Depth-resolved intersubject functional correlation of 7T BOLD signals reveals increased stimulus related information sharing across deep layers in premotor and parietal nodes for predictable actions.

Leonardo Cerliani[1,*], Ritu Bhandari[1,*], Lorenzo De Angelis[1,*], Wietske van der Zwaag[2], Pierre-Louis Bazin[1,3,,4], Valeria Gazzola[1,5,#] and Christian Keysers[1,5,#]

Affiliations:
1. Social Brain Lab, Netherlands Institute for Neuroscience, Meibergdreef 47, 1105 BA Amsterdam, Netherlands
2. Spinoza Centre for Neuroimaging, Amsterdam, the Netherlands
3. Integrative Model-Based Cognitive Neuroscience Research Unit, University of Amsterdam, Nieuwe Achtergracht 129, 1018 WS, Amsterdam, The Netherlands
4. Departments of Neurophysics and Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
5. Department of Psychology, University of Amsterdam, Amsterdam, Netherlands
* contributed equally to the role of first author
# contributed equally to the role of last author

Abstract

While the brain regions involved in action observation are relatively well documented in humans and primates, how these regions communicate to help understand and predict actions remains poorly understood. Traditional views emphasized a feed-forward architecture in which visual features are organized into increasingly complex representations that feed onto motor programs in parietal and then premotor cortices where the matching of observed actions upon the observer's own motor programs contributes to action understanding. Predictive coding models place less emphasis on feed-forward connections and propose that feed-back connections from premotor regions back to parietal and visual neurons represent predictions about upcoming actions that can supersede visual inputs when actions become predictable, with visual input then merely representing prediction errors. Here we leverage the notion that feed-back connections target deeper cortical layers than feed-forward connections to help adjudicate across these views. Specifically, we test whether observing sequences of hand actions in their natural order, which permits participants to predict upcoming actions, triggers more feed-back input to parietal regions than seeing the same actions in a scrambled sequence that hinders making predictions. Using submillimeter fMRI acquisition at 7T, we find that watching predictable sequences triggers more action-related activity (as measured using intersubject functional correlation) in deep layers of the parietal cortical area PFt than watching the exact same actions in scrambled and hence unpredictable sequence. In addition, functional connectivity analysis performed using intersubject functional connectivity confirms that this deep, action-related signals in PFt could originate from ventral premotor region BA44. This data showcases the utility of intersubject functional correlation in combination with 7T MRI to explore the architecture of social cognition under more naturalistic conditions, and provides evidence for models that emphasize the importance of feed-back connections in action prediction.

Introduction

The network of brain regions recruited by the observation of actions is relatively well described across humans and monkeys. In monkeys, using anatomical tracers, electrophysiology and fMRI, a temporo-parieto-frontal triad of mutually connected regions responds to the sight of hand actions (Nelissen et al. 2005, 2011; Rizzolatti and Sinigaglia 2016): regions lining the superior temporal sulcus (STS) are mutually connected with inferior parietal nodes (AIP/PFG) that are mutually connected with ventral premotor regions (F5a/c). In humans, exact homologies are difficult to establish, but studies and meta-analyses showed that observing goal-directed hand actions increase activity in lateral occipital/posterior temporal, parietal and premotor clusters that often peak in the lateral occipital cortex, parietal region PFt (similar to monkey PFG) and premotor region BA44 (similar to monkey F5a), and that these regions are also activated during the execution of similar actions (Gazzola and Keysers 2009; Caspers et al. 2010; Rizzolatti and Sinigaglia 2016).

While this network has long been implicated in participants' ability to perceive the actions of others, a more recent hypothesis has been that they play a key role in *predicting* the actions of others. Reviewing recent neuromodulation studies confirms that perturbing activity in these regions interferes with the participants' ability to predict upcoming actions in sequences (Keysers et al. 2018). Single cell recordings in parietal and premotor nodes in monkeys revealed that neurons responding to the sight of hand actions often have responses that differ based on the sequence of actions that can be predicted to follow (Bonini et al. 2010). Such recordings have also shown that responses to seeing a hand grasp behind an occluder are increased when an animal can predict the presence of an object behind that occluder (Umiltà et al. 2001).

To map the brain regions encoding information about the sequence of motor acts that form complex action sequences, we video-recorded familiar sequences of acts that jointly form meaningful action sequences (Fig. 1a and Movie 1), and compared brain activity - measured using fMRI at 3T - between conditions in which participants viewed the motor acts either in their original (and thus predictable) sequence (Intact condition) or in a randomized (and thus unpredictable) order (Scrambled condition), (Thomas et al. 2018). We reasoned that nodes of the action observation network can either only contain information about individual motor acts or also additional information about the sequence in which acts organize into complex actions with more distal goals. If a node only represents individual motor acts, it should respond similarly whether these acts are in their natural order (Intact) or not (Scrambled). If it contains information at the sequence level, we would expect its activity to be sensitive to the order in which motor acts are presented, and thus respond differentially.

To identify nodes with BOLD signals containing sequence level information, we used intersubject correlation (ISC, Nastase et al. 2019) to map and quantify stimulus locked information in the brain. Lerner et al. (Lerner et al. 2011) had pioneered this approach to identify sequence level information in language processing by having participants listen to narrated stories either in their intact order, or in scrambled order, in which for instance the same words appeared in random order. They found that while auditory cortices show similar levels of synchronization across listeners whether the story was heard intact or scrambled, frontal brain regions synchronized across listeners much more in the intact condition, showing that frontal regions integrate information across words into meaningful sequences. Adapting this elegant approach to actions, we found that left PFt, and to a lesser extent BA44, showed

higher ISC for intact than scrambled hand action sequences, establishing that these regions contain sequence level information (Thomas et al. (2018)). Temporal and occipital visual regions showed high ISC for Intact and Scrambled sequences, and thus encoded information about the observed acts, but did not show higher ISC in the intact compared to scrambled condition, suggesting that they do not contain significant levels of sequence level information. Together, these results sketch a hierarchical picture in which the temporal/occipital lobes process observed actions, but at the relatively short temporal scales of single motor acts, while PFt and BA44 process these acts in ways that integrate that information into longer - and arguably more meaningful - sequences.

How this network integrates individual motor acts into meaningful sequences is in the focus of a number of theoretical papers (Keysers and Perrett 2004; Kilner and Frith 2008; Friston et al. 2011; Keysers and Gazzola 2014), but empirical data to support these notions remains scarce. Early work emphasized feed-forward information flow in a hierarchical structure from the temporal lobe to parietal and premotor regions where action understanding and action prediction then occurs by triggering premotor action plans in mirror neurons. For instance, Gallese et al. in their seminal 1996 paper wrote: “The superior temporal sulcus representation would provide, in this case, an initial description of hand object interactions that would then be sent to F5 and matched with the 'motor vocabulary' of that area. […] When an individual is carrying out a goal-directed action, he 'knows' (predicts) its consequences. [Premotor] Mirror neurons could be the means by which this type of knowledge can be extended to actions performed by others” (Gallese et al. 1996, p606). Indeed, MEG data supported this notion by measuring increasing latencies across these nodes (Nishitani and Hari 2000). A problem with this feed-forward view is that it implies that for a person to respond to the actions of others, the sight and sound of that action would need to trigger activity in a feed-forward stream through high level sensory, parietal and then premotor regions. This would imply delays of ~200ms between witnessing an action and executing a response. However, pianists have been shown to adjust their own key presses to within 30ms of those of others (Keller et al. 2007) - an order of magnitude too fast for such a simple feed-forward architecture (Keysers and Gazzola 2014). More recently, it has been argued that the action observation system could instead use Bayesian inference in a predictive coding system that combines the feed-forward transmission of prediction errors with feed-back transmission of predictions across the mutually connected nodes of the system (Keysers and Perrett 2004; Kilner and Frith 2008; Friston et al. 2011; Keysers and Gazzola 2014). Specifically, we proposed that through their mutual connectivity, the premotor-parietal system, which orchestrates the temporal unfolding of complex action sequences during motor control, has the sequence level information spanning tens of seconds that psychologists call ‘motor schemata’, while the occipito/temporal nodes only represent individual actions, and receive predictions from the parietal lobe, and send prediction errors to the parietal lobe (Thomas et al. 2018). If the predictions match the incoming visual information (i.e. small or no prediction error) the sensory input from visual areas is attenuated - an attenuation we indeed measured in ERP responses over the occipital lobe (Thomas et al. 2018). If they do not match, the difference (large prediction error) is sent forward to parietal and premotor regions where they can activate alternative motor schemata (i.e. action sequences) that better match the visual input of the observed action sequences. If observed acts continuously violate expectations, participants report stopping to generate predictions and premotor cortices may then stop to send such predictions backwards towards parietal cortices (Thomas et al. 2018). A core - and testable - tenet of this predictive coding hypothesis is that the sight of actions organized in predictable sequences should induce more feed-back information flow instantiating predictions from BA44 to PFt while scrambled sequences should induce more feed-forward information flow instantiating prediction

errors (Fig. 1b). The traditional, hierarchical feed-forward view of action observation would not predict such an increase of feed-back information as subsequent motor acts in the sequence become predictable. In terms of time scales of information, the feed-forward input from visual regions would contain prediction errors at each transition across acts, in the scrambled condition in particular, that would have relatively fast frequencies (every ~2 s), while the feed-back input from premotor regions would contain a mix of frequencies that span from signals that predict the upcoming act (every ~2 s) and sequence level information at lower frequencies (with periods T up to ~1 min). Such a mix of frequencies has been observed in similar experiments on language (Honey et al. 2012).

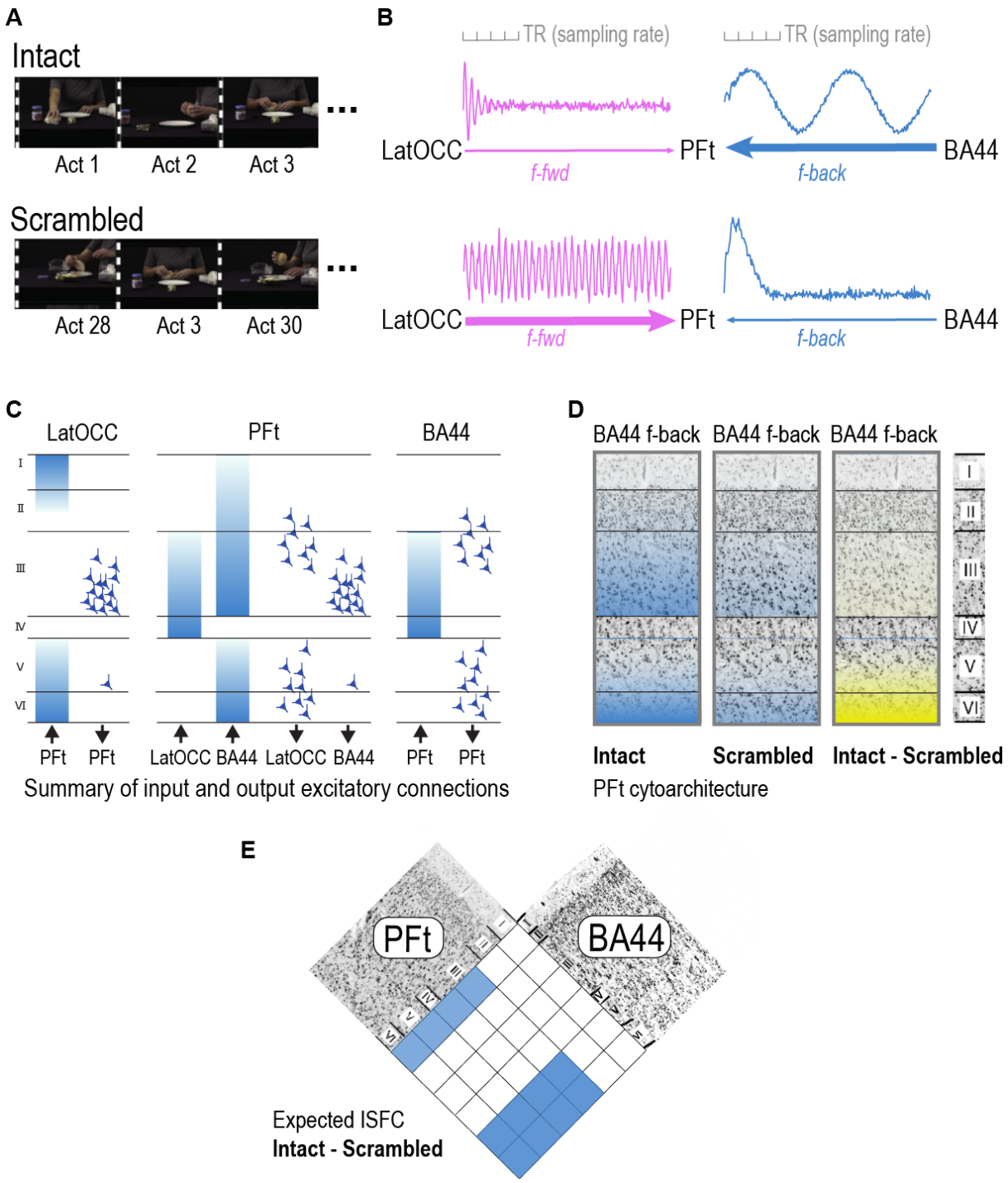

Figure 1: Stimuli and hypotheses.

