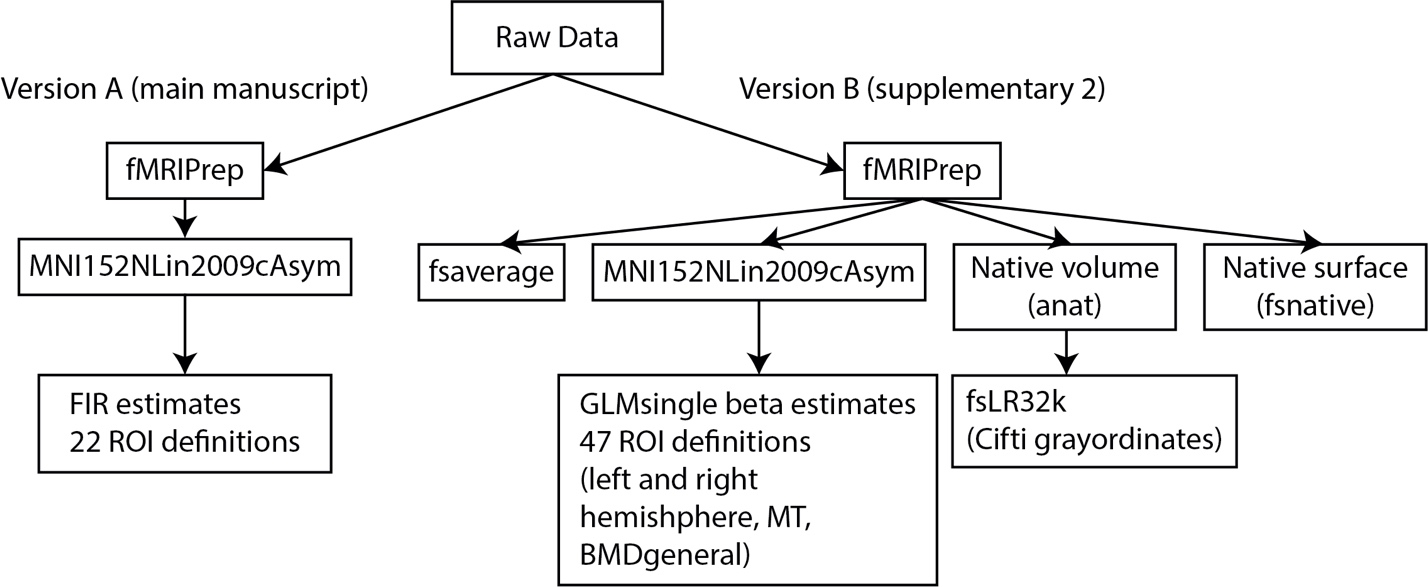
Supplementary #2 Material

1 Overview

This supplementary file details an additional preprocessed version (version B) of the BOLD Moments Dataset (BMD) released alongside the version presented in the manuscript (version A) (see Figure 1). Version B offers additional flexibility to use BMD in a researcher’s desired output space and ROI format. In brief, version B is preprocessed in five output spaces (MNI152NLin2009cAsym, anatomical, fsaverage, fsnative, fsLR32k), contains beta estimates computed with GLMsingle (Prince et al., 2022) for MNI152NLin2009cAsym and fsLR32k spaces, and defines 47 ROIs in MNI152NLin2009cAsym space (left and right hemispheres of the 22 ROIs defined in the main text and MT, plus one “BMDgeneral” ROI). We show whole brain noiseceiling reliability results in the volume-based MNI152NLin2009cAsym and surface-based fsLR32k spaces and high predictivity of a motion energy model in motion-selective ROIs (MT, hV4, V3AB, IPS0).

