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE (2011). [Functional network organization of the human brain](#). *Neuron* 72(4): 665-78.
- Prabhakaran V, Smith JA, Desmond JE, Glover GH, Gabrieli JD (1997). [Neural substrates of fluid reasoning: an fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test](#). *Cogn Psychol* 33(1): 43-63.
- Rilling JK, Glasser MF, Jbabdi S, Andersson J, Preuss TM (2012). [Continuity, divergence, and the evolution of brain language pathways](#). *Front Evol Neurosci* 3: 11.
- Robinson EC, Jbabdi S, Glasser MF, Andersson J, Burgess GC, Harms MP, Smith SM, Van Essen DC, Jenkinson M (2014). [MSM: A new flexible framework for Multimodal Surface Matching](#). *NeuroImage* Epub 2June, In Press, Accepted Manuscript.
- Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM (2014). [Automatic denoising of functional MRI Data: Combining independent component analysis and hierarchical fusion of classifiers](#). *Neuroimage*, 90: 449-468.
- Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, Eickhoff SB, Hakonarson H, Gur RC, Gur RE, Wolf DH (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(0): 240-256.

- Smith R, Keramatian K, Christoff K (2007). [Localizing the rostrolateral prefrontal cortex at the individual level](#). *Neuroimage* 36(4): 1387-1396.
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004). [Advances in functional and structural MR image analysis and implementation as FSL](#). *NeuroImage*, 23(S1): 208-19.
- Smith SM (2012). [The future of fMRI connectivity](#). *NeuroImage*, 62:1257-1266.
- Smith SM, Andersson J, Auerbach EJ, Beckmann CF, Bijsterbosch J, Douaud G, Duff E, Feinberg DA, Griffanti L, Harms MP, Kelly M, Laumann T, Miller KL, Moeller S, Petersen SE, Power J, Salimi-Khorshidi G, Snyder AZ, Vu A, Woolrich MW, Xu J, Yacoub E, Ügurbil K, Van Essen DC, Glasser MF (2013). [Resting-state fMRI in the Human Connectome Project](#). *NeuroImage* 80: 144-168.
- Smith SM, Vidaurre D, Beckmann CF, Glasser MF, Jenkinson M, Miller KL, Nichols TE, Robinson E, Salimi-Khorshidi G, Woolrich MW, Barch DM, Ügurbil K, Van Essen DC (2013). [Functional connectomics from resting-state fMRI](#). *Trends Cogn Sci*, 17(12):666-682.
- Smith SM, Glasser MF, Robinson E, Salimi-Khorshidi G, Duff E, Van Essen DC, Woolrich MW, Jenkinson M, Beckmann CF (2014). Methods for network modelling from high quality rfMRI data. *OHBM 2014 Annual Meeting*, Poster 1718.
- S.M. Smith, A. Hyvärinen, G. Varoquaux, K.L. Miller, and C.F. Beckmann (2014). [Group-PCA for very large fMRI datasets](#). *NeuroImage*, 101:738-749.
- Sotiropoulos SN, Moeller S, Jbabdi S, Xu J, Andersson JL, Auerbach EJ, Yacoub E, Feinberg D, Setsompop K, Wald LL, Behrens TEJ, Ügurbil K, Lenglet C (2013). [Effects of Image reconstruction on fibre orientation mapping from multichannel diffusion MRI: Reducing the noise floor using SENSE](#). *Mag Res Med* 70(6): 1682-9.
- Sotiropoulos SN, Jbabdi S, Xu J, Andersson JL, Moeller S, Auerbach EJ, Glasser MF, Hernandez M, Sapiro G, Jenkinson M, Feinberg D, Yacoub E, Lenglet C, Van Essen DC, Ügurbil K, Behrens TEJ (2013). [Advances in diffusion MRI acquisition and processing in the Human Connectome Project](#). *NeuroImage* 80: 125-143.
- Taylor JC, Wiggett AJ, Downing PE (2007). [Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas](#). *J Neurophysiol* 98(3): 1626-1633.
- Thayaparan K, Crossland MD, Rubin GS (2007). [Clinical assessment of two new contrast sensitivity charts](#). *Br J Ophthalmol* 91(6): 749-752.