(a) We presented participants with movies of everyday hand actions lasting ~1 min in length, filmed simultaneously with two cameras 45° apart, that were cut into ~30 individual motor acts lasting ~2 s (Table 1). In the intact condition, the motor acts were presented in their original order, but switching from one camera-view to the other at the transition between acts. In the scrambled condition, the acts were presented in randomized order. The change between camera-views were introduced in both conditions, because randomizing would otherwise have introduced visual transients in the Scrambled but not in the Intact movie.

(b) Based on predictive coding, we hypothesize that **for Intact sequences** (top row), the parietal region PFt should receive comparatively little prediction errors from high level visual regions in the lateral occipital cortex (LatOcc). This is because in predictable sequences, only the first act cannot be predicted, and should thus trigger a strong prediction error, while subsequent acts become increasingly predictable, leading to waning prediction-error input to PFt. In contrast, PFt should receive strong and sustained predictive feedback from ventral premotor regions, and especially from BA44 (Keysers and Perrett 2004; Kilner and Frith 2008; Friston et al. 2011; Keysers and Gazzola 2014). Input from visual regions should be dominated by higher frequencies representing the individual motor acts presented every ~2 s, while feedback from premotor regions would additionally include lower frequencies encoding sequence level information with periods up to the full ~1 min of the movie (Honey et al. 2012; Thomas et al. 2018). **For Scrambled sequences**, PFt should receive sustained prediction errors from visual cortices while feed-back predictions from the premotor area should increasingly wane as participants realize that the sequence is unpredictable.

(c) Adapting general organizational principles of the cortex summarized in Shipp (2007) by incorporating specific knowledge about the connectivity between ventral premotor and inferior parietal cortices obtained from monkey tracer studies (Gerbella et al. 2011) we can generate hypotheses about the layers in which the feed-forward information from LatOcc and the feed-back information from BA44 should arrive in PFt (gradients, middle panel), and where the cells in latOcc and BA44 from which this input originates are located (cells in the right and left panel, respectively). In addition, we can hypothesize where PFt signals back to LatOcc and forward to BA44 might terminate (gradients right and left panels), and where they might originate in PFt (cells in middle panel). Given that BOLD appears to be dominated by synaptic input rather than spiking output (Logothetis 2003), inputs (as illustrated by the gradients) would be expected to dominate the BOLD signals. Panel (c) is our own illustration of the schema in Shipp (2007) after modifying the gradients in PFt from BA44 by adding input to layer III based on Gerbella et al., 2011, and by aligning all layer thicknesses to those in PFt for readability although they vary across the brain regions.

(d) Predictions about the depth profile of stimulus locked information (as measured using ISC) in PFt for Intact and Scrambled conditions based on (b) and (c). Intact conditions should be dominated by feed-back from BA44 triggering high ISC in layers III, V, VI (blue gradients in left column), while Scrambled conditions should show less feedback from BA44 (weaker blue gradients in the middle column). Accordingly, we expect the contrast between the ISC in the two conditions to be particularly clear in the deep layers (V/VI) and layer III with more ISC for the intact sequences (yellow gradient in the right column). This feedback information includes sequence level information unfolding over the tens of seconds of the entire movie (approximately between 0.1 and 0.015 Hz) and is thus slow enough to be well within the Nyquist limit of our fMRI sampling (TR=4.2 s or ~0.24 Hz). In contrast, the feed-forward information representing prediction errors at each camera change (i.e. every 2 s on average), which should terminate in layer IV, and be stronger in the Scrambled condition, is too fast for its fluctuations to be accurately captures at our slower acquisition speed. Because intersubject functional connectivity depends on measuring the fluctuations over time of the signal rather than average BOLD activation, we hypothesize this difference in layer IV to be difficult to measure using the methods we employ(Nastase et al. 2019)..

(e) Predictions about the different (Intact vs Scrambled) functional connectivity between PFt and BA44 and LatOcc - measured using ISFC - based on the hypothesis described in (b) and (c): we expect PFt to show increased connectivity with BA44 for Intact > Scrambled with a feedback pattern of connectivity, while we expect it to show increased connectivity with LatOcc for Scrambled > Intact in accordance with a feedforward pattern of connectivity.

To summarize these predictions from the point of view of PFt, we thus have the following expectations (Fig. 1b): because participants do not know what sequence will be presented, PFt should initially receive visual input about the first acts of a sequence, an input that then rapidly wanes as PFt increasingly suppresses expected visual input. This input is dominated by the high frequency signals (T≈2s) of the individual motor acts. PFt should also receive predictions from BA44 in a broader frequency range, that include short term predictions about the unfolding of the current act (T≈2 s) to long term predictions about the entire sequence (T≈1 min). These BA44 feed-back predictions should remain high throughout the sequence, and dominate the overall signal during the Intact sequences. In contrast, for the Scrambled sequences, predictions cannot suppress the randomized presentation of the acts and PFt should thus receive sustained feed-forward prediction error input from visual regions at each camera change. Predictions from BA44 are likely to subside as participants realize that the sequence is not predictable, leading BA44 feedback to wane, so that the BOLD signal overall should be dominated by the feed-forward signal. Because the BOLD signal appears dominated by synaptic input over spiking output (Logothetis 2003), we neglect what signals PFt is likely to output to other regions in these considerations. Considering what we know about the laminar distribution of feed-forward and feed-back inputs from a variety of sensory brain regions organized in hierarchical fashion (Fig 1c, adapted from(Shipp 2007)), feed-forward input from visual regions should terminate in middle layers of PFt (layer IV) and feed-back input from BA44 should terminate mainly in deep layers of PFt (layers V/VI). However, monkey tracer studies have shown that the connections from F5a (the likely monkey homologue of BA44) to parietal regions similar to human PFt terminate both in deep layers V/VI and in layer III (Gerbella et al. 2011). Such deviation from the prototypical feedback connectivity pattern described by Felleman and Van Essen in the visual system (Felleman and Van Essen 1991) is indeed not usual for fronto-parietal connectivity across association areas (Rozzi et al. 2006; Gerbella et al. 2010, 2011). We can thus reformulate these predictions about the direction of information flow as predictions about the depth at which stimulus-locked information should dominate the BOLD signal in the Intact and Scrambled conditions if our model of predictive coding is true (Fig. 1d). With regard to feed-back information flow, which should be dominant in PFt in deep layers (layer V/VI) and in layer III, we expect more for Intact than Scrambled conditions, so that the contrast Intact-Scrambled should show increased ISC in layers III, V and VI. This feed-back information flow contains information at the sequence level, that fluctuates over tens of seconds, and our relatively slow 7T acquisition (TR=4.1s, Nyquist limit T=8.2s) should thus be adequate to capture its fluctuations over time, which is the basis for intersubject correlation analyses(Nastase et al. 2019). In contrast, the feed-forward information flow, that caries prediction errors, would peak at every camera change (T≈2s) in the Scrambled condition, and fluctuate faster than the 7T fMRI acquisitions we will use in this study (TR=4.1s, Nyquist limit T=8.2s). There might thus be higher *average* BOLD activity over the duration of the movies in the Scrambled condition in layer IV, where feed-forward should terminate, but the ISC method we use does not investigate average BOLD activity, but the synchronization of activity fluctuations between participants(Nastase et al. 2019). However, we expect to fail to measure the second by second fluctuations of such prediction errors accurately enough for ISC to evidence increased synchronization in layer IV for the Scrambled condition. Accordingly, for ISC, the most robust observable effect derived from the difference across conditions should be the increased stimulus-related information in deep layers (V/VI) and layer III for the Intact sequence. Such an increase is not predicted by traditional models of action observation that do not emphasize the role of feed-back connections, and can thus help adjudicate between these views.

Pioneering studies in the early visual cortex have shown that acquiring fMRI at submillimeter resolution at high-field strength (≥7T) can indeed test predictive coding hypotheses (Kok et al. 2016; Lawrence et al. 2018, 2019a, b; Finn et al. 2020; Aitken et al. 2020). Recent work in particular confirms that activations in deep layers is indeed measurable when feedback predictions about upcoming sensory events are likely to occur in the visual system (Kok et al. 2016; de Lange et al. 2018; Aitken et al. 2020), while activation in superficial layers seem to be measured when there is a mismatch between predicted and observed input and attentional signals (de Lange et al. 2018). Unfortunately, experiments testing whether deep layers carry signals that could be related to predictions have so far been restricted to early visual cortices and we do not know whether the same might be observed in parietal and premotor regions (Finn et al. 2020). Therefore here we acquired brain activity at 7T with an isotropic resolution of 0.8 mm in 9 participants, planning a partial-volume acquisition to include PFt, BA44 and LatOcc cortex of the left hemisphere (Supplementary Figure 1), while they observed intact and scrambled actions to test the validity of predictive coding models of action observation, specifically by testing for increased stimulus-dependent information in deep layers of PFt when seeing intact vs scrambled sequences. We choose PFt as our main region of interest because it is at the crossroad between feed-forward input representing prediction errors and feed-back information from BA44, and because it has a layering that is somewhat similar to the better explored visual system, while BA44 and its dysgranular organization (Amunts et al. 1999; Amunts and Zilles 2012) would make the same approach more tentative.

Methodologically, as in our previous 3T work (Thomas et al. 2018), we use intersubject correlation (ISC) to localize sequence-level information, and intersubject functional correlation (ISFC) to quantify information sharing across nodes in this study for two main reasons. First, to generate sequence level brain activity, we need longer sequences of actions (e.g. preparing breakfast) in the order of ~1min. Such stimuli are too long for traditional block-design fMRI experiments, but are ideal for ISC analyses (Nastase et al. 2019). Second, quantifying functional connectivity between brain regions within subjects is complicated by the fact that a sizable proportion of the BOLD signals are formed by noise that is often shared across brain regions (e.g. respiration and motion (Liu 2016)). That two brain regions synchronize their BOLD signal can thus be due to true exchange of stimulus relevant information across those regions or to the presence of common noise. Intersubject functional correlation (ISFC) provides a way to circumvent the connectivity-measure-inflating effect of such common noise across brain regions (Simony et al. 2016; Nastase et al. 2019) by leveraging the fact that BOLD signals in each voxel of the brain can be seen as a combination of (i) stimulus locked signals that represent stimulus-relevant information and (ii) signals unrelated to the stimulus (e.g. daydreaming and noise) that will not be stimulus locked. By correlating the signal in region PFt of one participant with the average time course of all other participants in region BA44 or LatOcc, ISFC can then isolate stimulus-locked information shared across these regions, because noise or stimulus unrelated thoughts would be unlikely to occur systematically at the same time in different participants, and therefore would average out and not correlate in time across participants. Here we will thus leverage ISC to test the presence of more ISC in deep layers of PFt for intact over scrambled sequences and ISFC to test that stimulus related information in deep layers of PFt is more shared with BA44 during intact vs scrambled sequences. Finally we can also leverage ISFC to explore whether information in the middle layers of PFt may indeed be more shared with LatOcc regions during scrambled sequences, although we hypothesize that this analysis will be limited by our ability to detect fast stimulus-dependent fluctuation of the BOLD signal in LatOcc due to our fMRI sampling frequency.

Table 1: List of sequences used as stimuli with total duration in seconds and number of motor acts.