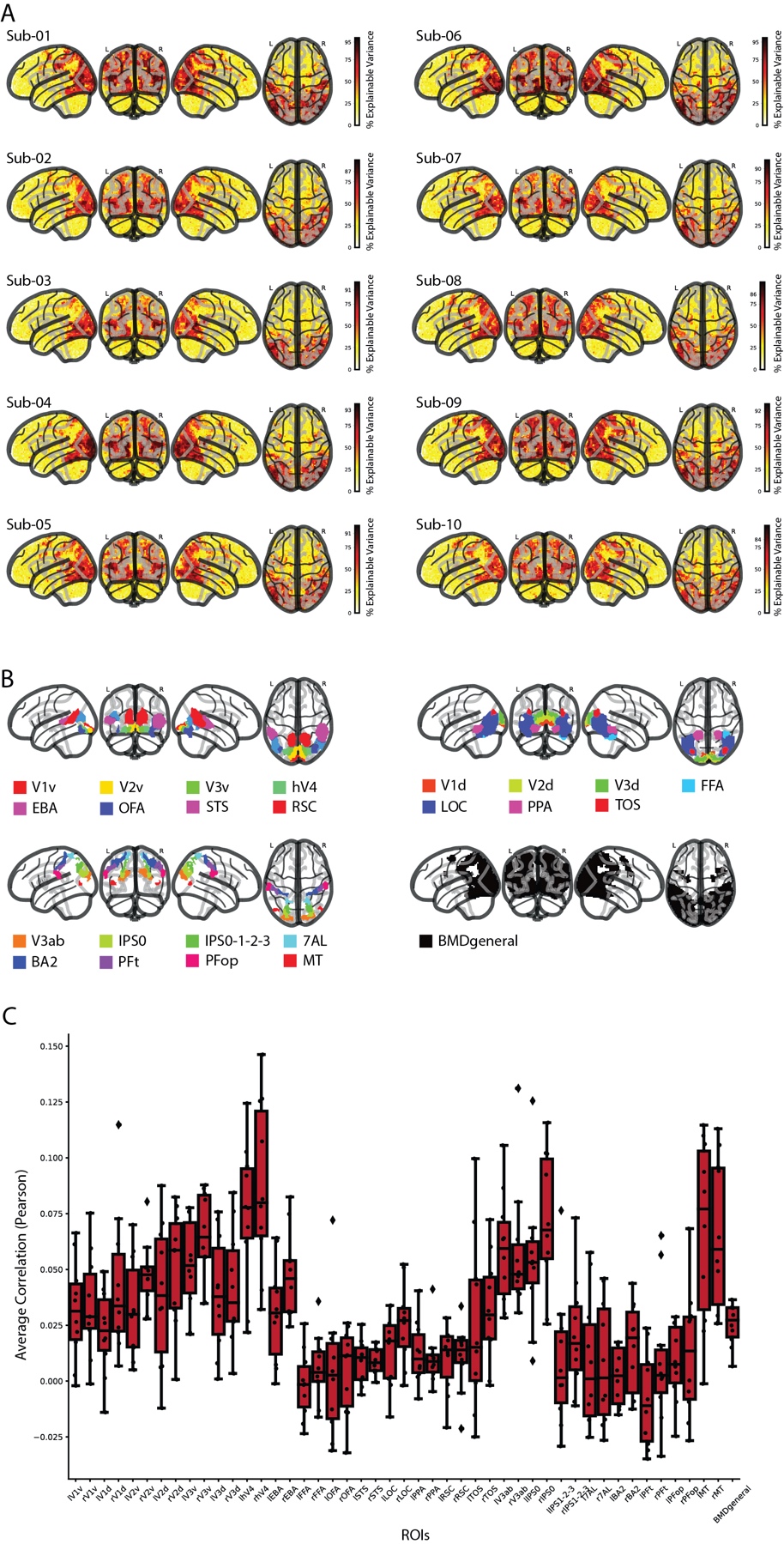
Both version A (presented in the manuscript) and version B (described here) are identical up to fMRIPrep preprocessing (Esteban et al., 2019). Details on the experimental design, participants, and MRI acquisition protocols can be found in the main text. Version A was preprocessed using the default 6 degrees of freedom for BOLD to T1w image registration (the flag –bold2t1w-dof) and one standard volumetric output space (MNI152NLin2009cAsym). Version B was preprocessed with 12 degrees of freedom for BOLD to T1w image registration and four output spaces comprising a standard and native volume output (MNI152NLin2009cAsym, anat) and a standard and native surface output (fsaverage, fsnative) with transformation matrices available between the spaces (see fMRIPrep preprocessing boilerplate text below). We then use version B’s preprocessed results to convert the data to a fifth output space (fsLR32k) using Ciftify (Dickie et al., 2019). Made for the Human Connectome Project (Glasser et al., 2013; Van Essen et al., 2013), CIFTI format organizes brain data in 2D surface-based “grayordinates” for cortical structures and 3D volume-based voxels for subcortical structures. Ciftify adapts the HCP preprocessing pipeline to provide excellent volume-to-surface registration, especially for inter-subject analyses, and access to a suite of HCP analysis and visualization tools (Dickie et al., 2019; Glasser et al., 2013; Robinson et al., 2018).

We provide single trial beta estimates using GLMsingle (Prince et al., 2022) in the volume-based MNI152NLin2009cAsym and surface-based fsLR32k output spaces. We also define 47 ROIs in the MNI152NLin2009cAsym volume space for greater research and modeling flexibility. The 47 ROIs are similar to the 22 ROIs described in the main manuscript but are separated by left/right hemisphere, include the motion-selective MT ROI, include a “BMDgeneral” ROI that broadly defines reliably activated cortex across all BMD subjects, and enforce ROIs to have an equal number of voxels across subjects to facilitate inter-subject modeling.



**Figure 1: Overview of preprocessing pipelines.** Version A (left) of BMD was preprocessed with fMRIPrep into a standard volumetric output space, modeled with FIR functions, and supplemented with 22 ROI definitions. Details are provided in the main manuscript. Version B (right) of BMD was preprocessed with fMRIPrep into two volume-based and two surface-based output spaces. Single trial beta estimates using GLMsingle and 47 ROI definitions were provided in the standard volume-based space. Data was transformed into a fifth output space (fsLR32k) using the Ciftify toolbox and scripts from the Human Connectome Project preprocessing pipeline.