- Tricomi EM, Delgado MR, Fiez JA (2004). [Modulation of caudate activity by action contingency](#). *Neuron* 41(2): 281-292.
- Üğurbil K, Xu J, Auerbach EJ, Moeller S, Vu A, Duarte-Carvajalino JM, Lenglet C, WU X, Schmitter S, Van de Moortele PF, Strupp J, Sapiro G, De Martino F, Wang D, Harel N, Garwood M, Chen L, Feinberg DA, Smith SM, Miller KL, Sotiropoulos SN, Jbabdi S, Andersson JL, Behrens TEJ, Glasser MF, Van Essen DC, Yacoub E (2013). [Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project](#). *NeuroImage* 80: 80-104.
- Van Essen DC (2005). [A Population-Average, Landmark- and Surface-based \(PALS\) atlas of human cerebral cortex](#). *Neuroimage* 28(3):635-62.
- Van Essen DC, Üğurbil K, Auerbach E, Barch D, Behrens TEJ, Bucholz R, Chang A, Chen L, Corbetta M, Curtiss SW, Della Penna S, Feinberg D, Glasser MF, Harel N, Heath AC, Larson-Prior L, Marcus D, Michalareas G, Moeller S, Oostenveld R, Petersen SE, Prior F, Schlaggar BL, Smith SM, Snyder AZ, Xu J, Yacoub E, for the WU-Minn HCP Consortium (2012a). [The Human Connectome Project: A data acquisition perspective](#). *NeuroImage* 62: 2222-2231.
- Van Essen DC, Glasser MF, Dierker DL, Harwell J, Coalson T (2012b). [Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases](#). *Cereb Cortex* 22(10): 2241-62.
- Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Üğurbil K (2013). [The WU-Minn Human Connectome Project: An Overview](#). *NeuroImage* 80: 62-79.
- Wheatley T, Milleville SC, Martin A (2007). [Understanding animate agents: distinct roles for the social network and mirror system](#). *Psychol Sci* 18(6): 469-474.
- White SJ, Coniston D, Rogers R, Frith U (2011). [Developing the Frith-Happe animations: a quick and objective test of Theory of Mind for adults with autism](#). *Autism Res* 4(2): 149-154.
- Wendelken C, Nakhachenko D, Donohue SE, Carter CS, Bunge SA (2008). [Brain is to thought as stomach is to ??: investigating the role of rostrolateral prefrontal cortex in relational reasoning](#). *J Cogn Neurosci* 20(4): 682-693.
- Woolrich MW, Ripley BD, Brady M, Smith SM (2001). [Temporal autocorrelation in univariate linear modeling of FMRI data](#). *NeuroImage* 14(6): 1370-86.
- Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM (2009). [Bayesian analysis of neuroimaging data in FSL](#). *NeuroImage* 45(1 Suppl): S173-S186.

Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL (2011). [The organization of the human cerebral cortex estimated by intrinsic functional connectivity](#). *J Neurophysiol* 106(3): 1125-1165.

Zaitsev M, Dold, Sakas G, Hennig J, Speck O (2006). [Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system](#). *NeuroImage* 31:1038–1050.

** Further details on the HCP data acquisition and analysis methods as of March 2013 are available in eight manuscripts published in a special Connectome issue of *NeuroImage* ([Van Essen et al., 2013](#), [Üğurbil et al., 2013](#), [Glasser et al. 2013](#), [Smith et al., 2013](#), [Barch et al., 2013](#), [Sotiropoulos et al., 2013](#), [Marcus et al., 2013](#), and [Larson-Prior et al., 2013](#)).

Appendices

Download Appendices at the following links:

[Appendix 1. HCP scan protocols](#)

[Appendix 2. Matlab code for voxel-wise correction of dMRI gradients](#)

[Appendix 3. File Names and Directory Structure for 1200 Subjects Data](#)

[Appendix 4. HCP Protocol Standard Operating Procedures \(SOPs\)](#)

[Appendix 5. MR Data Acquisition Information for an Exemplar Subject](#)

[Appendix 6. Task fMRI and tMEG E-Prime Key Variables](#)

[Appendix 7. tMEG: Extracting trial subsets from data produced by tmegpreproc pipeline](#)