	Action	Seconds	Acts
1	Inflating and tying a balloon.	51	27
2	Making a paper boat.	94	32
3	Preparing bread with butter and jam.	79	40
4	Sewing a button.	66	42
5	Writing a gift card.	83	39
6	Rolling a cigarette.	72	30
7	Arranging flowers in a vase.	82	39
8	Framing a picture.	112	39
9	Sending a letter.	42	34
10	Replacing battery in a torch.	51	27
11	Applying nail polish.	49	23
12	Squeezing oranges.	62	40
13	Sharpening a pencil.	83	44
14	Replacing a pillow cover.	44	35
15	Removing nail polish.	64	32
16	Preparing a sandwich.	77	27
17	Toasting bread.	65	30
18	Folding a shirt.	38	20

Methods

Participants

Data were acquired from 14 subjects (10 males, 4 females, Age mean∓std = 24.9∓2.8, range=21-30). Three subjects were excluded due to missing data and two additional subjects were removed due to excessive head motion. Therefore 9 subjects (6 males, 3 females, Age mean∓std= 24.8∓2.6, range=21-30) were included in our analyses. Upon arrival the subjects were screened for MRI compatibility and were familiarized with the MRI setup. All subjects reported being right handed and having no psychological or neurological disorders. All of them had a normal or corrected to normal vision. Subjects were compensated 10 Euros/hour. The study was approved by the Lab Ethic Review Board of University of Amsterdam (2016-BC-6837) and all subjects signed the informed consent before the beginning of the experimental session.

Stimuli

Eighteen movies containing different daily actions (e.g. preparing sandwiches with butter and jam; see Table 1 for the full list) were simultaneously recorded by two video cameras (Sony MC50, 29 frames/s) at an angle of 45 degrees. The videos were edited using ADOBE Premiere ProCS5 running on Windows. Each movie was subdivided into shots containing one meaningful motor act each (e.g. taking bread, opening the butter dish, scooping butter with a knife, etc.). This was done on recordings from both camera angles. These motor acts (mean/standard deviation duration 2s ± 1s) were then assembled to build two types of ~1minute long stimuli (Fig. 1).

For the *Intact* (I) presentation, the natural temporal sequence in which the acts were recorded was maintained, but a camera angle change was introduced between every two consecutive acts by alternate sampling from the recordings of the two cameras. In the *Scrambled* (S) versions, the acts remained the same, but the order of the acts was randomly re-arranged, and a camera angle change was introduced between every two consecutive acts. Camera angle changes were imposed at each act transition in both types of movies to compensate for the visual transients that would otherwise be present only in the scrambled movies.

This resulted in 18 intact and 18 scrambled movies. These movies were presented to the subjects in 4 blocks. Block 1 (220 fMRI scans), 2 (160 fMRI scans) and 3 (190 fMRI scans) each consisted of 5 unique intact movies and the corresponding 5 scrambled movies. Block 4 (145 fMRI scans) consisted of the remaining 3 intact movies and the 3 corresponding scrambled movies. The stimuli in each block were separated by inter-movie interval varying from 8 to 12 s. The unique combination of videos in each block was kept constant across participants to ensure that each subject always viewed the scrambled version of a movie before its intact version.

Experimental procedure

Each subject was invited twice, around one week apart (Mean=5 days, Range = 2 -10 days). On the first day, they viewed blocks 1 and 2 of Intact and Scrambled movies, in two separate but consecutive 3DEPI fMRI data acquisitions. This was followed by the acquisition of anatomical images

(MP2RAGE and T1-3DEPI). Subjects then viewed the same 2 blocks of the movies again while we collected new 3DEPI images. The order of the blocks was pseudo-randomized across subjects. During the second day, subjects saw blocks 3 and 4 of Intact movies and their temporally Scrambled versions. This was followed by a T1-3DEPI image acquisition followed by repetition of the blocks. Similar to the first session, blocks were randomized across subjects. Videos were presented using the Presentation software (Neurobehavioral Systems, Inc., Albany, CA, USA). No behavioural response was required during the four sessions, but participants were instructed to carefully observe the videos. We ensured that subjects were not in the scanner for more than 90 minutes during each session.

Image acquisition

All MRI images were acquired on a Philips 7T Achieva at the Spinoza Center in Amsterdam (https://www.spinozacentre.nl/) using a 32-channel receive and 2-channel transmit volume coil (Nova Medical, USA). On each experimental day, we acquired 4 runs of partial fMRI brain images, and one anatomical ($T_1$23DEPI) image, with the same field of view and in-plane resolution of the 3D EPI fMRI images. Images were centered on the inferior posterior parietal lobe (rectangle in Figure 2A), but also covered lateral portions of the temporal, occipital and inferior frontal lobe. We manually adjusted the bounding box to approximately cover the same brain regions in all subjects. On the first day only, we additionally acquired one high-resolution whole-brain anatomical image (MP2RAGE). All images were acquired in sagittal orientation.

fMRI acquisition: 3D EPI sequence at 0.8 x 0.8 x 0.8 mm resolution, FOV 169 x 150, 30 slices, reconstructed at 0.75 mm, each 3D volume acquired in 4.11 s, EPI factor = 27, SENSE(AP) = 3.9, flip angle = 18º, TR/TE = 54/28 ms.

partial FOV anatomical acquisition: $T_1$23DEPI (van der Zwaag et al. 2018a, b) sequence at 0.8 x 0.8 x 0.8 mm resolution, FOV 169 x 150, 30 slices, reconstructed at 0.75 mm, SENSE(AP) = 3.9, inversion time (inv1/inv2) 0.8/2.7 s, flip angle 20/16º, TR/TE = 52/25 ms.

whole-brain anatomical acquisition: MP2RAGE (Marques et al. 2010) sequence at 0.64 x 0.64 x 0.64 mm resolution, FOV 220 x 220, 256 slices, reconstructed at 0.625 mm, inversion time (inv1/inv2) 0.8/2.7 s, flip angle 7/2º, TR/TE = 6.2/2.2 ms.

Image preprocessing

In our preprocessing pipeline, we aimed at preserving the ability to detect the cortical depth of the BOLD signal by minimizing the amount of interpolation steps for the fMRI images, and planned the other necessary preprocessing steps involving anatomical images around this aim. Specifically, the native space of the fMRI images was kept as the reference for all subject-level and group-level inter-modal registration and analysis, an approach also adopted by previous studies (Huber et al. 2017) to preserve the layer-specific information in high-resolution fMRI images.

The whole preprocessing pipeline can be accessed at the official github repository for this study (https://github.com/ldeangelisphys/layerfMRI). The scripts (in bash, Python, R) employ routines from a variety of MRI image processing packages: FSL (https://fsl.fMRIb.ox.ac.uk/fsl), nighres (https://nighres.readthedocs.io/) ANTs (https://stnava.github.io/ANTs/), ITK-SNAP and associated command-line utilities (http://www.itksnap.org/).

Image reconstruction

The PAR/REC stacks exported from the Philips scanner were converted to NIFTI images using dcm2niix v1.0.20201102 (https://github.com/rordenlab/dcm2niix) and reoriented to the standard RPI coordinate system using `fslorient2std`, to enable inspecting the images in a familiar orientation. Facial information was removed from whole-brain anatomical images using pydeface (https://github.com/poldracklab/pydeface). Finally, the T1w and the quantitative T1 map images were reconstructed from the two inversions of the MP2RAGE acquisitions using the nighres `nighres.intensity.mp2rage_t1_mapping` function and the appropriate acquisition parameters. All the other parameters of the functions were left at the default values. The T1 images were subsequently used to estimate the cortical depth and the registration with the fMRI images.

Skull stripping

Removal of the skull was necessary to carry out image registration across MRI modalities within subject and with the MNI template. A precise extraction of the cortical sheet was instead carried out later (see below "Cortical depth estimation"). For the full anatomical acquisition, we obtained the best results by first correcting the second inversion image of the MP2RAGE (inv2) for bias field using ANTs' `N4BiasFieldCorrection`, then feeding this and the quantitative T1 map into the nighres' `nighres.brain.intensity_based_skullstripping` function to carry out skull stripping. In the partial brain anatomical images, we carried out the bias field correction on the mean inv2 acquisition for each participant, and then thresholded this image to an intensity of 1.5e5. In the partial brain fMRI images, we thresholded the intensity-normalized mean fMRI image to either 0.05 or 0.1 (2/9 and 7/9 participants, respectively), and then generated the brain mask using `ants.get_mask` with `cleanup` option set to 1 (erosion with a radius of 2 voxels). These values of intensity threshold were manually chosen based on visual inspection of the skull stripping, after observing that the default parameters would produce a suboptimal result (i.e. either removing brain tissue or not identifying voxels in the skull for removal) which would have had consequences during image registration.

FMRI image preprocessing

FMRI images underwent a minimal preprocessing pipeline, which consisted of brain extraction, motion correction to the middle image of each run using FSL MCFLIRT and detrending using a gaussian-weighted line as implemented in FSL FEAT. To preserve the cortical depth specificity of the fMRI signal, we did not apply spatial smoothing to the images that were used for the main analysis related to the cortical depth. Instead we applied a 6mm FWHM spatial smoothing only for the initial

group-level analysis which served to identify the regions of interest later used in the cortical depth analysis. Details are provided in the Analysis section.

Registration

All registrations between anatomical (T1w) images were carried out using ANTs' SyN algorithm (symmetric diffeomorphic registration preceded by an affine transformation) with mutual information as optimization metric (Avants et al. 2008, 2011; Klein et al. 2009) as implemented in ANTsPy. For the registration between fMRI and partial FOV anatomical image, we used the provided version of the SyN algorithm optimized for BOLD images (SyNBoldAff). The registration between the partial anatomical image (T1w) and the fMRI image was estimated using the mean fMRI image of each run and the partial anatomical image acquired in the same session. Additionally to these within-subject registrations, we also estimated the registration between the full anatomical image and the standard MNI (the 1 mm isotropic `icbm152_2009_brain.nii.gz` provided in nilearn - https://nilearn.github.io/) using the same SyN algorithm and optimization metric.

Cortical depth estimation

The estimation of the cortical depth from the high-resolution whole-brain anatomical images was carried out with the functions provided by nighres. This complex procedure involves several steps which have been extensively detailed in previous works (Bazin et al. 2014; Huntenburg et al. 2018). Here we briefly describe the pipeline adopted for our images.

The skull stripped version of the inv2 and quantitative T1 map were obtained using the brain mask from the previous skull stripping of the bias field corrected inv2 image. We then estimated two probability maps: one for the dura mater (based on the inv2 image), one for the sulcal ridges (based on the T1 map). Together with the initial brain mask, these additional maps provided further information for carrying out brain segmentation (Bazin et al. 2018) of the whole-brain anatomical images.

Brain segmentation was then carried out using a Multi-compartment Geometric Deformable Model (MGDM) (Fan et al. 2008; Bogovic et al. 2013) implemented in the `nighres.brain.mgdm_segmentation` function. The inputs to the function were the skull-stripped T1 map and a stack containing the dura and sulcal probabilistic maps. Since this procedure is based on labelled atlas priors (provided in nighres), the GM output map refers only to the cortex and excludes deep brain nuclei. The boundaries of the cortex are then extracted based on the results of the segmentation using a single object Geometric Deformable Model, and the cortical ribbon is reconstructed using the CRUISE algorithm (Han et al. 2004).

The CRUISE reconstruction results in the definition of the boundaries between GM/WM and GM/CSF, which are then used to estimate cortical depth using a volume-preserving method (Waehnert et al. 2014, 2016; Trampel et al. 2019). Per the defaults of nighres, cortical depth values - ranging from 0 to 1 - are presented in the following results as higher towards the pial surface.

Analytic strategy

We conducted our fMRI analysis in the framework of inter-subject correlation (ISC - Hasson et al. 2004) and the related inter-subject functional connectivity (ISFC) (Simony et al. 2016; Nastase et al. 2019). ISC is a relatively novel analytical paradigm where the significance of a brain region for a certain task is determined by estimating the similarity - using Pearson correlation - of the brain activity between different participants involved in the same task over time - e.g. while watching the same movie - rather than by regressing a hemodynamic model of the task on brain activity.

Generally, ISC is carried out at the voxel level after alignment of the entire 4D fMRI volume in a common space like the MNI. This however represented a limitation for the setting of our experiment, since the spatial interpolations required to move the native fMRI data into a common space (coregistration) decrease the spatial resolution of the BOLD signal. In order to benefit from the ISC analytic strategy in our data, while at the same time retaining the potential to resolve the fMRI signal as a function of cortical depth, we devised a new procedure that we denominate "dual-ISC", which allows an ISC analysis to be carried out on the fMRI data in native space.