2 Version B MNI152NLin2009cAsym Preprocessing and ROI Definition



**Figure 2: Whole-brain noiseceiling and regions of interest (ROIs)** A) For each subject and voxel in the whole brain, we show the noiseceiling as percent of explainable variance using the testing set videos. The colorbar in each plot is normalized between 0 and 100, and the highest colorbar tick is that subject's maximum explainable variance in the whole brain. B) We show the 46 non-overlapping parcels (combined left and right hemispheres) and BMDgeneral (black) in a glass brain. Each subject's 8 category-selective ROIs (EBA, OFA, STS, RSC, FFA, LOC, PPA, and TOS) are functionally defined by extracting the top 50% most active voxels within the respective parcel. All subjects share the same ROI definition from the remaining parcels. BMDgeneral is defined independently from the ROIs and reflects a group-averaged region of cortex that reliably responds to videos in the BOLD Moments experiment. C) The boxplots depict the correlation (Pearson) between the predicted brain responses with the true responses of the testing set using stimulus features computed from a motion energy model. The boxes show the median response across subjects (horizontal line), 25th and 75th percentile (lower and upper box boundary), and whiskers extending to maximum and minimum values within 1.5 times the interquartile range. Individual subject results are shown as black points, and outliers are shown as diamonds (n=10 subjects for all ROIs).

All fMRI data were organized in the standardized BIDS format (K. J. Gorgolewski et al., 2016) and preprocessed using fMRIPrep (Esteban et al., 2019). The data were slice time corrected, co-registered to the subject's T1w anatomical scan, and spatially normalized to a standard volumetric MNI152NLin2009cAsym template brain. The main experimental runs were then temporally interpolated from their acquisition TR of 1.75 seconds to a TR of 1 second to time-lock volume sampling to stimulus presentations. A General Linear Model (GLM) was used to estimate single trial beta estimates (Prince et al., 2022). The beta responses were then z-scored across video conditions.

2.1 General Linear Model

2.1.1 Functional Localizer

We use GLMsingle (Prince et al., 2022) to model the hemodynamic response function (HRF) to the video localizer for each subject separately. The subject’s preprocessed data in MNI152NLin2009cAsym space was spatially smoothed with a 9mm full width half maximum of the gaussian kernel. The data was then temporally interpolated from an acquisition TR of 1.75s to an interpolated TR of 1s to timelock image acquisition to block onset. Each block, although composed of 6 3s videos (except for the fixation blocks, where no videos were shown), was modeled as a single stimulus. The onsets and durations (18s) of the Body, Face, Object, Scene, Scrambled, and Fixation blocks, along with the temporally interpolated and smoothed fMRI time series, were input to the general linear model. GLMsingle (1) chose an optimal HRF from a library at each voxel, (2) identified a number of nuisance regressors from principal component analysis of a noise pool that explain a maximum amount of variance, and (3) performed fractional ridge regression at each voxel to estimate single trial betas.

2.1.2 Main Experiment

For each subject, we fit beta estimates to each single-trial fMRI response in the main experiment using GLMsingle (Prince et al., 2022). The preprocessed data in MNI152NLin2009cAsym space was temporally interpolated from an acquisition TR of 1.75s to an interpolated TR of 1s to time-lock stimulus onset to image acquisition (e.g., 1.75s does not evenly go into the inter-trial interval of 4s). In this way, we acquire fMRI scans at different timepoints along the BOLD signal (with respect to stimulus presentation) and, after interpolating, achieve a regular sampling of the BOLD signal time-locked to stimulus onset for easier analysis. The interpolated fMRI time series, stimuli onsets, and stimuli durations (modeled as a 0s impulse) for each session separately were input to the general linear model. GLMsingle estimated single trial beta values by (1) fitting an optimal HRF to each voxel from a library of HRFs, (2) identifying nuisance regressors from a noise pool that maximally explain variance, and (3) implementing fractional ridge regression to improve estimates in a rapid event-related design. Responses to testing and training videos were estimated separately.

In this way, we obtained one beta estimate for each stimulus presentation for each subject. This resulted in a total of 4,020 beta estimates per subject (3 beta estimates x 1,000 training videos and 10 beta estimates x 102 testing videos).

Nan-indices corresponding to outside the subject’s brain mask were identified and removed. For the training and testing data separately, the beta estimates were then z-scored across video conditions such that the response profile for each stimulus presentation at each voxel (i.e., a vector of length 1000 for training data or of length 102 for testing data) had a mean value of 0 and standard deviation of 1.

2.2 Regions of Interest Definition

We computed a non-overlapping set of 46 ROIs (regions of interest) (23 ROIs separated by left and right hemispheres) previously known to be driven by dynamic stimuli spanning visual and parietal cortices (Gazzola & Keysers, 2009; Le et al., 2017; Logothetis & Sheinberg, 1996; R. Peeters et al., 2009; R. R. Peeters et al., 2013; Rizzolatti & Sinigaglia, 2010; Silver & Kastner, 2009; VanRullen & Thorpe, 2001). Note that these ROI definitions differ slightly compared to those detailed in the main manuscript (version A). We first created a non-overlapping parcellation in the standard MNI152NLin2009cAsym space identical across subjects composed of parcels resampled from Wang and colleagues, Glasser and colleagues, and Julian and colleagues (Glasser et al., 2016; Julian et al., 2012; Wang et al., 2015). Finally, we used the t-contrasts from each subject’s functional localizer results to identify the top 50% of voxels within the corresponding functional parcel from Julian and colleagues (Julian et al., 2012) (bodies > objects: EBA; objects > scrambled: LOC; scenes > objects: PPA, RSC, STS; faces > objects: OFA, FFA, STS). This ROI definition method facilitated inter-subject modeling approaches by ensuring all ROIs were defined for each subject and each ROI contained the same number of voxels across subjects. Furthermore, the parcellation shared across subjects (before taking each subject’s top 50% of voxels in the parcels from Julian and colleagues (Julian et al., 2012)) allowed modeling approaches that incorporate voxel-level spatial information, since the parcel indices are the same for each subject.

In detail, we defined ROIs V1v, V1d, V2v, V2d, V3v, V3d, hV4, V3a, V3b, IPS0, IPS1, IPS2, IPS3 from Wang and colleagues (Wang et al., 2015) (maxprob\_vol\_{h}h.nii, where {h} is “l” or “r”), L\_2 (here referred to as BA2), 7AL, PFt, PFop, and MT from Glasser and colleagues (Glasser et al., 2016) (HCPMMP1\_on\_MNI152\_ICBM2009a\_nlin.nii), and EBA, LOC, PPA, RSC, STS, OFA, FFA, and STS from Julian and colleagues (Julian et al., 2012) ({h}{parcel}.img from the n=30 group, where {h} is “l” or “r” and {parcel} is the parcel name). We group V3a and V3b into V3ab and IPS1, IPS2, and IPS3 into IPS1-2-3 due to subtle differences in functional preferences that can be difficult to resolve with our in-the-wild naturalistic stimuli (Georgieva et al., 2009; Konen & Kastner, 2008; Press et al., 2001; Smith et al., 1998). All ROIs were resampled into our functional volumetric dimensions and separated by left and right hemispheres. Voxels outside a common brain mask computed across subjects were removed from the parcel. There was no overlap between the parcels within Wang and colleagues (Wang et al., 2015) or between the parcels within Glasser and colleagues (Glasser et al., 2016).

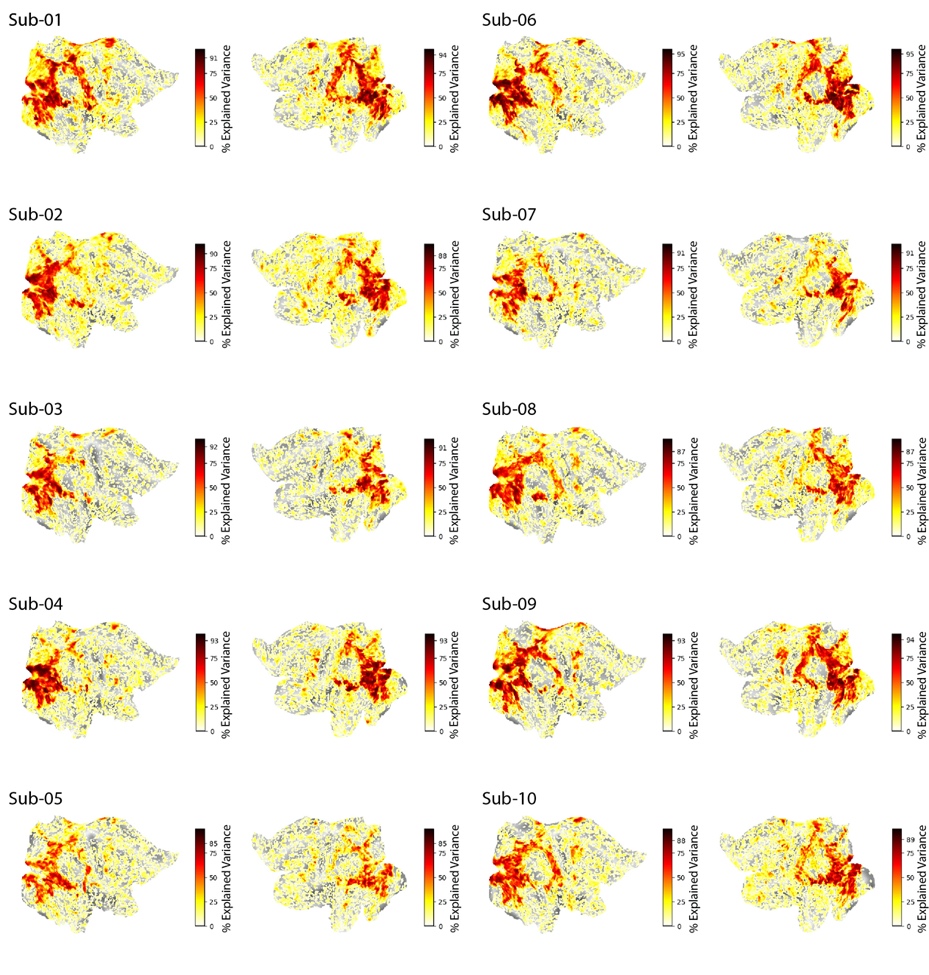
To address minimal overlap between the parcels derived from Julian and colleagues (Julian et al., 2012), the in-question voxels were assigned to the parcels that had a greater inter-subject agreement from our functional localizer experiment. Specifically, a t-test (two-sided, independent) was computed between the condition beta estimates to define category-selective contrasts: bodies > objects (body selective; EBA), objects > scrambled (object selective; LOC), scenes > objects (scene selective; PPA, RSC, TOS), and faces > objects (face selective; OFA, FFA, STS). Similar to the Group-constrained Subject Specific (GSS) procedure (Julian et al., 2012), the t-contrast maps were binarized at a p-value cutoff (p < 0.05, uncorrected), where voxels below this cutoff were assigned 0 and voxels above this cutoff were assigned 1. The binarized t-contrast maps were averaged across subjects and smoothed (6mm full width half maximum of the gaussian kernel) to obtain four probability maps (one for each contrast) that contains information on inter-subject agreement at each voxel. If both LOC and EBA overlapped at voxel A, for example, we indexed voxel A’s value in both the objects > scrambled probability map (for LOC) and the bodies > objects probability map (for EBA) and assigned voxel A to the parcel with the higher value (i.e., the higher inter-subject agreement). No parcels within a contrast overlapped.