The **first ISC** is carried out in MNI space after smoothing the fMRI data. The results from this ISC are used to identify the locations on the cortex where brain activity is significantly consistent across subjects during the observation of either Intact or Scrambled movies. The group-level binary masks of the clusters identified in this analysis are then transformed into the single subject native space, where they are used to estimate an average time course for each cluster using the original, non-smoothed data. Crucially, within each cluster the time course averaging is carried out separately for voxels at different cortical depths. This allows to carry out a **second ISC**, where the unit of analysis is not each single voxel's fMRI time course, but rather the averaged time course for all the voxels at a certain cortical depth in each of the clusters of activity. Similar to spatial smoothing in conventional-resolution fMRI, the averaging also provides a way of denoising the fMRI signal before conducting the analyses. We will now describe the steps of this procedure in detail.

First ISC - Localizer

The preprocessed fMRI images of each run for each subject were spatially smoothed with a gaussian kernel of 6 mm FWHM (8 times the original voxel size). The previously estimated transformations (fMRI to partial T1w, partial to full T1w, full T1w to MNI) were aggregated into one single transformation using `ants.apply_transforms`, to limit the amount of interpolation steps. This composite transformation was applied to the 4D images using linear interpolation. This allowed us to carry out the first standard voxelwise ISC analysis to define a group-level localizer of brain activity associated with the observation of intact or scrambled movies, without biasing the analysis for one of the others.

We then used the registered 4D images to perform an InterSubject Correlation (ISC) analysis. This approach requires correlating the time courses recorded in every voxel across different subjects to identify those that are synchronized across subjects, therefore reflecting shared stimulus-dependent fluctuations across subjects (Nastase et al. 2019). We use ISC to determine which regions are involved in the processing of either intact or scrambled clips that we show to each participant.

The ordering and timing at which the clips are shown was randomized across participants. Before carrying the intersubject correlation, we then extract the segment (i.e. series of volumes) corresponding to each clip, we standardized each segment (removed mean and divided by the standard deviation) and concatenate them following the same order for every participant. This generated one 4D dataset containing all the data segments corresponding to intact movies concatenated together in a cardinal order and one containing all the scrambled movies. Note that although for both conditions each movie was displayed twice, such repetitions are treated as two independent movies and concatenated accordingly.

Following this processing of the data, we ran ISC in a leave-one-out approach. This procedure consists in correlating the time course of a voxel for one subject with the average time course corresponding to all the other subjects (Nastase et al. 2019). We then use a non parametric approach (bootstrap) to identify the brain regions with a significant ISC (Chen et al. 2016). Given the relatively low sample size (9 subjects), we also perform a traditional parametric t-test, which is predicted to produce less false positives than non-parametric metrics in the case of limited sample sizes (De Angelis et al. 2021). Having observed that both thresholding methods yield consistent results, we threshold the ISC map with the bootstrap method, with a p-threshold of $\alpha = 0.001$ and a minimum cluster size of $k = 50$.

Having observed that both thresholding methods yield consistent results, we estimated the significance of median ISC values using bootstrap (Chen et al. 2016) across 5000 subject-wise resampling (with replacement). ISC values were considered significant if they lied outside the mid 99.95% (equivalent to a two-tailed $p \leq 0.001$) of the bootstrap distribution, after adjusting (i.e. shifting) the distribution for the median ISC of the actual sample. We then apply the cluster size filter with a custom made python code based on the sklearn-image library (cluster connectivity = 2, as in SPM). This result was used to produce a binary mask in MNI space encompassing the brain locations where a significant degree of synchronized activity was found across participants during the observation of either intact or scrambled movies representing goal-directed actions.

Second ISC - Cortical depth specific

The localizer in MNI space was taken into the native fMRI space of each run for each subject using the inverse of the previously estimated composite transformation from fMRI to MNI space. Concerning this and all other transformation to the native space, since we were transforming volumes containing labels (either significantly active regions from ISC1, atlas regions or cortical depth) the information of which should be preserved as such (i.e. integers), we used a nearest neighbour interpolation (`genericLabel` option in ANTsPy, recommended by ANTs team for images of this kind).

As detailed in the Introduction, we had a strong anatomical hypothesis about the location of the regions where we expected to observe a cortical-depth specific effect of observing intact vs. scrambled goal-directed actions. Specifically, we focused our investigation of the inferior posterior parietal lobe on area PFt (Caspers et al. 2006), of the ventral premotor cortex to the posterior inferior frontal gyrus Brodmann area 44 (Amunts et al. 1999) and the higher-order lateral visual cortices. We defined our regions of interest using a logical AND between the functional localizer obtained by the first ISC and the cytoarchitectonic maps from Juelich (Eickhoff et al. 2005) for PFt and BA44.

Since these cytoarchitectonic maps do not provide labels in the lateral occipital lobe, we used in this case the lateral occipital cortex in the Harvard-Oxford atlas (https://fsl.fMRIb.ox.ac.uk/fsl/fslwiki/Atlases), which provided a good approximation to the cluster yielded by the localizer ISC in the lateral occipital lobe, and did not include any other cluster of activity identified in the localizer ISC (See Fig. 2a). The Juelich maps and the lateral occipital cluster were transformed from the MNI into the native fMRI space using the same procedure used for the localizer maps. The final regions of interest were computed as the conjunction between the localizer and the Juelich maps. Additionally, the cortical depth maps estimated in each subject's whole-brain MP2RAGE scan were taken into the native space using the same procedure.

Using these ROIs, we extracted the time courses to feed into the second ISC. The time courses were extracted from the fMRI data in native space after basic preprocessing (motion correction + trend removal and no spatial smoothing). For each region of interest in PFt, BA44 and the latOCC region, the encompassed voxels were binned according to their cortical depth. For each cortical depth bin, we concatenated the segments of the time course containing movie stimuli, transformed the intensity values in Z-scores and averaged the resulting values across voxels (as in the first ISC). This resulted in a time course containing the brain activity elicited by the movies for each cortical depth bin, each region of interest and each subject.

Given the nominal resolution of our data (0.8 mm), it is not possible to bin the cortical depth to selectively isolate the fMRI signal specific to each cortical layer. However for our experiment we did not aim to estimate the specific activity of each layer, but rather to estimate differences between intact and scrambled movies in deep/superficial with respect to middle layers. For this reason, rather than settling on a specific number of cortical depth bins, we carried out the second ISC analysis after sampling the cortical depth in 6,8 and 10 bins, and verified the consistency across the results. Correction for multiple comparisons (using false discovery rate - FDR) was adapted to the choice of number of bins. We did not explore binnings lower than 6 since this would have likely removed any possibility of identifying signal potentially associated with layer IV. Inference on the difference between the ISC values for Intact and Scrambled in each cortical depth bin and for each region of interest was carried out using a paired t-test.

ISFC

In addition to the ISC, we performed a depth-resolved inter-subject functional connectivity (ISFC) analysis. In this analysis, we compute the functional connectivity across each bin (1 to 6) and each region of interest (PF, BA44 and latOCC). We compute this functional connectivity with a leave-one-out ISFC approach. Precisely, this procedure consists in computing the correlation of the time course associated to bin_j in the region of interest JU_i for subject s, with the leave-one-out average time course associated to bin_l in the region of interest JU_k (averaged over all subjects except s). As a result, we have for each combination of bin and JU and for each subject s an estimate of the intersubject functional connectivity between bin_j in JU_i and bin_l in JU_k:

$$ISFC\,[JU_i\,bin_j;\ JU_k\,bin_l] \;=\; \frac{1}{N}\sum_s \rho[\,tc_s\,(JU_i\,bin_j)\,,\ \overline{tc_{N-s}}\,(JU_k\,bin_l)\,]$$

where $tc_s(JU_i\, bin_j)$ is the time course associated bin_j in JU_i for subject s and ρ is a Pearson correlation.

We transformed such correlation values to z scores (using fisher-z transform) and used them as input for a one sample t-test to assess whether each condition exhibits non-null functional connectivity across these regions. Additionally, we perform a paired sample t-test between the intact and scrambled condition, to determine whether the specific condition has an impact on the functional connectivity that we determine with ISFC.

Alongside the ISFC, we carry a more traditional PPI-like functional connectivity analysis (Friston et al. 1997). In this approach, we perform all the correlations in a within-subject fashion and subsequently compute statistics across subjects, i.e.,

$$FC\,[JU_i\, bin_j;\, JU_k\, bin_l] \;=\; \frac{1}{N}\sum_s \rho[\, tc_s(JU_i\, bin_j)\,,\, tc_s(JU_k\, bin_l)\,]$$

where $tc_s(JU_i\, bin_j)$ is the time course associated bin_j in JU_i for subject s and ρ is a Pearson correlation. Inference on the difference between the ISFC values for Intact and Scrambled in each cortical depth bin and for each region of interest was carried out using a paired t-test.

Results

Premotor, parietal and lateral occipital cortices show significant ISC for intact or scrambled sequences

First, we aimed to macroscopically localize premotor, parietal and posterior temporal/lateral occipital patches of interest with activity that carries movie-related information in intact or scrambled movies. To do so, we transformed the preprocessed 4D fMRI data into the MNI space, smoothed the BOLD signal, and calculated the leave-one-out ISC for each voxel separately for intact and scrambled sequences. This revealed the expected network including the lateral premotor cortex and BA44, the primary somatosensory cortex and the inferior parietal lobule (including PFt) as well as lateral extrastriate occipital regions (latOCC) (Fig. 2). This network resembles that found using the same stimuli in conventional resolution 3T fMRI (Thomas et al. 2018). We then identified voxels containing significant ISC for the intact or scrambled movie in PFt, BA44 and latOCC as regions of interest, to then resolve the ISC as a function of cortical depth. By selecting them with a logical OR between the results of the ISC carried out either on intact or scrambled movies (i.e. taking all voxels surviving the threshold for $ISC_{intact} > 0$, or for $ISC_{scrambled} > 0$), we avoid biasing the regions of interest towards intact or scrambled movies in further analyses (Kriegeskorte et al. 2009).

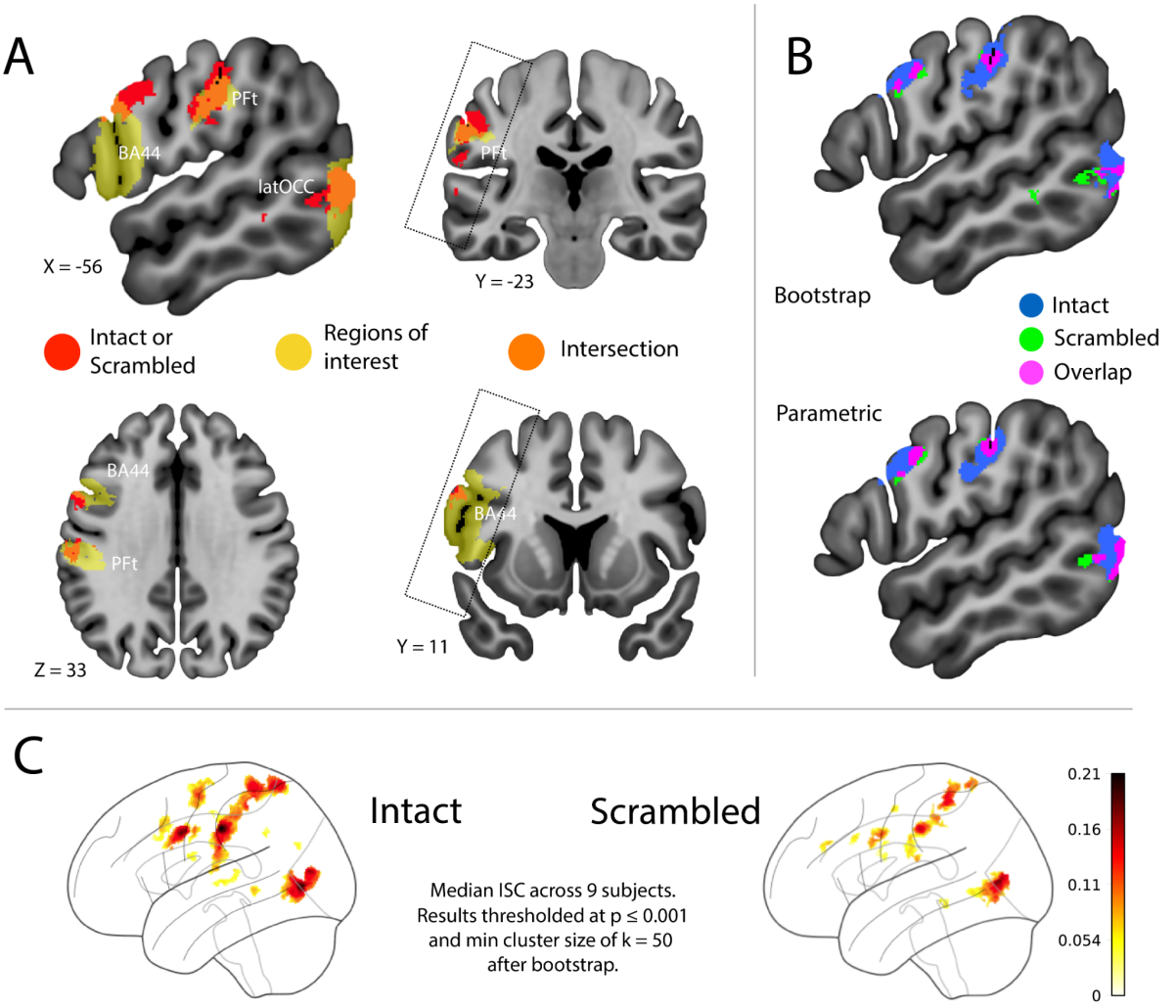

Fig 2 Results of the Localizer-ISC. The first ISC was carried out in MNI space, to define patches of interest with brain activity relating to hand action observation during either Intact or Scrambled movies, to be later used in the second, depth-resolved ISC in native space. **A.** Results (significant voxels) for a logical OR between the contrasts $ISC_{Intact} > 0$ *or* $ISC_{Scrambled} > 0$ (red and orange), overlaid onto the initial regions of interest (yellow): BA44 and PFt from the Anatomy toolbox, lateral occipital (latOCC) from the Harvard-Oxford cortical atlas. The latOCC was used as an anatomical region that corresponds to the posterior mid temporal gyrus clusters often observed in previous studies of action observation. Significantly active voxels co-located with voxels in the regions of interest are denoted in orange. The rectangle crossing the coronal images approximates the field of view common to our participants afforded by our scans. Lack of ISC outside of this box should and cannot be interpreted as suggesting that these regions do not show significant ISC. The exact extent of the field of view across subjects can be found in Fig S1 **B.** Comparison of the results obtained using bootstrap (p_{unc}<0.001, k=50 minimum voxels per cluster) and parametric inference (p_{unc}<0.001, k=50 minimum voxels per cluster). Scrambled activity was overall associated with smaller clusters mostly encompassed in regions activated by Intact movies. **C.** Median ISC parameter estimates across 9 subjects. Significance was established according to the null distribution of median ISC values obtained using bootstrap. Median ISC values lying beyond the 99.99th percentile of the null distribution (i.e. p ≤ 0.001) were considered significant. In addition, a minimum cluster size of k=50 voxel was applied to determine significance.

PFt shows increased stimulus related information in deep layers for intact > scrambled sequences

The second ISC analysis aimed at resolving the cortical depth of the fMRI BOLD signal synchronization in PFt for Intact vs Scrambled movies. Therefore, this analysis was carried out using the preprocessed but unsmoothed data in the native fMRI space of each participant. For each subject an average time course was calculated across all voxels within each depth-bin in PFt. Then, a leave-one-out ISC was carried out for each bin, i.e. the signal of each subject *i* in a depth-bin was correlated with the average signal of all other subjects in the same depth bin of the same ROI.

The results of this depth-resolved ISC for PFt are displayed in Figure 3a and show how the difference in stimulus locked signals between the observation of Intact and Scrambled movies depends on the cortical depth. As predicted (Fig. 1), during the observation of Intact movies, the synchronization of the activity in deep layers (at depths aligning with layers V/VI) of PFt significantly increased. This increase extended into middle depths, aligning with layer III, in line with the tracer studies in monkeys that suggest that feedback from BA44 to PFt also terminates in layer III. For completeness, although we did not have specific hypotheses for ISC in BA44 and LatOcc, we also present the differential ISC as a function of depths for these regions in Supplementary Fig. S2. For BA44, the activity was significantly more synchronized in the superficial layers for the intact, and in the deepest layers for the scrambled movies (See Supplementary Fig S2). Similar results are obtained when the cortical depth was sampled in either 8 or 10 bins (See Supplementary Fig. S3). No significant changes were observed in our latOcc patch of interest. Examining the individual contrast values in PFt (Fig. 3b) shows that the individual participants showed relatively consistent effects in our sample. Finally, examining the actual ISC values in the two conditions (Fig. 3c and Supplementary Figure S2) confirms the general trend found in the literature, that ISC is highest in earlier sensory regions (latOcc peak average ISC around r = 0.3), lower in parietal cortices (PFt peak average ISC around r = 0.2), and lowest in frontal cortices (BA44 peak average ISC around r = 0.1).

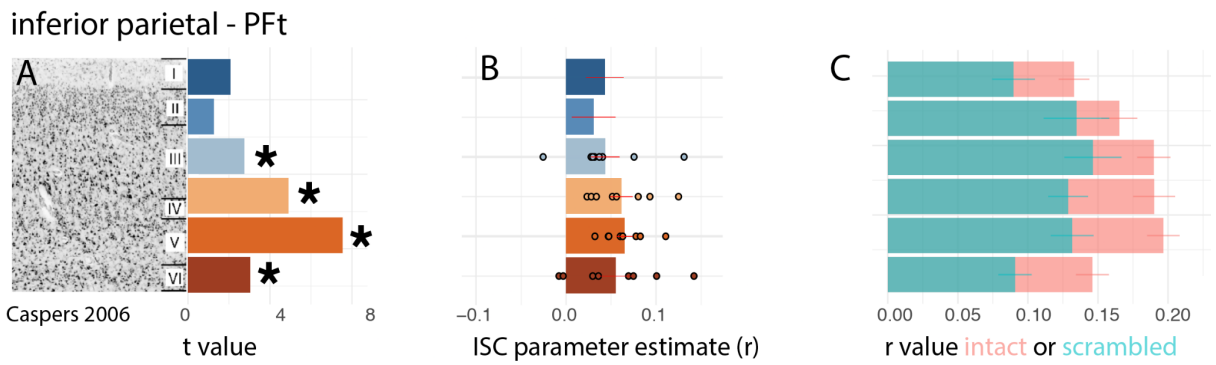

Fig 3 - Depth-resolved ISC for the contrast Intact - Scrambled movies in PFt. (a) t-values of the contrast intact > scrambled in PFt as a function of depth bin. T-values derive from a matched pair t-test of the r→z transformed leave-one-out ISC values per bin. Significant t-values after FDR correction (q(FDR) = 0.05) across the 6 bins are indicated with an asterix. Results are shown next to a histological reconstruction of the 6 cortical layers of PFt in postmortem brains from Caspers et al. (Caspers et al. 2006) for visual guidance only. The consistency between these results and those obtained with 8 or 10 bins is shown in Supplementary Figure S3. (b) The difference in leave-one-out r-value between intact and scrambled movies as a function of depth in PFt. Results were generated by Fisher r→z transformation before averaging and calculating s.e.m., and then back-transformed into r-values for illustration. For bins with significant differences in (a) we show the value of each

participant. Error bars are s.e.m. (c) r-value of the leave-one-out ISC as a function of bin for the intact (pink) and scrambled (green) condition separately. Stars are from panel (a).

Functional connectivity is increased in deep voxels across PFt and BA44

After we found evidence for differential ISC values for Intact and Scrambled movies at different cortical depth within PFt, we probed the interaction between PFt and either BA44 or latOCC in the two conditions. We hypothesized that during the observation of Intact movies, brain activity would be more enhanced in the feed-back stream from BA44 to PFt, terminating at depth aligning with layers V/VI and III (Gerbella et al. 2011), with the deep-layers particularly associated with predictive brain activity (Shipp 2007; Kok et al. 2016; de Lange et al. 2018; Aitken et al. 2020). In contrast, for the information transfer across latOCC and PFt, because the latOCC appears to represent information at the short temporal scale of acts, too fast for the the Nyquist limit of our acquisition, we do not expect measurable changes in ISFC. Had we had higher temporal resolution, we might have expected the scrambled condition in particular, to increase ISFC from latOCC towards the intermediate depth of PFt aligning with the granular layers.

Fig. 4 (left column) represents the results of the ISFC. In accordance with our hypothesis about predictive information being most visible in deep-layers, the information transfer, as assessed using ISFC, between PFt and BA44 is increased for intact compared to scrambled conditions at greater depth, in bins that align with layers V and VI of PFt and BA44 (Fig4a). In addition, we also find increased ISFC in depth-bins that align with layer III in PFt. In both regions, the increase of ISFC also extends to depths that align with the relatively thin layer IV. In PFt, we also observe a small but significant increase of ISFC between the most superficial depth of PFt and the deepest layers of BA44, as might be expected based on the pattern of feed-back connections described in the anatomical literature for other brain regions (Fig. 1c,(Shipp 2007)). In line with the idea that the additional exchange of information between PFt and latOCC conveying prediction errors in the scrambled condition would occur at a time-scale too fast for our acquisition to capture, we did not find a significant contrast in ISFC across conditions between these regions. Examining the ISFC values in the individual conditions for PFt x BA44 (Fig. 4 left middle and bottom row) confirms that although some significant ISFC occurs in all conditions, the intact condition stands out with more consistent ISFC in the intact condition.

To illustrate the advantage of ISFC over traditional within-subject functional connectivity, we used the same data, but calculated the correlation between the time course of one area and that of another within each participant, rather than across a participant and the average of the others (Fig. 4 right column). Contrasting this within subject correlation across our two conditions is then equivalent to a traditional PPI (Psycho-Physiological Interaction(Friston et al. 1997)), but could not reveal a specific change in our sample. The advantages of ISFC over a traditional PPI are in line with previous studies and is thought to be due to the fact that calculating correlations across participants rather than within participants reduces the impact of common noise across voxels(Simony et al. 2016; Nastase et al. 2019).

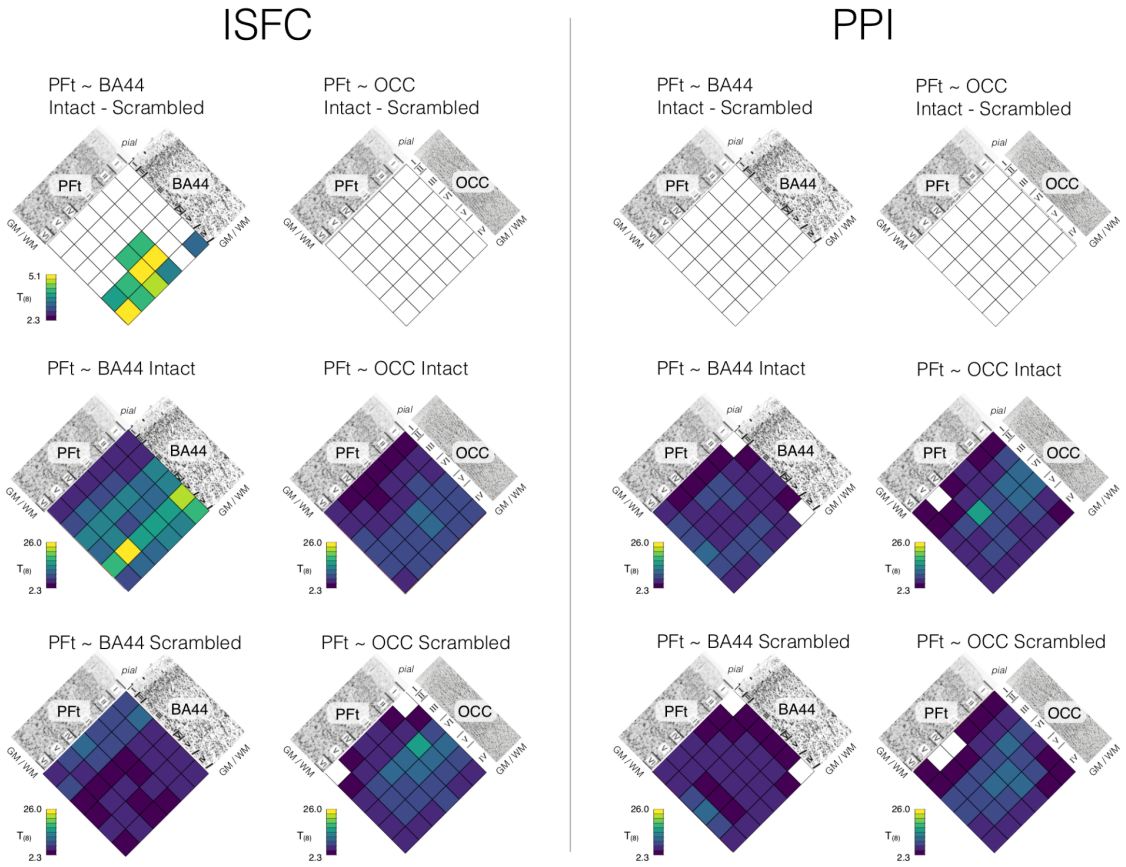

Fig. 4 - Results of the ISFC. Left column: Significant ISFC differences for the Intact - Scrambled contrast (top row), as well as for the Intact>0 (middle) and Scrambled>0 (bottom) contrast separately. **Right column:** Same but using a traditional PPI analysis using the same dataset. Note: ISFC calculates the correlation between time-courses in region A of each of the N participants and region B averaged across the other participants, and then contrasts these N values across conditions using a paired t-test. PPI calculates the correlation between regions A and B in the same participant for each of the N participants, and then contrasts these N values across conditions. The same color scale was used throughout to permit the comparison of t-values across

Discussion

Our aim was to shed further light onto the functional architecture of action observation by providing depth-resolved fMRI data that could help evaluate one of the core notions of predictive coding: that parietal nodes receive increased feed-back from premotor region BA44 when actions organize into predictable sequences. Specifically, we wanted to test whether sequence level information - isolated through the contrast intact-scrambled - (1) can be found in layers of PFt known to receive feedback connections from BA44, namely layers V/VI and III (Gerbella et al. 2011), and (2) is shared with deep layers of BA44. The analytical methods of ISC and ISFC were used to tackle the first and the second question, respectively.