Finally, we addressed the minimal overlap between parcels across the three atlases. If non-functionally defined parcels (i.e., the parcels derived from Wang and colleagues (Wang et al., 2015) and Glasser and colleagues (Glasser et al., 2016)) overlapped with functionally-defined parcels (i.e., from Julian and colleagues (Julian et al., 2012)), preference was given first to the non-functionally defined parcel. Otherwise, the voxel may end up not being assigned to any ROI after subject-specific ROI definition (described below) because the functionally-defined parcels are generous in size, reflecting group-level inter-subject agreement. If the overlapping parcels were all non-functionally defined, preference would have gone to the smaller parcel to preserve its size, but there was only overlap between functionally and non-functionally defined parcels. In this way, we obtained 46 non-overlapping parcels identical for each subject.

We then functionally define the category-selective ROIs for each subject by identifying the top 50% most active (i.e., highest t-values, uncorrected) voxels inside the ROIs respective t-contrast map masked by the corresponding parcel (Mineroff et al., 2018). This method achieves both a subject-specific functional definition, maintains the relative size between ROIs, and ensures the same number of voxels within an ROI across subjects. No constraint of contiguity was enforced.

We additionally define a swathe of cortex that showed consistently reliable responses to videos across subjects in this study, here termed BMDgeneral. BMDgeneral was algorithmically defined in five steps: (1) compute split-half correlations across stimuli repetitions to obtain p-values at each voxel, (2) binarize the volume at p < 0.05 (uncorrected), indicating voxels with a value of 0 and 1 have poor and good split-half reliability, (3) average each subject's binarized mask to obtain a probability map reflecting the inter-subject agreement of each voxel's reliability, (4) smooth the probability map (6mm at full width half maximum), and (5) identify clusters (cluster threshold = 50 voxels, 8mm between peaks, statistic threshold of 0.1). All identified clusters (here, 4 clusters) are collectively identified as BMDgeneral. BMDgeneral may or may not overlap with the 46 ROIs.

3 Version B CIFTI Preprocessing



**Figure 3: Left and right cortex noiseceiling.** For each subject and grayordinate vertex in the left and right flattened hemispheres, we show the noiseceiling as percent of explainable variance using the testing set videos. The colorbar in each plot is normalized between 0 and 100, and the highest colorbar tick is that subject's maximum explainable variance in the whole brain. Values are thresholded at 1.

After fMRIPrep preprocessing, data was converted to CIFTI format using the Ciftify framework (version 2.3.3) (Dickie et al., 2019), which adapts scripts from the Human Connectome Project (HCP) preprocessing pipeline (Glasser et al., 2013). Ciftify allows registration of data to fsLR32k output space in a HCP compatible directory structure while still respecting BIDS organization (K. J. Gorgolewski et al., 2016)for the use of BIDS-apps (e.g., fMRIPrep). First, the “ciftify\_recon\_all” function converts a subject’s freesurfer output directory (output from fMRIPrep) into CIFTI format, registers their anatomical data from to a surface-based mesh using the MSMSulc algorithm (Robinson et al., 2018), then resamples to a 32k standard space (fsLR32k). Second, the “cifity\_subject\_fmri” maps the preprocessed functional scans to the subject’s fsLR32k space. Voxels along the cortical ribbon are weighted and projected to vertices on the surface (e.g., “grayordinates”), and subcortical voxels (including cerebellum) are resampled to volumetric MNI space. No smoothing was applied.