As in our previous 3T MRI study (Thomas et al. 2018), at the macroscopic level we found that premotor, parietal and lateral occipital clusters of voxels encoded information about the observed motor acts, as measured by significant intersubject correlation while viewing the intact *or* scrambled movies. This was true whether we used non-parametric methods advised in the literature(Chen et al. 2016), or parametric methods we have recently shown to be more sensitive while at the same time controlling for Type-I errors(De Angelis et al. 2021). We then focused on PFt because predictive coding models and traditional feed-forward models of action observation make predictions that differ regarding the recruitment of deep layers in PFt: only such models predict that feed-back information should increase in intact over scrambled sequences, and hence that the contrast intact > scrambled should reveal increased ISC in deep layers (V,VI) and in layer III of PFt. To focus on voxels in PFt that represent the movies, we thus transformed the coordinates of the MNI voxels with significant information in at least one of the conditions at the group level (intact or scrambled) into the participants' native, unsmoothed space to define our PFt patch of interest. This patch was not defined based on a particular relationship between intact and scrambled ISC, and is thus not biased towards either condition(Kriegeskorte et al. 2009). In native space, using unsmoothed data, we then calculated the leave-one-out ISC between a participant's average activity in that bin and the average of the other participants in the same bin, and compared ISC in the intact and scrambled condition. As in our previous 3T study(Thomas et al. 2018), we found that PFt showed higher ISC while participants viewed intact than scrambled sequences. Importantly, for the first time we resolved the depth-profile of such sequence-level information. In accordance with predictive coding and the notion that predictive top-down information should be strongest in deeper layers (Kok et al. 2016; de Lange et al. 2018; Aitken et al. 2020), we found the significant difference with ISC intact > ISC scrambled to be observed in the deeper half of PFt, peaking at a depth that aligns with layer V in postmortem histological examinations of the brains of other individual(Caspers et al. 2006). In line with anatomical tracing studies in monkeys, suggesting that feedback connections from BA44 to PFt should terminate in layer III in addition to layers V/VI, this increased ISC was significant also at depth compatible with the location of layer III. Although we had no particular hypotheses for the depth profile of latOcc and BA44, as in our previous 3T study, we found that latOCC does not show significant differences in ISC between intact and scrambled sequences, irrespective of cortical depth, while BA44 did show significantly more ISC for intact vs scrambled conditions in superficial layers. See Supplementary Note 1 for further details.

By combining our comparatively higher spatial resolution with ISFC - to avoid the issue of shared noise across brain regions(Simony et al. 2016; Nastase et al. 2019) - we could also verify the prediction that the sequence level information at PFt depths aligning with layers III, V and VI is shared with the deep layers of BA44, as would be expected if BA44 were one of the sources of the additional, sequence-level information in PFt. Because this increased ISFC is calculated across different subjects, it indicates that it is signal relating to the intactness of the sequence in the stimuli that synchronizes across BA44 and PFt (Simony et al. 2016; Nastase et al. 2019). The depth of this channel of increased ISFC dovetails remarkably well with recent observations that predictive processes in other domains of visual processing are accompanied by an increased recruitment of the deeper layers of the cortical visual system (Kok et al. 2016; de Lange et al. 2018; Aitken et al. 2020), and with anatomical consideration that the dominant origin of feedback originates from deep layers in higher cortical region(Shipp 2007).

Although it is difficult to equate BOLD signal at a certain cortical depth with activity in a particular cortical layer, we did observe some increases in ISFC between superficial depths of PFt with deeper cortical depths of BA44. This would be in line with observations that feedback signals also terminate in Layer I in many cortical regions (Shipp 2007), but does not fit perfectly with the monkey data(Gerbella et al. 2011), that describes them to terminate in layers III, V and VI for the likely homologous regions. This could reflect a difference in species, or the fact that the monkey tracer studies, which are based on a small number of injections, may have missed such a perhaps weaker connection. It should be noted that the depth bins aligning with layer III in PFt, in which we measure increased ISC and ISFC for intact sequences, inevitably also align with the depth of the thin layer IV in PFt. This makes it impossible to verify our prediction that layer IV, thought to receive feed-forward information from visual regions, should not show increased ISC or ISFC for the Intact sequences. Even increasing the number of bins in our depth analysis cannot overcome this issue for two reasons: first, because capillaries flow through layer IV to reach layers V, VI so that any increase in ISC/ISFC in neurons in layers V or VI will also increase the ISC/ISFC in the BOLD signal in layer IV (Stephan et al. 2019); second, because resampling data in an increasing number of depth-bins after acquisition at 0.8mm voxel size will lead to vertical blurring.

Our study has a number of additional limitations. First, as often at high-field fMRI, we used fewer participants scanned twice rather than more participants (Baker et al. 2020; Cai et al. 2021) due to the difficulty of finding participants that can limit head movement to ensure stability at 0.8mm resolution. This may limit the degree to which we can extrapolate our findings to the general population, although the similarity of our results with those of a larger group acquired at 3T at the macroscopic level (Thomas et al. 2018), and the limited spread of our participant's r-values illustrates that the effects we report were quite robustly observed across participants. Second, to maintain our processing pipeline as close as possible to the actual data, we did not attempt to remove the bleeding of activation from deeper to more superficial layers due to the organization of the cortical vascularization (Stephan et al. 2019). Future studies using more elaborate models of how blood flows through the depth of the cortex may provide a clearer picture on whether signal changes in the intermediate levels aligning with layer IV across PFt and BA44 do or do not show local changes in ISFC. Third, the comparatively slow acquisition rate of our fMRI signal makes it impossible for us to tune accurately in the prediction error signals that we expect to occur at each camera change in our datasets. We are currently analysing data acquired using ECoG in epileptic patients to try to shed light on this faster temporal scale.

Finally, although we use depth resolved fMRI to quantify the level of feed-back input to PFt across conditions in ways that have become relatively frequent in the 7T literature, attributing activity at a particular depth to feed-back vs feed-forward information flow hinges (a) on the separability of these information flows along the cortical depth and (b) on our knowledge of where these inputs terminate. While in primary sensory regions such as V1 both these conditions appear to be met, with feed-forward signals peaking in Layer IV while feedback signals alter activity elsewhere (Self et al. 2019), the connectivity pattern between parietal and premotor regions in monkeys are, as we mentioned, more complex, with some frontal input to parietal regions terminating throughout all layers in a lateral connectivity pattern while others are multilayered, avoiding layer VI, albeit with a pattern that differs from the prototypical pattern described in early visual regions (Felleman and Van Essen 1991; Rozzi et al. 2006; Shipp 2007; Gerbella et al. 2010, 2011). While studies like ours have thus been argued to be necessary to start exploring whether the prevalence of predictive feed-back signals in deep layers observed in the visual system applies elsewhere (Finn et al. 2020), there is also

urgent need to perform the kind of laminar recordings that have benchmarked the foundations of this approach in the visual system(Self et al. 2019) across posterior parietal and premotor regions.

Despite these limitations, we hope that this study will serve as a proof of concept that depth-resolved analyses can be performed during naturalistic viewing in higher cognitive brain regions using intersubject correlation-based approaches, and that we may contribute to what Finn and colleagues called for: “layer fMRI is now at a point where we can expand from tightly controlled experiments in sensory cortex with clear hypotheses—which were necessary to show feasibility of the technique—to more exploratory, data-driven investigations of functional dynamics both across the cortical hierarchy as well as within higher-order regions themselves. […] Data acquired during naturalistic stimulation—e.g., movie watching—lend itself to both connectivity and activation analyses” (Finn et al. 2020). Ultimately such studies may literally contribute to a deeper understanding of social cognition.

Author contributions

Following the Contributor Roles Taxonomy (CRediT) available at https://casrai.org/credit/. The relative amount of markers highlights the degree of contribution within each role.

Author	RB	LC	LdA	WvdZ	PLB	VG	CK
Conceptualization	**					***	***
Data curation	*	***					
Formal analysis	*	***	**				
Funding acquisition	***					***	***
Investigation	***			**			
Methodology	***	***	***	**	**	***	***
Project Administration	**	**				***	***
Resources	***					***	***
Software		***	**		**		
Supervision		*				***	***
Validation		***	***				
Visualization		***				*	*
Writing original draft	*	***	**			***	***
Writing review and editing	*	*	*	*	*	*	*

Acknowledgements

We thank Giuseppe Luppino for guiding our hypotheses about the layers of PFt that should receive BA44 input based on the tracing studies in monkeys from his lab. We thank Teresa de Sanctis for help with preparing the stimuli. We thank all the members of the Spinoza centre where the scanning was conducted for providing technical support. We thank Robert Turner for several helpful discussions on depth specific 7T fMRI.

Funding

Funding: The work was funded by the BIAL foundation grant 255/16 and the European Commission ERC grant VicariousBrain (ERC-StG 312511). CK was funded by VICI grant 453-15-009, VG by ERC grant ERC-StG 758703.

References

Aitken F, Menelaou G, Warrington O, et al (2020) Prior expectations evoke stimulus-specific activity in the deep layers of the primary visual cortex. PLoS Biol 18:e3001023. https://doi.org/10.1371/journal.pbio.3001023

Amunts K, Schleicher A, Burgel U, et al (1999) Broca's region revisited: cytoarchitecture and intersubject variability. J Comp Neurol 412:319–341

Amunts K, Zilles K (2012) Architecture and organizational principles of Broca's region. Trends Cogn Sci 16:418–426. https://doi.org/10.1016/j.tics.2012.06.005

Avants BB, Epstein CL, Grossman M, Gee JC (2008) Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12:26–41. https://doi.org/10.1016/j.media.2007.06.004

Avants BB, Tustison NJ, Song G, et al (2011) A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage 54:2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025

Baker DH, Vilidaite G, Lygo FA, et al (2020) Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. Psychol Methods. https://doi.org/10.1037/met0000337

Bazin P-L, van der Zwaag W, Bhandari R, et al (2018) Sulcal ridge alignment for laminar fMRI at 7T

Bazin P-L, Weiss M, Dinse J, et al (2014) A computational framework for ultra-high resolution cortical segmentation at 7Tesla. Neuroimage 93 Pt 2:201–209

Bogovic JA, Prince JL, Bazin P-L (2013) A Multiple Object Geometric Deformable Model for Image Segmentation. Comput Vis Image Underst 117:145–157

Bonini L, Rozzi S, Serventi FU, et al (2010) Ventral premotor and inferior parietal cortices make distinct contribution to action organization and intention understanding. Cereb Cortex N Y N 1991 20:1372–1385. https://doi.org/10.1093/cercor/bhp200

Cai Y, Hofstetter S, van der Zwaag W, et al (2021) Individualized cognitive neuroscience needs 7T: Comparing numerosity maps at 3T and 7T MRI. NeuroImage 237:118184. https://doi.org/10.1016/j.neuroimage.2021.118184

Caspers S, Geyer S, Schleicher A, et al (2006) The human inferior parietal cortex: Cytoarchitectonic parcellation and interindividual variability. NeuroImage 33:430–448. https://doi.org/10.1016/j.neuroimage.2006.06.054

Caspers S, Zilles K, Laird AR, Eickhoff SB (2010) ALE meta-analysis of action observation and imitation in the human brain. NeuroImage 50:1148–67. https://doi.org/10.1016/j.neuroimage.2009.12.112

Chen G, Shin Y-W, Taylor PA, et al (2016) Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. NeuroImage 142:248–259. https://doi.org/10.1016/j.neuroimage.2016.05.023

De Angelis L, Gazzola V, Keysers C (2021) Parametric tests for Leave-One-Out Inter-Subject~Correlations in fMRI provide adequate Type I error control while providing high sensitivity. bioRxiv. https://doi.org/10.1101/2020.07.16.206235

de Lange FP, Heilbron M, Kok P (2018) How Do Expectations Shape Perception? Trends Cogn Sci 22:764–779. https://doi.org/10.1016/j.tics.2018.06.002

Eickhoff SB, Stephan KE, Mohlberg H, et al (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps

and functional imaging data. Neuroimage 25:1325–1335

Fan X, Bazin P-L, Prince JL (2008) A Multi-Compartment Segmentation Framework With Homeomorphic Level Sets. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 1–6

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47

Finn ES, Huber L, Bandettini PA (2020) Higher and deeper: Bringing layer fMRI to association cortex. Prog Neurobiol 101930. https://doi.org/10.1016/j.pneurobio.2020.101930

Friston K, Mattout J, Kilner J (2011) Action understanding and active inference. Biol Cybern 104:137–160. https://doi.org/10.1007/s00422-011-0424-z

Friston KJ, Buechel C, Fink GR, et al (1997) Psychophysiological and Modulatory Interactions in Neuroimaging. NeuroImage 6:218–229. https://doi.org/10.1006/nimg.1997.0291

Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. Brain J Neurol 119 (Pt 2):593–609. https://doi.org/10.1093/brain/119.2.593

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. Cereb Cortex 19:1239–1255. https://doi.org/10.1093/cercor/bhn181

Gerbella M, Belmalih A, Borra E, et al (2010) Cortical Connections of the Macaque Caudal Ventrolateral Prefrontal Areas 45A and 45B. Cereb Cortex 20:141–168. https://doi.org/10.1093/cercor/bhp087

Gerbella M, Belmalih A, Borra E, et al (2011) Cortical connections of the anterior (F5a) subdivision of the macaque ventral premotor area F5. Brain Struct Funct 216:43–65. https://doi.org/10.1007/s00429-010-0293-6

Han X, Pham DL, Tosun D, et al (2004) CRUISE: cortical reconstruction using implicit surface evolution. Neuroimage 23:997–1012

Hasson U, Nir Y, Levy I, et al (2004) Intersubject synchronization of cortical activity during natural vision. Science 303:1634–40. https://doi.org/10.1126/science.1089506

Honey CJ, Thesen T, Donner TH, et al (2012) Slow Cortical Dynamics and the Accumulation of Information over Long Timescales. Neuron 76:423–434. https://doi.org/10.1016/j.neuron.2012.08.011

Huber L, Handwerker DA, Jangraw DC, et al (2017) High-Resolution CBV-fMRI Allows Mapping of Laminar Activity and Connectivity of Cortical Input and Output in Human M1. Neuron 96:1253-1263.e7

Huntenburg JM, Steele CJ, Bazin P-L (2018) Nighres: processing tools for high-resolution neuroimaging. GigaScience 7:. https://doi.org/10.1093/gigascience/giy082

Keller PE, Knoblich G, Repp BH (2007) Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. Conscious Cogn 16:102–111. https://doi.org/10.1016/j.concog.2005.12.004

Keysers C, Gazzola V (2014) Hebbian learning and predictive mirror neurons for actions, sensations and emotions. Philos Trans R Soc B Biol Sci 369:20130175–20130175. https://doi.org/10.1098/rstb.2013.0175

Keysers C, Paracampo R, Gazzola V (2018) What neuromodulation and lesion studies tell us about the function of the mirror neuron system and embodied cognition. Curr Opin Psychol 24:35–40. https://doi.org/10.1016/j.copsyc.2018.04.001

Keysers C, Perrett DI (2004) Demystifying social cognition: a Hebbian perspective. Trends Cogn Sci 8:501–7. https://doi.org/10.1016/j.tics.2004.09.005

Kilner JM, Frith CD (2008) Action observation: inferring intentions without mirror neurons. Curr Biol CB 18:R32-3. https://doi.org/10.1016/j.cub.2007.11.008

Klein A, Andersson J, Ardekani BA, et al (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46:786–802. https://doi.org/10.1016/j.neuroimage.2008.12.037

Kok P, Bains LJ, van Mourik T, et al (2016) Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. Curr Biol 26:371–376. https://doi.org/10.1016/j.cub.2015.12.038

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular Analysis in Systems Neuroscience: {T}he Dangers of Double Dipping. Nat Neurosci 12:535–540

Lawrence SJ, Norris DG, de Lange FP (2019a) Dissociable laminar profiles of concurrent bottom-up and top-down modulation in the human visual cortex. eLife 8:. https://doi.org/10.7554/eLife.44422

Lawrence SJD, Formisano E, Muckli L, de Lange FP (2019b) Laminar fMRI: Applications for cognitive neuroscience. NeuroImage 197:785–791. https://doi.org/10.1016/j.neuroimage.2017.07.004

Lawrence SJD, van Mourik T, Kok P, et al (2018) Laminar Organization of Working Memory Signals in Human Visual Cortex. Curr Biol CB 28:3435-3440.e4. https://doi.org/10.1016/j.cub.2018.08.043

Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J Neurosci Off J Soc Neurosci 31:2906–15. https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Liu TT (2016) Noise contributions to the fMRI signal: An overview. NeuroImage 143:141–151. https://doi.org/10.1016/j.neuroimage.2016.09.008

Logothetis NK (2003) The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci Off J Soc Neurosci 23:3963–3971

Marques JP, Kober T, Krueger G, et al (2010) MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. NeuroImage 49:1271–1281. https://doi.org/10.1016/j.neuroimage.2009.10.002

Nastase SA, Gazzola V, Hasson U, Keysers C (2019) Measuring shared responses across subjects using intersubject correlation. Soc Cogn Affect Neurosci. https://doi.org/10.1093/scan/nsz037

Nelissen K, Borra E, Gerbella M, et al (2011) Action observation circuits in the macaque monkey cortex. J Neurosci Off J Soc Neurosci 31:3743–3756. https://doi.org/10.1523/JNEUROSCI.4803-10.2011

Nelissen K, Luppino G, Vanduffel W, et al (2005) Observing others: multiple action representation in the frontal lobe. Science 310:332–336. https://doi.org/10.1126/science.1115593

Nishitani N, Hari R (2000) Temporal dynamics of cortical representation for action. Proc Natl Acad Sci U S A 97:913–918. https://doi.org/10.1073/pnas.97.2.913

Rizzolatti G, Sinigaglia C (2016) The mirror mechanism: a basic principle of brain function. Nat Rev Neurosci 17:757–765. https://doi.org/10.1038/nrn.2016.135

Rozzi S, Calzavara R, Belmalih A, et al (2006) Cortical Connections of the Inferior Parietal Cortical Convexity of the Macaque Monkey. Cereb Cortex 16:1389–1417. https://doi.org/10.1093/cercor/bhj076

Self MW, van Kerkoerle T, Goebel R, Roelfsema PR (2019) Benchmarking laminar fMRI: Neuronal spiking and synaptic activity during top-down and bottom-up processing in the different layers of cortex. NeuroImage 197:806–817. https://doi.org/10.1016/j.neuroimage.2017.06.045

Shipp S (2007) Structure and function of the cerebral cortex. Curr Biol 17:R443–R449. https://doi.org/10.1016/j.cub.2007.03.044

Simony E, Honey CJ, Chen J, et al (2016) Dynamic reconfiguration of the default mode network during narrative comprehension. Nat Commun 7:12141. https://doi.org/10.1038/ncomms12141

Stephan KE, Petzschner FH, Kasper L, et al (2019) Laminar fMRI and computational theories of brain function. Neuroimage 197:699–706

Thomas RM, De Sanctis T, Gazzola V, Keysers C (2018) Where and how our brain represents the temporal structure of observed action. NeuroImage 183:677–697. https://doi.org/10.1016/j.neuroimage.2018.08.056

Trampel R, Bazin P-L, Pine K, Weiskopf N (2019) In-vivo magnetic resonance imaging (MRI) of laminae in the human cortex. NeuroImage 197:707–715. https://doi.org/10.1016/j.neuroimage.2017.09.037

Umiltà MA, Kohler E, Gallese V, et al (2001) I know what you are doing. a neurophysiological study. Neuron 31:155–165. https://doi.org/S0896-6273(01)00337-3 [pii]

van der Zwaag W, Buur PF, Fracasso A, et al (2018a) Distortion-matched T1 maps and unbiased T1-weighted images as anatomical reference for high-resolution fMRI. NeuroImage 176:41–55. https://doi.org/10.1016/j.neuroimage.2018.04.026

van der Zwaag W, Buur PF, Fracasso A, et al (2018b) Examples of sub-millimeter, 7T, T1-weighted EPI datasets acquired with the T123DEPI sequence. Data Brief 20:415–418. https://doi.org/10.1016/j.dib.2018.08.030

von Economo C (1927) Zellaufbau der Grosshirnrinde des Menschen. Zehn Vorlesungen mit 61 Abbildungen. Springer, Berlin

Waehnert MD, Dinse J, Schäfer A, et al (2016) A subject-specific framework for in vivo myeloarchitectonic analysis using high resolution quantitative MRI. Neuroimage 125:94–107

Waehnert MD, Dinse J, Weiss M, et al (2014) Anatomically motivated modeling of cortical laminae. Neuroimage 93 Pt 2:210–220