Note that preprocessing BMD through the entire HCP preprocessing pipeline is expected to obtain even better results due to the availability of high resolution T2w and fieldmap scans. We use Ciftify here to maintain a BIDS compliant dataset (for use with other BIDS-apps) and a common fMRIPrep preprocessing root between output spaces all while still making the data immediately available in the advantageous CIFTI format.

3.1 General Linear Model

3.1.1 Main Experiment

We estimate single-trial beta values at each vertex on the fsLR32k cortical surface for each subject using GLMsingle (Prince et al., 2022). The unsmoothed time series output from Ciftify (Dickie et al., 2019) was temporally interpolated to a TR of 1s (acquisition TR=1.75s) in order to temporally align image acquisition with stimulus onset. The video stimulus trials were entered into the general linear model (GLM) as regressors of interest. Null trials were not entered into the GLM. GLMsingle then computed single-trial beta estimates by optimally fitting one of twenty hemodynamic response functions (HRFs) to each voxel, calculating nuisance regressors from noisy vertices, and incorporating fractional ridge regression for improved beta estimates in a rapid event-related experimental designs. The GLM was computed per session, and testing and training runs were computed independently. This procedure resulted in 4,020 single-trial beta estimates for each subject.

In this way, we obtained one beta estimate for each stimulus presentation for each subject. This resulted in a total of 4,020 beta estimates per subject (3 beta estimates x 1,000 training videos and 10 beta estimates x 102 testing videos).

4 Noiseceiling Calculation

We compute the percent of explainable variance at each voxel for each subject as an estimate of the noiseceiling (equation 2) (Allen et al., 2022; Prince et al., 2022). First, the noise, signal and total variance of the beta estimates is computed. The noise variance is computed as the mean variance of the beta estimates of the within-video presentation trials. The total variance is computed as the variance of the beta estimates across all video presentation trials. The signal variance is computed as the total variance minus the noise variance. The signal variance was positively rectified, where negative values were assigned 0 and positive values were preserved. Next, the noiseceiling signal-to-noise ratio (SNR) (*ncsnr*) is computed as the fraction of signal standard deviation () to noise standard deviation () (equation 1),

(equation 1)

where the standard deviation is equal to the square root of the variance (). Finally, the percentage of explainable variance (*noiseceiling*) is calculated as,

(equation 2)

where *ncsnr* is the noiseceiling signal-to-noise ratio (SNR) and *numTrials* is the number of video presentation trials (*numTrials*=10 for the testing set and *numTrials*=3 for the training set). This measure of voxel reliability differs from the split-half reliability measure used in the main manuscript (version A) but produce very similar results. The reliability measure proposed here is computationally less expensive than a split-half computation and has no stochastic elements.

5 Motion Energy Features Computation and Encoding Model

Motion energy features were used to predict brain activity in response to BMD’s 3 second naturalistic videos. The motion energy model (Adelson & Bergen, 1985; Nishimoto et al., 2011; Watson & Ahumada, 1985) consists of a series of spatial and temporal Gabor filters intended to capture local motion and direction in a video stimulus, thus making it a highly interpretable method to model video dynamics. The motion energy encoding model accuracy (Figure 2C) shows highest prediction accuracy in motion selective ROIs, namely MT (Born & Bradley, 2005; Nishimoto & Gallant, 2011), hV4 (Kamitani & Tong, 2006; Roe et al., 2012), V3AB (Konen & Kastner, 2008; Smith et al., 1998), and IPS0 (Konen & Kastner, 2008). These results support that single trial beta estimates of BMD’s 3 second naturalistic videos capture motion information.

Motion energy features for each BMD video stimulus was computed using the MATLAB code available here: <https://github.com/gallantlab/motion_energy_matlab> (Nishimoto et al., 2011; Nishimoto & Gallant, 2011). For each 268x268 video, the frames were converted from RGB to LAB color space, and only the L (luminance) channel was retained. The luminance channel was then passed through a three-dimensional bank of spatiotemporal Gabor filters consisting of two spatial dimensions and one temporal dimension. Similar to the filter bank used in (Nishimoto et al., 2011) to model naturalistic movies, the three-dimensional filters are defined at five spatial frequencies (0, 2, 4, 8, 16, and 32 cycles/image), three temporal frequencies (0, 2, and 4Hz), and eight directions (0, 45, 90, 135, 180, 225, 270, and 315 degrees) with the exception that the 0 Hz temporal filter is defined at only 0, 45, 90, and 135 degrees directions and the 0 cycles/image spatial filter is defined at 0 degree orientation. Local motion-energy features were computed by taking the square root of the sum of the squared outputs of each pair of filters with orthogonal phases. The logarithm of the output from these filters was computed to scale large values, and the temporal dimension of the output was downsampled to 1 second to match the fMRI sampling rate (i.e., the interpolated TR of 1 second) of the BOLD time series. The output was then z-scored across time. In total, this procedure resulted in a matrix of size 3 x 6555 (seconds x motion energy features).