Supplementary Online Materials

Supplementary Figure S1 : Field of view of the fMRI acquisitions

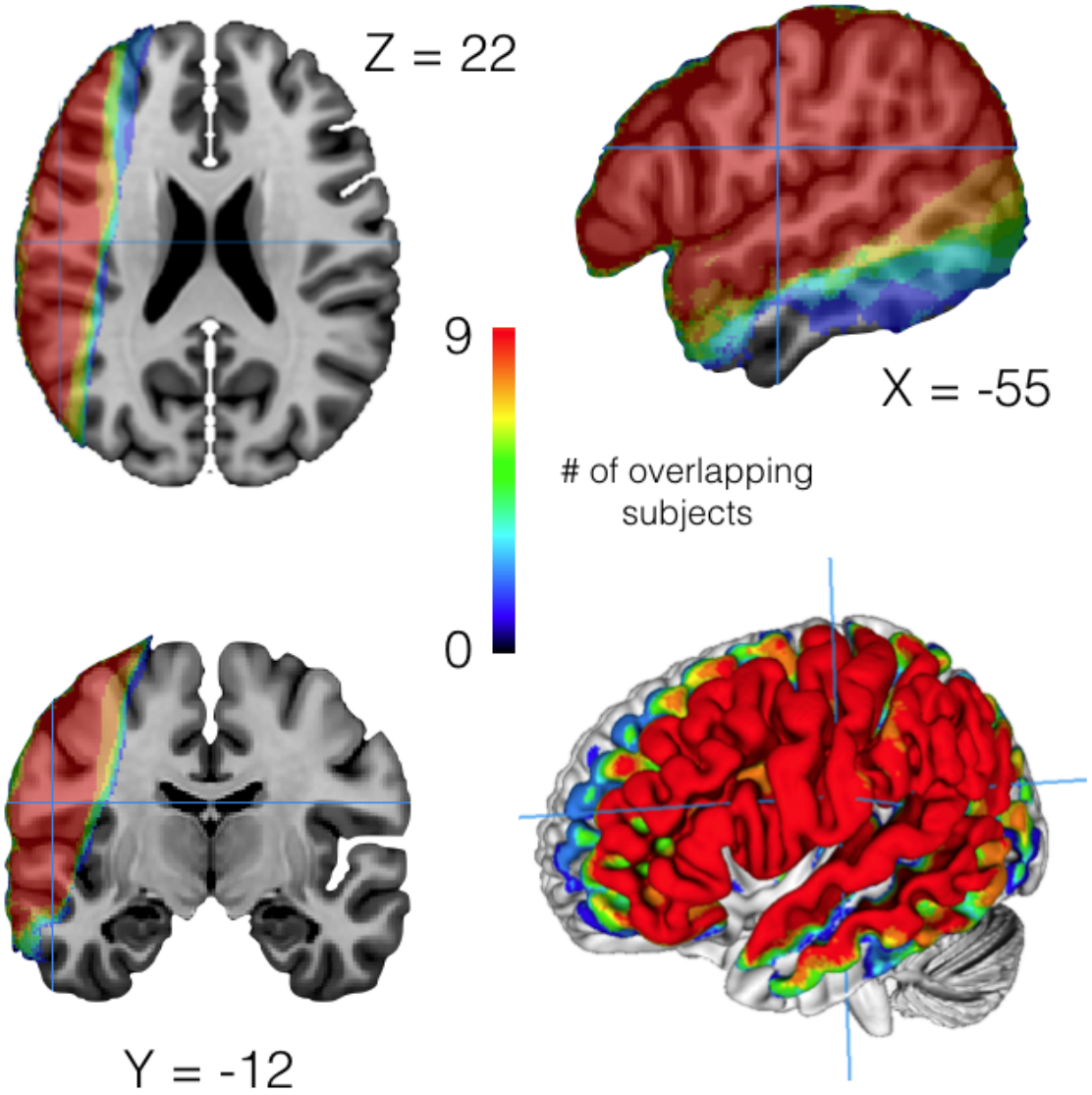

Figure S1: Coverage of the left hemisphere. We centered our planning of the 7T acquisition on the parietal area PFt. Because the planning was done subject by subject, not the exact same coverage in terms of MNI coordinates was achieved in all participants. Here we therefore show the overlap in coverage after transforming the field of view into MNI space. As can be seen, we achieved consistent coverage in inferior parietal and frontal regions as well as in anterior lateral occipital cortex.

Supplementary Note 1 and Figure S2: ISC as a function of depth in LatOcc and BA44

Although we had no particular hypotheses for the depth profile of latOcc and BA44, as in our previous 3T stud(Thomas et al. 2018), we found that latOCC does not show significant differences in ISC between intact and scrambled sequences, irrespective of cortical depth (Supplementary Figure S2a). This is in line with other studies showing that more anterior regions integrate information over longer periods than early sensory regions, be they visual or auditory(Lerner et al. 2011; Honey et al. 2012). In BA44 we found a more complex pattern of differences between intact and scrambled sequences. The two most superficial depth bins showed ISC that was higher in intact than scrambled conditions, while the deepest depth bin showed the opposite effect. While finding intact>scrambled in BA44 was expected, that this difference was most marked at superficial depth was not: based on previous findings in visual brain regions, we expected sequence-level information to be strongest in the depth of BA44(Kok et al. 2016; de Lange et al. 2018; Aitken et al. 2020). Furthermore, that the deepest parts of BA44 would show more ISC for scrambled sequences was entirely unexpected and we feel uncertain about how to interpret this finding. Some participants reported that viewing the scrambled sequences created an oddly dissonant experience of cognitively understanding what action was being performed at a more abstract level despite the perceptual frustration of a visual sequence that constantly violated expectations in the timing of the acts. In a way, the sum of the acts made sense at an abstract level, but their sequence did not. Perhaps, one could thus speculate that this hybrid pattern in BA44, with relatively increased ISC in superficial voxels for intact and in deep voxels for scrambled sequences could represent a switch in the dominant processes, with the more superficial voxels processing sequence level information embedded in the temporal stream of the visual input (which is more reliable in the intact sequence) and deeper voxels processing top-down input from cognitive processing that is more detached from the temporal stream of the visual input, and becomes more critical when the visual input itself does not make sense (in the scrambled condition). However, such an interpretation is very tentative indeed.

It is also worth noting that ISFC with PFt increases for intact sequences compared to scrambled sequences in the depth of BA44 while we find here that ISC in BA44 is actually not increased at the same depth in BA44 for intact compared to scrambled sequences. While this may appear contradictory at first sight, it is important to note that ISFC changes in BA44 focus on signals consistent in timing across participants that is also shared with PFt, while ISC simply quantifies signals consistent across participants. It thus appears that for the intact sequences, although in the depth of BA44 signals overall are less consistent across participants than for scrambled sequences, the part of the variance over time shared with PFt does increase.

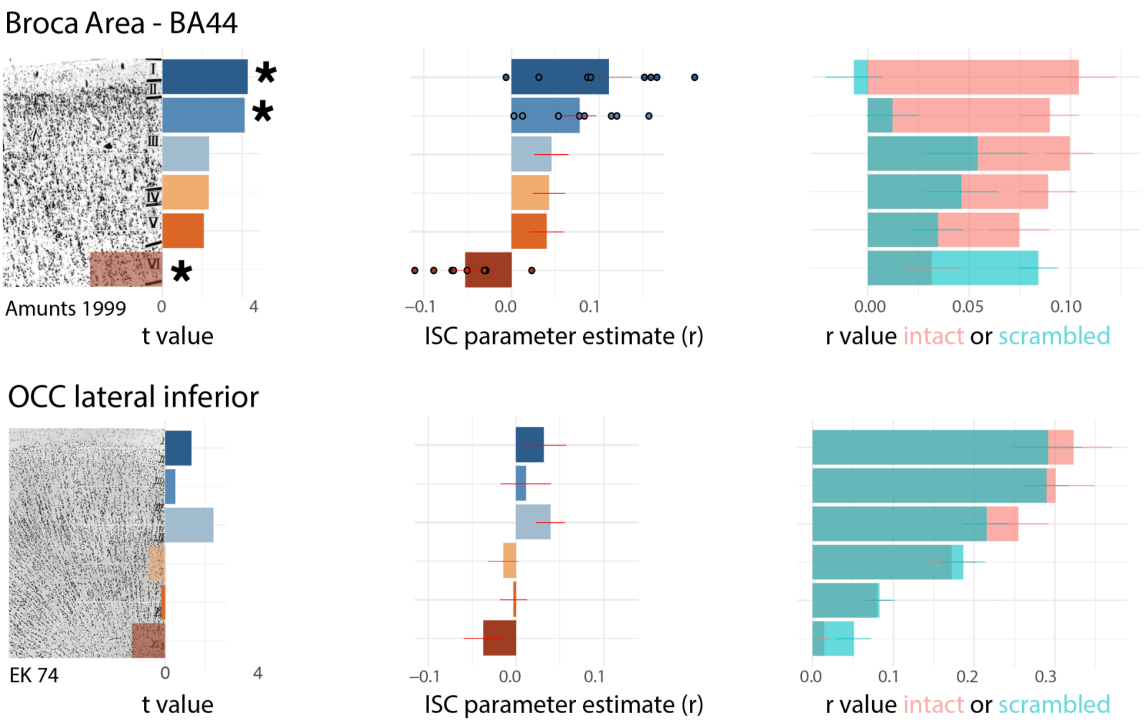

Figure S2: ISC as a function of depth in LatOcc and BA44. All conventions as in Figure 3, but for Lateral Occipital cortex, using a postmortem analysis from (von Economo 1927), or BA44 using a postmortem analysis from Amunts et al.(1999).

Supplementary Figure S3 : Consistency of ISC results across binning

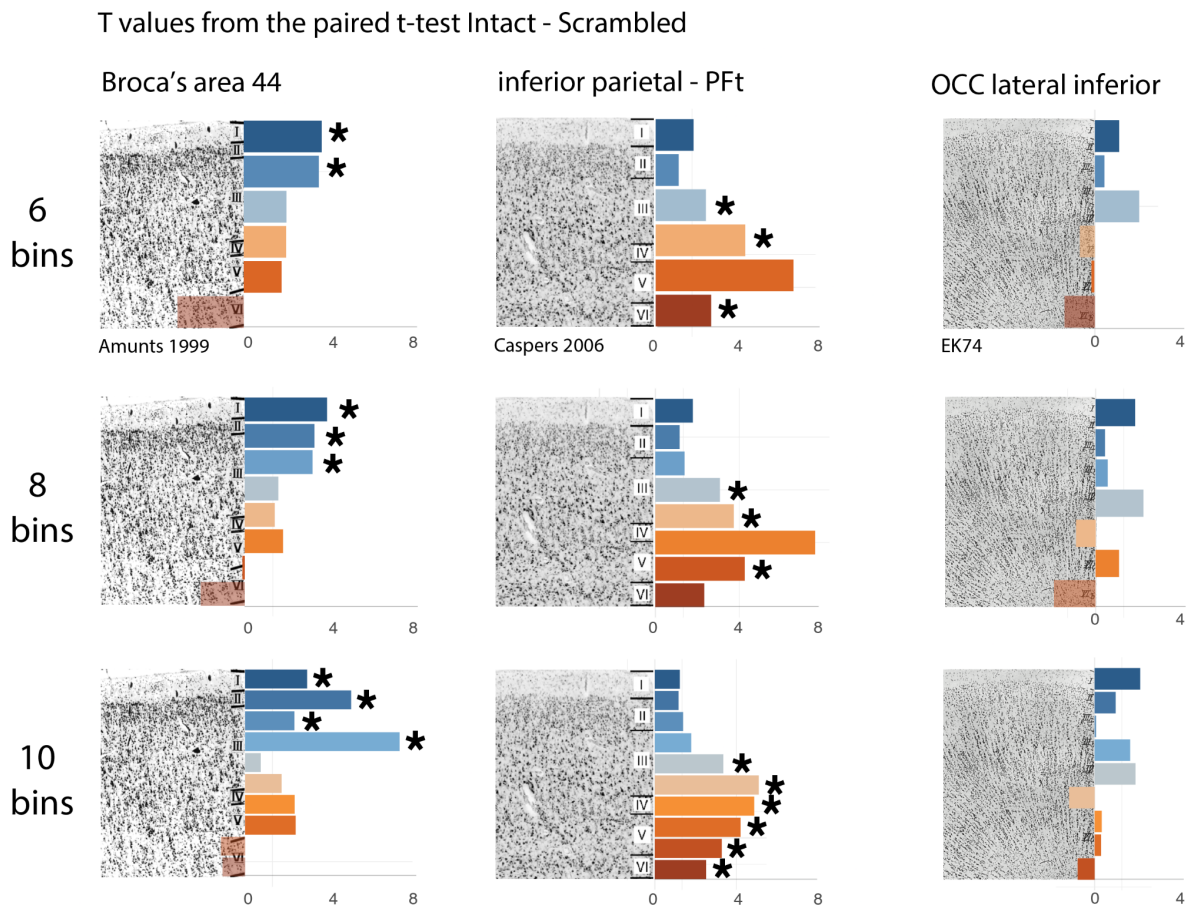

Fig. S3 Consistency across binning choices. Bars represent the t value from comparing the ISC estimates in either condition using a paired t-test. Significance was determined using FDR correction (q(FDR) = 0.05 - relative to the number of bins) across all bins within each region, and is indicated by the asterisk.

Supplementary Materials 4 : Quality control images for preprocessing steps: skull stripping, registration, transformation of localizer/atlas following registration, layering

This folder of the github repo (https://github.com/ldeangelisphys/layerfMRI/tree/main/QC_images) contains quality control images documenting the outcome of the following procedures:

Procedure	Example QC image
summary_skullstrip: outcome of the skull-stripping for the partial FOV fMRI and low-res anatomical, as well as of the high-res anatomical image.	https://github.com/ldeangelisphys/layerfMRI/blob/main/QC_images/summary_skullstrip/sub_02_skullstrip.pdf
summary_registration and **summary_inv_registration**: registration between different spaces: fMRI native (4 partial FOV sessions per subject), low-res	https://github.com/ldeangelisphys/layerfMRI/blob/main/QC_images/summary_registration/sub_02_registration.pdf https://github.com/ldeangelisphys/layerfMRI/blob/main/QC_images/summary_inv_registration/sub_02_inv_registration.pdf
summary_layering: high-res anatomical skull strip, cortical segmentation and cortical depth estimation	https://github.com/ldeangelisphys/layerfMRI/blob/main/QC_images/summary_layering/sub_02_layering.pdf
summary_trans2native: transformation of the cortical depth maps into the native fMRI space	https://github.com/ldeangelisphys/layerfMRI/blob/main/QC_images/summary_trans2native/sub_02_native_space.pdf