The motion energy features were then used in a voxelwise linear encoding model (Naselaris et al., 2011) to predict the brain activity (beta estimates) in 47 regions of interest (ROIs) from the version B preprocessed data in MNI152NLin2009cAsym space (Figure 2C). Specifically, the motion energy features for each video were concatenated along the three seconds and underwent principal component analysis (PCA) to reduce dimensionality to the top 100 components. PCA was fit to the training videos and applied to both the training and testing videos. A linear model was then fit to the training video features to predict the response at the voxel. The learned weights of the linear model were then applied to the testing video features. The encoding model accuracy was computed as the correlation of the vector of predicted responses of the test set with the vector of true responses of the test set.

6 fMRIPrep Preprocessing Boilerplate Text

We reproduce the fMRIPrep boilerplate text describing version B’s preprocessing details below:

“”

Results included in this manuscript come from preprocessing performed using fMRIPrep 23.0.2 (Esteban et al., 2019, 2022; RRID:SCR\_016216), which is based on Nipype 1.8.6 (Esteban, Oscar et al., 2022; K. Gorgolewski et al., 2011; RRID:SCR\_002502).

#### Preprocessing of B0 inhomogeneity mappings

A total of 6 fieldmaps were found available within the input BIDS structure for this particular subject. A B0 nonuniformity map (or fieldmap) was estimated from the phase-drift map(s) measure with two consecutive GRE (gradient-recalled echo) acquisitions. The corresponding phase-map(s) were phase-unwrapped with prelude (FSL 6.0.5.1:57b01774).

#### Anatomical data preprocessing

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008; RRID:SCR\_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (6.0.5.1:57b01774, RRID:SCR\_002823, Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 7.3.2, RRID:SCR\_001847, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR\_002438, Klein et al., 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was were selected for spatial normalization and accessed with TemplateFlow (23.0.0, Ciric et al., 2022): ICBM 152 Nonlinear Asymmetrical template version 2009c [Fonov et al., 2009, RRID:SCR\_008796; TemplateFlow ID: MNI152NLin2009cAsym].

#### Functional data preprocessing

For each of the 62 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1:57b01774, Jenkinson et al., 2002). The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. BOLD runs were slice-time corrected to 0.834s (0.5 of slice acquisition range 0s-1.67s) using 3dTshift from AFNI (Cox & Hyde, 1997, RRID:SCR\_005927). The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with twelve degrees of freedom to account for distortions remaining in the BOLD reference. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al., 2014) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al., 2002). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer’s aseg segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components’ time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by (Patriat et al., 2017). All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using mri\_vol2surf (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.9.1 (Abraham et al., 2014, RRID:SCR\_001362), mostly within the functional processing workflow. For more details of the pipeline, see [the section corresponding to workflows in *fMRIPrep*’s documentation](https://fmriprep.readthedocs.io/en/latest/workflows.html).

“”

References:

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://www.frontiersin.org/articles/10.3389/fninf.2014.00014

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *JOSA A*, *2*(2), 284–299. https://doi.org/10.1364/JOSAA.2.000284

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), Article 1. https://doi.org/10.1038/s41593-021-00962-x

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Born, R. T., & Bradley, D. C. (2005). Structure and Function of Visual Area Mt. *Annual Review of Neuroscience*, *28*(1), 157–189. https://doi.org/10.1146/annurev.neuro.26.041002.131052

Ciric, R., Thompson, W. H., Lorenz, R., Goncalves, M., MacNicol, E. E., Markiewicz, C. J., Halchenko, Y. O., Ghosh, S. S., Gorgolewski, K. J., Poldrack, R. A., & Esteban, O. (2022). TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models. *Nature Methods*, *19*(12), Article 12. https://doi.org/10.1038/s41592-022-01681-2

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4–5), 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Dickie, E. W., Anticevic, A., Smith, D. E., Coalson, T. S., Manogaran, M., Calarco, N., Viviano, J. D., Glasser, M. F., Van Essen, D. C., & Voineskos, A. N. (2019). ciftify: A framework for surface-based analysis of legacy MR acquisitions. *NeuroImage*, *197*, 818–826. https://doi.org/10.1016/j.neuroimage.2019.04.078

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Esteban, O., Markiewicz, C. J., Goncalves, M., Provins, C., Kent, J. D., DuPre, E., Salo, T., Ciric, R., Pinsard, B., Blair, R. W., Poldrack, R. A., & Gorgolewski, K. J. (2022). *fMRIPrep: A robust preprocessing pipeline for functional MRI* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.7430291

Esteban, Oscar, Markiewicz, Christopher J., Burns, Christopher, Goncalves, Mathias, Jarecka, Dorota, Ziegler, Erik, Berleant, Shoshana, Ellis, David Gage, Pinsard, Basile, Madison, Cindee, Waskom, Michael, Notter, Michael Philipp, Clark, Daniel, Manhães-Savio, Alexandre, Clark, Dav, Jordan, Kesshi, Dayan, Michael, Halchenko, Yaroslav O., Loney, Fred, … Ghosh, Satrajit. (2022). *nipy/nipype: 1.8.3* (1.8.3) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.596855

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Gazzola, V., & Keysers, C. (2009). The Observation and Execution of Actions Share Motor and Somatosensory Voxels in all Tested Subjects: Single-Subject Analyses of Unsmoothed fMRI Data. *Cerebral Cortex*, *19*(6), 1239–1255. https://doi.org/10.1093/cercor/bhn181

Georgieva, S., Peeters, R., Kolster, H., Todd, J. T., & Orban, G. A. (2009). The Processing of Three-Dimensional Shape from Disparity in the Human Brain. *Journal of Neuroscience*, *29*(3), 727–742. https://doi.org/10.1523/JNEUROSCI.4753-08.2009

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

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

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, *60*(4), 2357–2364. https://doi.org/10.1016/j.neuroimage.2012.02.055

Kamitani, Y., & Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Current Biology*, *16*(11), 1096–1102. https://doi.org/10.1016/j.cub.2006.04.003

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Konen, C. S., & Kastner, S. (2008). Representation of Eye Movements and Stimulus Motion in Topographically Organized Areas of Human Posterior Parietal Cortex. *Journal of Neuroscience*, *28*(33), 8361–8375. https://doi.org/10.1523/JNEUROSCI.1930-08.2008

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, *1*(1), 76–85.

Le, A., Vesia, M., Yan, X., Crawford, J. D., & Niemeier, M. (2017). Parietal area BA7 integrates motor programs for reaching, grasping, and bimanual coordination. *Journal of Neurophysiology*, *117*(2), 624–636. https://doi.org/10.1152/jn.00299.2016

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621. https://doi.org/10.1146/annurev.ne.19.030196.003045

Mineroff, Z., Blank, I. A., Mahowald, K., & Fedorenko, E. (2018). A robust dissociation among the language, multiple demand, and default mode networks: Evidence from inter-region correlations in effect size. *Neuropsychologia*, *119*, 501–511. https://doi.org/10.1016/j.neuropsychologia.2018.09.011

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073

Nishimoto, S., & Gallant, J. L. (2011). A Three-Dimensional Spatiotemporal Receptive Field Model Explains Responses of Area MT Neurons to Naturalistic Movies. *Journal of Neuroscience*, *31*(41), 14551–14564. https://doi.org/10.1523/JNEUROSCI.6801-10.2011

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. *Current Biology*, *21*(19), 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031

Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *NeuroImage*, *144*, 74–82. https://doi.org/10.1016/j.neuroimage.2016.08.051

Peeters, R. R., Rizzolatti, G., & Orban, G. A. (2013). Functional properties of the left parietal tool use region. *NeuroImage*, *78*, 83–93. https://doi.org/10.1016/j.neuroimage.2013.04.023

Peeters, R., Simone, L., Nelissen, K., Fabbri-Destro, M., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2009). The Representation of Tool Use in Humans and Monkeys: Common and Uniquely Human Features. *Journal of Neuroscience*, *29*(37), 11523–11539. https://doi.org/10.1523/JNEUROSCI.2040-09.2009

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Press, W. A., Brewer, A. A., Dougherty, R. F., Wade, A. R., & Wandell, B. A. (2001). Visual areas and spatial summation in human visual cortex. *Vision Research*, *41*(10), 1321–1332. https://doi.org/10.1016/S0042-6989(01)00074-8

Prince, J. S., Charest, I., Kurzawski, J. W., Pyles, J. A., Tarr, M. J., & Kay, K. N. (2022). Improving the accuracy of single-trial fMRI response estimates using GLMsingle. *eLife*, *11*, e77599. https://doi.org/10.7554/eLife.77599

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), Article 4. https://doi.org/10.1038/nrn2805

Robinson, E. C., Garcia, K., Glasser, M. F., Chen, Z., Coalson, T. S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., Hutter, J., Price, A., Cordero Grande, L., Hughes, E., Tusor, N., Bayly, P. V., Van Essen, D. C., Smith, S. M., Edwards, A. D., … Rueckert, D. (2018). Multimodal surface matching with higher-order smoothness constraints. *NeuroImage*, *167*, 453–465. https://doi.org/10.1016/j.neuroimage.2017.10.037

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., Lu, H., & Vanduffel, W. (2012). Toward a Unified Theory of Visual Area V4. *Neuron*, *74*(1), 12–29. https://doi.org/10.1016/j.neuron.2012.03.011

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Sciences*, *13*(11), 488–495. https://doi.org/10.1016/j.tics.2009.08.005

Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M., & Hennig, J. (1998). The Processing of First- and Second-Order Motion in Human Visual Cortex Assessed by Functional Magnetic Resonance Imaging (fMRI). *Journal of Neuroscience*, *18*(10), 3816–3830. https://doi.org/10.1523/JNEUROSCI.18-10-03816.1998

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

VanRullen, R., & Thorpe, S. J. (2001). The Time Course of Visual Processing: From Early Perception to Decision-Making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461. https://doi.org/10.1162/08989290152001880

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual-motion sensing. *JOSA A*, *2*(2), 322–342. https://doi.org/10.1364/JOSAA.2.000322

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424